Imperial College London

Imperial College Business School

Machine Learning Applications in Augmenting Mispricing Recognition and Mean Reversion Strategy

2021-2022 Applied Project
Quantitative Analysis - Investment Strategy
MSc Investment & Wealth Management

Chin-Fu Kuo
CID: 02116577

Contents

1 Client Specification 1
2 Introduction 2
3 Data & Methodology 3
3.1 Datasets 3
3.1.1 Source 3
3.1.2 Data Preprocessing 3
3.1.3 Differences In Data & Methodology 3
3.2 Implementation Of Mean Reversion Strategies 4
3.2.1 OLS Regression Applied Strategy (Benchmark) 4
3.2.2 8 Machine Learning Augmented Strategies 4
3.3 Performance Metrics 6
4 Results 7
4.1 Models Trained To Recognise Mispricing 7
4.2 Performance Comparisons Across Strategies 7
4.3 Performance Comparisons Across Periods 8
4.4 Feature Importance 9
4.5 Explaining Returns With 6-Factor Time Series Regressions 9
4.6 Correlations Between 9 Strategies 10
5 Conclusion 10
6 References 11
7 Appendix 12
7.1 Datasets, Definitions & Source 12
7.2 Notation, Model Outline & Performance Metrics 13
7.3 The Optimal Selection For Hyper-parameters 14
7.4 Performance Statistics (For The Whole Sample Period) 16
7.5 Performance Statistics (1998 - 2012 vs 2013 - 2022) 18
7.6 Features Importance 20
7.7 6 Factor Time-Series Regressions 22
7.8 Correlations Between 9 Strategies & Graphs for Cumulative Returns 23

1 Client Specification

Our client is an eminent asset management firm based in the USA, which requests to stay anonymous. The firm seeks to strengthen the mean reversion strategy currently implemented by its quantitative investment team. The client notices the potential of the well-known paper by Söhnke Matthias Bartram and Mark Grinblatt with the title “Agnostic Fundamental Analysis Works”. The paper proposes a peer-implied valuation approach based on the OLS regression model to estimate fair equity values for listed companies. Recognising the mispricing by calculating the degree of deviation of a company’s market capitalisation from its peer-implied fair equity value brings a new possibility, also a potential “Quantamental” approach, to the mean reversion strategy.

The client aims to comprehend (1) whether the mispricing signal in the USA stock market still works effectively implemented with data in recent years; (2) whether the performance of the valuation approach proposed by Bartram and Grinblatt (2018 & 2021) can be further enhanced by applying machine learning methods in boosting the mispricing recognition; (3) the performance statistics of mean reversion strategy based on the Bartram and Grinblatt (2018 & 2021) approach after augmented by different machine learning models. (4) whether any known market anomalies (risk factors) can explain the strategy.

2 Introduction

Prior to initiating further research according to the requirements, this work illustrates the assumptions made, the methodology applied, and the results produced by Bartram and Grinblatt (2018 & 2021).

To generate the peer-implied fair equity value, Bartram and Grinblatt (2018 & 2021) make two assumptions. First, they assume that the stock market as a whole has a high degree of efficiency, meaning most firms' market capitalisations are usually close to their fair equity values. Instead, an individual firm's market capitalisation can deviate from its fair equity value. Second, the Law of One Price holds under most circumstances, which means deviations from fair value tend to contract than expand.

Based on the two assumptions, Bartram and Grinblatt (2018 & 2021) propose an approach to estimating firms' fair equity value via relative valuation. First, they run a cross-sectional OLS regression of firms' 21 accounting items $x_{j,1,t},\ldots,x_{j,21,t}$ on their market capitalisations $\mathrm{MV_{j,t}}$ to obtain the coefficients $b_{1,t},\ldots,b_{21,t}$ and the intercept a_{t}, i.e.:

$$\mathrm{MV_{j,t}} = a_t + b_{1,t} \cdot x_{j,1,t} + \cdots + b_{21,t} \cdot x_{j,21,t} \tag{1}$$

In which j denotes the firm, t denotes time. Next, they plug each firm's accounting items into the model with the above coefficients to calculate its peer-implied fair equity value, $\mathrm{FV}_{j,t}$.

$$\mathrm{FV}_{j,t} = a_t + b_{1,t} \cdot \mathrm{x}_{j,1,t} + \cdots + b_{21,t} \cdot \mathrm{x}_{j,21,t} \tag{2}$$

Firm $j's$ date t mispricing signal, $\mathrm{MS}_{j,t}$, can thus be defined as the percentage difference between its date t peer-implied fair equity value, $\mathrm{FV}_{j,t}$ and its market capitalisation, $\mathrm{MV_{j,t}}$.

$$\mathrm{MS}_{j,t} = \frac{FV_{j,t} - MV_{j,t}}{MV_{j,t}} \tag{3}$$

Bartram and Grinblatt (2018 & 2021) rank firms at the end of each month according to the mispricing signals and classify them into 5-quintile groups. In the following month, they construct a mean reversion strategy by imposing a long position on the most underpriced group and executing a short position on the most overpriced group, with a holding period of 1 month. During the backtesting period, 310 return months, from March 1987 to December 2012, the average monthly spread of their long/short equally-weighted portfolio reaches 0.42%, with 58.1% of months having positive returns over the period.

Overall, the findings are persistent with the original paper. The mispricing anomaly continuously exists during the backtesting period, 291 return months, from April 1998 to June 2022. Replacing the OLS regression model with ML models raises the strategy's average returns, reduces volatility, and enhances the mispricing recognition's precision. The fact that they have a lower turnover ratio also reduces the transaction costs, making the strategy more feasible and profitable.

3 Data & Methodology

3.1 Datasets

3.1.1 Source

Overall, equity stock is the only asset class involved in this work. Three primary datasets, which range 291 months, from April 1998 to June 2022, are utilised. Their information is outlined in Appendix 7.1. All currently active firms and delisted firms once listed in the US stock market are included.

3.1.2 Data Preprocessing

The sample construction in this work almost follows Bartram and Grinblatt (2018 & 2021). First, to avoid investing in illiquid or tiny stocks, only firms listed on the NYSE, AMEX or NASDAQ-NMS over the 1998 - 2022 period are selected. Besides, stocks with a monthly market capitalisation of less than $300 million or a monthly stock price of less than $10 are excluded. Second, non-common stocks, secondary listing and foreign listing stocks are excluded. (Matthias et al., 2021) Third, for all remaining firms, non-missing values for market capitalisation and all accounting items are required, except for the item “cash dividends”, since its values are sometimes mistakenly recorded as missing in non-dividend-paying months. For these errors, the missing values are replaced with 0. Then, any data point with a missing value is dropped. Fourth, since the quarterly accounting items are tagged with their data date instead of the available date, all values of accounting items for each firm are manually lagged for 3 months to avoid look-ahead bias. Finally, all quarterly accounting items are converted to monthly frequency by filling the missing values in gap months between quarter ends with the values from the prior-quarter accounting items. After the data preprocessing step, the final dataset contains 7,611 unique firms in total and 1,981 unique firms on average monthly.

3.1.3 Differences In Data & Methodology

First, Bartram and Grinblatt (2018 & 2021) use the price, return and the number of outstanding shares data from the CRSP database. By contrast, the Compustat database is used as the source for the data in this work. Next, Bartram and Grinblatt (2018 & 2021) sum the most recent four quarterly values of accounting items from the income and cash flow statements when training the OLS regression model. In contrast, this step is not done in this work to avoid distorting the characteristics of the data. Finally, compared to the dataset used by Bartram and Grinblatt (2018 & 2021), a stricter market capitalisation and stock price threshold is applied in this work to filter out illiquid and micro-cap stocks.

3.2 Implementation Of Mean Reversion Strategies

All strategies are backtested with the monthly frequency data and equally-weighted portfolio. Compared to Bartram and Grinblatt (2018 & 2021), which only sort stocks into quintile groups, sorting stocks into 5-quintile and 10-decile groups based on mispricing signals is applied in this work. The notations used for the 9 strategies are outlined in Appendix 7.2.

3.2.1 OLS Regression Applied Strategy (Benchmark)

Due to the data source and methodology differences, it is inappropriate to directly compare the performance statistics from Bartram and Grinblatt (2018 & 2021) with those of the ML augmented models in this work. Therefore, an OLS regression model is created as the benchmark. Apart from the differences mentioned, the OLS regression model closely follows the approach proposed by Bartram and Grinblatt (2018 & 2021).

At the end of each month, a cross-sectional OLS regression of firms' 21 contemporaneous available accounting items on their market capitalisations is created to calculate the fair equity value and the mispricing signals used to classify stocks. The signal generation starts in April 1998 and ends in May 2022.

The following month, a long (short) position is implemented for the most underpriced (overpriced) groups based on the corresponding mispricing signals. The holding period is 1 month. The above process is repeated month by month over the whole sample period. Thus, the trading period ranges from May 1998 to June 2022.

3.2.2 8 Machine Learning Augmented Strategies

The Random Forest (Breiman, 2001), XGBoost (eXtreme Gradient Boosting, Chen et al., 2016), Neural Network (Hopfield, 1982), and LSTM (Long-short Term Memory, Hochreiter et al., 1997) models are used. The standard approach of ML applications and out-of-sample prediction (Gu et al., 2020) is firmly followed in this work, except for the hyper-parameter tuning process. Eight ML algorithms are used to augment the valuation approach from Bartram and Grinblatt (2018 & 2021). The corresponding training and testing windows for the ML models are provided in Appendix 7.2.

Except for the LSTM model, a fixed 6-month rolling window that moves forward by 1 month each time is applied for training the ML models. Models are trained by each firm's 21 accounting items (feature variables) within each rolling training window, and their target variable is each firm's market capitalisations within the same window. The month after the training window serves as the out-of-sample testing data in which the trained models will be fed with each firm's 21 accounting items and predict each firm's fair equity value, as indicated in Figure 1.

Since aiming to take the most advantage of its capability in capturing the pattern hidden in sequential data, fixed sliding training and testing windows of 12 months that moves forward by 1 month each time are implemented on the LSTM model. In this case, the sliding window scheme means the LSTM model is trained with the first 12-month target and feature variables. The model will then be fed with the 2nd to 13th months' feature variables to predict the fair equity value in the 13th month and so on, as indicated in Figure 2. Besides, since the shape of LTSM's input array is fixed, when the model is trained or used to predict fair equity values, the stocks with less than 12 months of data within the window must be excluded.

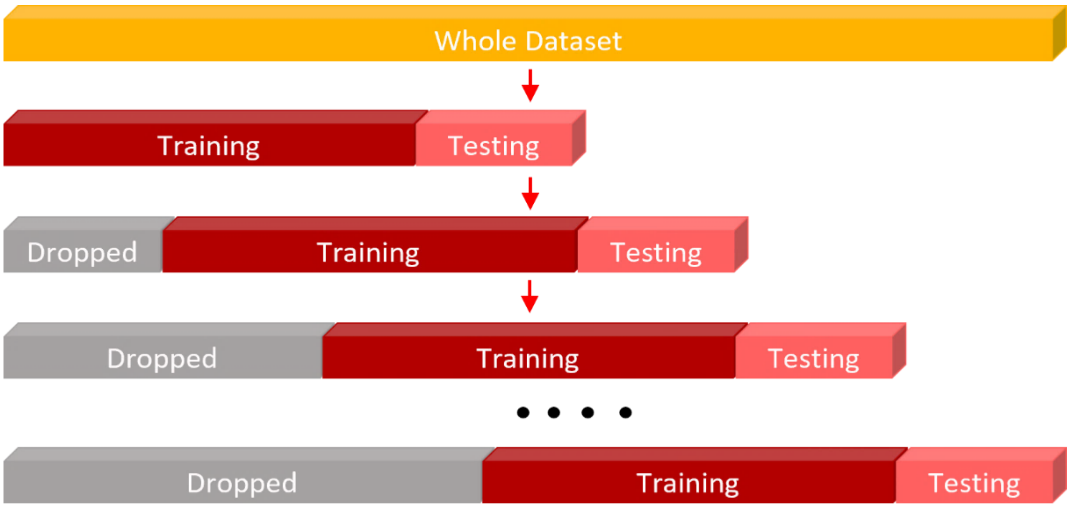

Figure 1: Fixed Window Rolling Scheme

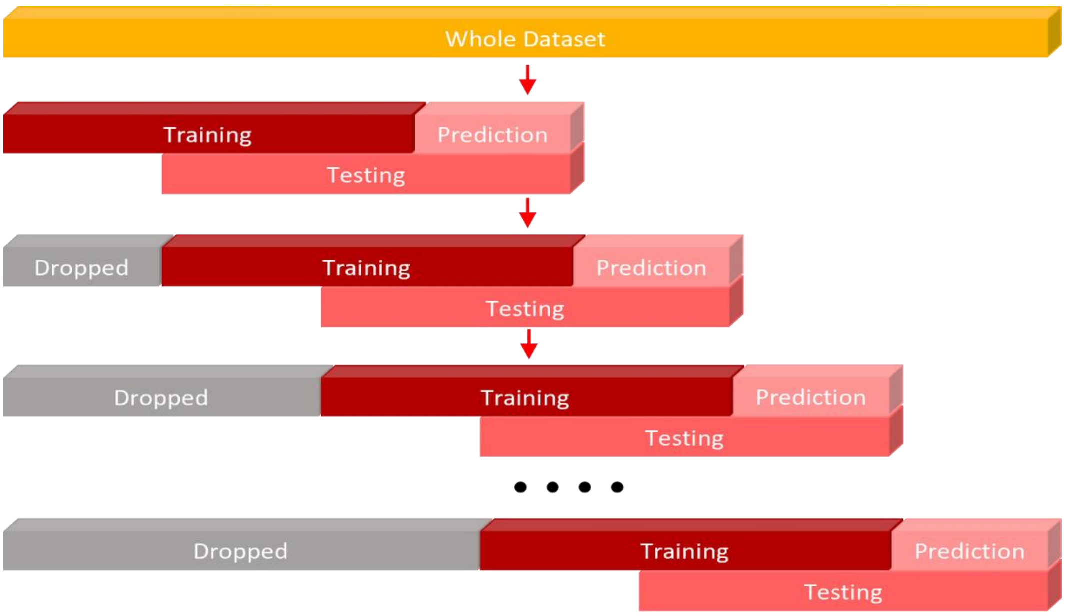

Figure 2: Fixed Sliding Window Scheme for LSTM Model

Additionally, due to the Neural Network and LSTM models' favour in a mean-zero dataset, a standardisation process is executed on each feature variable individually within the training and testing windows before being fed into the models. In contrast, this step is not done for Random Forest and XGBoost models.

After the training and predicting step, the trained models calculate the mispricing signal, and the stocks are sorted based on it. A long (short) position will be implemented for the most underpriced (overpriced) groups in the next month, with a holding period of 1 month. The whole rolling or sliding process is repeated over the data period. It is worth mentioning that since the ML models, except for the LSTM, are trained on a 6-month period, their signal generation starts in October 1998 and ends in May 2022. Thus, their trading period ranges from November 1998 to June 2022. By contrast, the LSTM model is trained on a 1-year period, so its signal generation starts in April 1999 and ends in May 2022, and its trading period ranges from May 1999 to June 2022.

Regarding the hyper-parameters tuning process, the standard procedure for time-series data is that tuning should be done on the validation dataset between the training and testing set for each rolling or sliding window and must not shuffle the data. However, due to the extreme computational complexity (a totally of 2,274 runs of hyper-parameters tuning needed) in this work, all models' hyper-parameters are only grid searched in the final 2 months of non-shuffled data of their first training dataset. Then, the optimal hyper-parameter setup found for each model is used for all remaining windows. The optimal choice of the hyper-parameters for each model is described in Appendix 7.3.

3.3 Performance Metrics

In the original paper, Bartram and Grinblatt (2018 & 2021) apply (1) the average monthly return and (2) the fraction of months with positive returns as the primary metrics to examine the effectiveness of the mispricing signal. Besides, they also compare the strategy's performance across periods to check whether the mispricing signal has weakened over time.

By contrast, in this work, the main purpose is to compare the performance of the ML augmented strategies, find the best model to boost the mispricing recognition's precision and explore the strategy's feasibility in the real world, so the transaction costs and volatility also need considering. Therefore, in addition to the metrics used by Bartram and Grinblatt (2018 & 2021), (3) the annualised standard deviation, (4) the max drawdown, (5) the Sharpe ratio, (6) the average No. of transactions executed per month, and (7) the average market capitalisation of the portfolios are applied as the performance metrics. The 8 ML augmented strategies will be evaluated with the 7 performance metrics and compared with the OLS benchmark. The notations used for the 7 performance metrics are outlined in Appendix 7.2.

4 Results

Overall, 6 steps were undertaken to obtain the empirical results.

4.1 Models Trained To Recognise Mispricing

The OLS and all 8 ML strategies are provided or trained with the data to calculate the peer-implied fair equity value and the corresponding mispricing signals. For the 8 ML models, they have been applied with the optimal hyper-parameters that are tuned using the final 2 months' non-shuffled data of their first training dataset. Their hyper-parameters remain the same for all the training and testing windows over the sample period.

4.2 Performance Comparisons Across Strategies

Based on the mispricing signals generated by each strategy, the stocks are sorted into 5-quintiles and 10-deciles corresponding to each strategy. The investments are made in the extreme two groups the next month. The long/short portfolios are denoted as Q5 (the most underpriced group) / Q1 (the most overpriced group) for the quintile classified portfolios and D10/D1 for the decile classified portfolios. In each group, the stock holding is equally weighted. The performance statistics for Q5/Q1 and D10/D1 are provided in Appendix 7.4.

Our OLS benchmark has very similar properties and performance to the OLS model created by Bartram and Grinblatt (2018 & 2021). The average monthly spread of our OLS benchmark is 0.41%, with 58.9% of months having positive returns. It can be a reasonable benchmark for ML augmented strategies.

Overall, all ML augmented strategies raise the accuracy of mispricing recognition compared to the OLS benchmark and thus bring higher spreads between the underpriced and the overpriced, a higher percentage of months with positive returns, lower volatility, and a higher Sharpe ratio. The findings are similar to the results derived from Matthias, Marina and Marc (2021), who apply RF and XGB models to detect the mispricing in the Eurozone and train the models with a 4-year term of the fixed rolling training window.

No matter how precise the ML models' mispricing recognition can be, the short position in the most overpriced group still generates negative returns like the OLS benchmark does, consistent with the results from Bartram and Grinblatt (2018 & 2021) in the US stock market. Compared to the OLS benchmark, the ML models can further reduce the returns of the overpriced group, raising the spread returns.

It is worth mentioning that among all the ML strategies, the RF strategy achieves almost the best performance on all metrics and even the lowest holding turnover ratio. Its transaction costs are only 51% and 63% of the OLS benchmark implemented with 5-quintile and 10-decile strategies, respectively. The NN5 model fits the mispricing recognition approach the best in all NN models, and its performance is exceptionally close to that of the RF strategy. However, the NN5 strategy, on average, executes 17% or 28% more transactions per month than the RF model, indicating that the NN5 strategy would have higher transaction costs.

Besides, compared to the OLS benchmark, all ML strategies are more prone to recognise firms with a smaller market capitalisation as underpriced stocks and firms with a larger market capitalisation as overpriced stocks. As mentioned, firms with a smaller market capitalisation may possess the property of lower liquidity, which may limit the profitability of the strategy's long position, especially the strategy implemented with the 10-decile classification. Even if the 10-decile classification brings an even higher spread and lower transaction costs to the strategy, it also focuses more on the firms with a smaller market capitalisation. The trade-off between the group size and the trading feasibility can be further discussed.

Although the LSTM strategy performs a bit better than the OLS benchmark, it still performs worse than other ML strategies. The fair equity value it predicts is relatively closer to each firm's contemporaneous market capitalisation, primarily because the LSTM model not only learns from the feature variables but the market changing pattern of the 12-month period. Its property of learning the time series evolution shrinks the difference between its predicting fair equity value and market capitalisation, leading to worse performance than other ML strategies.

4.3 Performance Comparisons Across Periods

To help inspect whether the effectiveness of the mispricing recognition factor has declined in recent years, the whole sample period is split into two periods: the first period ranges from April 1998 to December 2012, and the second period is from January 2013 to June 2022. The performance statistics for each strategy implemented with the 5-quintile portfolio in the 2 periods are calculated respectively and listed in Appendix 7.5. Each strategy's Q5-Q1 spread lessens by around 50% in unison after switching from the 1998-2012 to 2013-2022 period, primarily because the long position's returns reduce by 20% to 30%.

On the other hand, it is worth noticing that the OLS benchmark's Q1 return decreases after switching from the 1998-2012 to 2013-2022 period. In contrast, all ML models' Q1 return increases, decreasing their Q5-Q1 spread. The difference may result from the fact that the ML models' optimal hyper-parameters are grid searched with the samples in 1998 and remain the same for all windows. Since the market regime may change a lot after such a long time, these hyper-parameters may not be optimal anymore. Thus, if hyper-parameters are tuned using the samples within the 2013-2022 period, ML models' Q1 return would decrease a bit, and their Q5-Q1 spread returns would be higher than the returns shown.

In summary, the mispricing recognition factor still works so far, and its effectiveness has weakened in recent years compared to the performance from 1998 to 2012. However, the degree of decline in its effectiveness may be smaller than shown in the tables.

4.4 Feature Importance

Model's SHAP (SHapley Additive exPlanations, Lundberg et al. 2017) values or impurity-based importance (Breiman et al., 1984) are computed to help understand which feature variables matter to models in fair equity value estimate. Figures in Appendix 7.6 illustrate the average feature importance for each model over the whole sample period.

The OLS benchmark, NN2, NN3 and NN5 models primarily rely on only 1 to 3 feature variables. Instead, the RF, XGB, NN4, and NN6 models draw information from 6 to 10 feature variables. Additionally, the LSTM model seems to extract information from all 21 feature variables. The ML models, except for the LSTM, are somehow consistent in selecting feature variables as the source of information. Cash & Short-Term Investments, Common/Ordinary Equity, Non-operating Income, Income Taxes, and Property Plant & Equipment are the 5 feature variables that most ML models use to estimate the fair equity value.

4.5 Explaining Returns With 6-Factor Time Series Regressions

The time-series regressions of 6 known market anomalies, i.e. Mkt_RF, SMB, HML, Mom, ST_Rev and LT_Rev (described in Appendix 7.1) on 4 representative strategies are performed to examine whether these factors can explain the strategies. The results are reported in Appendix 7.7.

The Q1, Q5, and Q5-Q1 portfolios of the OLS Benchmark, RF, NN5, and LSTM strategies generate significant positive alpha, meaning that the mispricing signal indeed captures some positive returns that the known anomalies cannot explain. Besides, most of their returns come from SMB and HML factors, consistent with our expectations and the statistics of the average market capitalisation. Interestingly, compared to most of the strategies that lightly but statistical-significantly rely on the ST_Rev factor, the RF strategy has an insignificant relationship with the factor, and the RF possess the highest alpha. Considering the length of the holding window is only 1 month, it is pretty reasonable that the relationships between most strategies and LT_Rev are statistically insignificant.

Furthermore, it is worth mentioning that the Q5-Q1 returns for all strategies have a significant negative relationship with the market risk factor, which is an excellent selling point for institutional clients that especially favour counter-cyclical investments, such as insurance companies and pension funds.

4.6 Correlations Between 9 Strategies

Finally, a correlation matrix between strategies and graphs for cumulative returns of all strategies are provided in Appendix 7.8 to help understand the interactions between strategies. Due to the difference in mispricing signal recognition where OLS only capture linear relationship, but ML models consider the nonlinearity, the correlations between the OLS benchmark and each ML strategy are significantly lower than those between the ML strategies.

5 Conclusion

Overall, the mispricing recognition based on the valuation approach by Bartram and Grinblatt (2018 & 2021) still works in the US stock market so far but has gradually weakened, and the 8 machine learning augmented strategies perform better than the OLS benchmark on all metrics. Among them, the random forest augmented strategy brings us the highest Sharpe ratio and the lowest transaction costs.

Although the returns of the strategy can be partially explained by the SMB, HML and ST_Rev factors, the strategies still generate statistical-significantly positive alpha. The long/short version of the strategy has a statistical-significantly negative beta, an appealing aspect to institutional clients.

6 References

Bartram, S. M., Grinblatt, M., 2018. Agnostic fundamental analysis works. Journal of Financial Economics, Vol. 128 (1), April 2018, 125–147.

Bartram, S. M., Grinblatt, M., 2021. Global market inefficiencies. Journal of Financial Economics 139, 234-259.

Matthias X. H., Marina K. & Marc S. R., 2021. Finance Research Letters, forthcoming

Breiman, L., 2001. Random forests. Machine Learning 45, 5-32.

Chen, T., Guestrin, C., 2016. XGBoost: a scalable tree boosting system. KDD 2016: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.

Hopfield, J., 1982. Neural networks and physical systems with emergent collective computational abilities. Proceedings of the National Academy of Sciences, 79(8), pp.2554-2558.

Hochreiter, S. and Schmidhuber, J., 1997. Long Short-Term Memory. Neural Computation, 9(8), pp.1735-1780.

Gu, S., Kelly, B., Xiu, D., 2020. Empirical asset pricing via machine learning. The Review of Financial Studies 33, 2223-2273.

Lundberg, S., Lee, S., 2017. A unified approach to interpreting model predictions. Advances in Neural Information Processing Systems 47, 68–77.

Breiman, L., Friedman, J. H., Olshen, R. A., Stone, C. J., 1984. Classification and regression trees. Wadsworth and Brooks/Cole, Monterey.

7 Appendix

7.1 Datasets, Definitions & Source

Variable	Definition	Source
ACOQ	Current Assets - Other - Total - Quarterly	S&P Compustat Fundamentals Dataset
AOQ	Assets - Other - Total - Quarterly	
ATQ	Assets - Total - Quarterly	
CEQQ	Common/Ordinary Equity - Total - Quarterly	
CHEQ	Cash and Short-Term Investments - Quarterly	
DLTTQ	Long-Term Debt - Total - Quarterly	
DVPQ	Dividends - Preferred/Preference - Quarterly	
IBADJQ	Income Before Extraordinary Items - Adjusted for Common Stock Equivalents - Quarterly	
IBQ	Before Extraordinary Items - Quarterly	
LCOQ	Current Liabilities - Other - Total - Quarterly	
LOQ	Liabilities - Other - Total - Quarterly	
LTQ	Liabilities - Total - Quarterly	
NIQ	Net Income (Loss) - Quarterly	
NOPIQ	Non-operating Income (Expense) - Quarterly	
PIQ	Pretax Income - Quarterly	
PPENTQ	Property Plant and Equipment - Total (Net) - Quarterly	
PSTKQ	Preferred/Preference Stock (Capital) - Total - Quarterly	
SALEQ	Sales/Turnover (Net) - Quarterly	
TXTQ	Income Taxes - Total - Quarterly	
XIDOQ	Extraordinary Items and Discontinued Operations - Quarterly	
DVY	Cash Dividends (Cash Flow) - Yearly	
PRCCM	Price - Close - Monthly	
TRT1M	Total Return - Monthly	
CSHOM	Shares Outstanding – Issue - Monthly	
Fed Rate	Federal Funds Effective Rate - Monthly	Fred Economic Data
Mkt_RF	Monthly market index excess return	Ken French Data Library
SMB	Monthly Small Minus Big (SMB) portfolio return	
HML	Monthly High Minus Low (HML) portfolio return	
Mom	Monthly Momentum portfolio return	
ST_Rev	Monthly Short-term reversal portfolio return	
LT_Rev	Monthly long-term reversal portfolio return	

Table 1: Variable Definitions & Sources

7.2 Notation, Model Outline & Performance Metrics

Strategy	Notation
OLS Regression Applied Strategy	OLS
Random Forest Regression Model Augmented Strategy	RF
XGBoost Regression Model Augmented Strategy	XGB
2 Hidden Layers Neural Networks Augmented Strategy	NN2
3 Hidden Layers Neural Networks Augmented Strategy	NN3
4 Hidden Layers Neural Networks Augmented Strategy	NN4
5 Hidden Layers Neural Networks Augmented Strategy	NN5
6 Hidden Layers Neural Networks Augmented Strategy	NN6
1 Hidden Layer LSTM Model Augmented Strategy	LSTM

Table 2: Notation for Strategies

Model	Fixed Rolling Training Window	Fixed Testing Window
RF	6 months	1 month
XGB	6 months	1 month
NN2	6 months	1 month
NN3	6 months	1 month
NN4	6 months	1 month
NN5	6 months	1 month
NN6	6 months	1 month
Model	**Fixed Sliding Training Window**	**Fixed Testing Window**
LSTM	12 months	12 months

Table 3: Machine Learning Models Outline

Performance Metrics	Notation
Average Monthly Return	AMR
Percentage of Months with Positive Returns	M>0
Annualised Standard Deviation (Volatility)	Vol
Max Drawdown	MDD
Sharpe Ratio	SR
Average No. of Transactions Executed per Month	T/M
Average Market Capitalisation	Mkt Cap

Table 4: Notation for Performance Metrics

7.3 The Optimal Selection For Hyper-parameters

All hyper-parameters listed below are tuned with a grid search algorithm. The figures in the brackets are the hyper-parameter spaces that we searched through. The figures outside the brackets are the optimal hyper-parameters.

Params	RF Regression Model	XGB Regression Model
n_estimators	350 [100, 200, 300, 350, 400]	50 [50, 100, 150, 200, 250, 300]
max_depth	2 [2, 3, 4, 5, 6, 7, 8]	5 [4, 5, 6, 7, 8, 9]
max_features	4 [2, 3, 4, 5, 6, 7, 8]	-
max_leaf_nodes	4 [2, 3, 4, 5]	-
min_samples_leaf	1 [1, 2, 3, 4, 5]	-
min_samples_split	6 [3, 4, 5, 6, 7, 8]	-
eta	-	0.02 [0.001, 0.01, 0.02, 0.03]
alpha	-	0 [0, 0.1, 0.2, 0.3, 0.4]
gamma	-	0 [0, 0.1, 0.2, 0.3, 0.4]
reg_lambda	-	0.2 [0.1, 0.2, 0.3, 0.4, 0.5]
subsample	-	0.5 [0.2, 0.3, 0.4, 0.5, 0.6]
colsample_bytree	-	0.6 [0.4, 0.5, 0.6, 0.7, 0.8]

Table 5: Hyper-parameters Selected for RF & XGBoost Models

For the number of nodes in each layer, the searching space ranges from 1 to 30. The activation function is searched from [relu, tanh, sigmoid]. The dropout rate ranges from 0.1 to 0.7. Learning rate ranges [0.0001, 0.001, 0.01]. L2 Regularisation is only applied to kernel and bias ranges [0.001, 0.01, 0.1]. The number of epochs ranges from 1 to 50.

Params	NN2 Model	NN3 Model	NN4 Model	NN5 Model	NN6 Model
Activation Function	ReLu	ReLu	ReLu	ReLu	ReLu
Learning Rate	0.01	0.01	0.01	0.01	0.001
L2 Regularisation	0.001	0.1	0.01	0.001	0.01
Epochs	50	50	2	6	6
Input Layer	15	20	11	4	12
Dropout Rate	0.1	0.2	0.3	0.4	0
Hidden Layer 1	4	13	44	13	16
Dropout Rate 1	0.2	0.2	0.3	0.1	0.6
Hidden Layer 2	18	11	15	19	11
Dropout Rate 2	0.3	0.1	0.4	0.2	0.6
Hidden Layer 3	-	17	17	2	11
Dropout Rate 3	-	0	0.4	0.5	0.1
Hidden Layer 4	-	-	16	16	18
Dropout Rate 4	-	-	0.1	0.6	0.6
Hidden Layer 5	-	-	-	9	3
Dropout Rate 5	-	-	-	0.1	0
Hidden Layer 6	-	-	-	-	18
Dropout Rate 6	-	-	-	-	0.6

Table 6: Hyper-parameters Selected for Neural Network Models

For the number of nodes in each layer, the searching space ranges from 1 to 30. The dropout rate ranges from 0.1 to 0.7. Learning rate ranges [0.0001, 0.001, 0.01]. L2 Regularisation is only applied to kernel and bias ranges [0.001, 0.01, 0.1]. The number of epochs ranges from 1 to 50.

Params	**LSTM Model**
Learning Rate	0.01
L2 Regularisation	0.01
Epochs	50
Input Layer (LSTM Layer)	34
Dropout Rate	0.4
Hidden Layer 1	27
Dropout Rate 1	0.3

Table 7: Hyper-parameters Selected for LSTM Models

7.4 Performance Statistics (For The Whole Sample Period)

Performance stats on the period of ‘98 to ‘22 for all strategies applied with the 5-quintile sorted portfolio are listed below:

	AMR	M>0	Vol	MDD	SR	T/M	Mkt Cap (bn)
OLS							
Q5	1.92%	67.9%	16.97%	11.8%	1.40	209	4.0
Q1	1.51%	65.5%	19.24%	24.8%	0.70	207	43.0
Q5 - Q1	0.41%	58.9%	6.41%	22.7%	0.33	416	24.1
RF							
Q5	2.52%	70.7%	16.43%	12.4%	2.00	138	0.6
Q1	0.96%	61.7%	16.50%	21.9%	0.44	76	70.3
Q5 - Q1	1.56%	73.1%	4.70%	2.1%	1.97	213	36.9
XGB							
Q5	2.45%	71.0%	16.32%	10.1%	1.96	187	1.8
Q1	1.05%	61.7%	17.46%	28.9%	0.46	142	58.5
Q5 - Q1	1.40%	73.8%	4.82%	3.0%	1.72	329	31.2
NN2							
Q5	2.49%	70.7%	17.24%	12.8%	1.86	227	1.7
Q1	0.99%	62.1%	16.80%	24.6%	0.51	243	61.1
Q5 - Q1	1.50%	69.0%	5.11%	4.3%	1.74	470	32.5
NN3							
Q5	2.55%	70.3%	16.86%	10.5%	1.99	188	1.1
Q1	1.01%	62.1%	16.91%	26.0%	0.43	225	62.0
Q5 - Q1	1.54%	71.4%	5.00%	2.7%	1.82	413	32.8
NN4							
Q5	2.55%	71.0%	16.87%	12.1%	1.98	148	0.7
Q1	1.01%	62.8%	17.10%	25.0%	0.44	162	67.6
Q5 - Q1	1.54%	71.0%	4.96%	2.9%	1.85	310	35.5
NN5							
Q5	2.53%	69.7%	16.83%	12.1%	1.96	140	0.6
Q1	0.97%	62.1%	16.36%	22.2%	0.47	131	73.8
Q5 - Q1	1.56%	71.4%	4.79%	2.3%	1.93	271	38.7
NN6							
Q5	2.55%	69.7%	16.82%	10.8%	1.96	156	0.9
Q1	1.05%	61.0%	17.51%	28.2%	0.46	136	65.3
Q5 - Q1	1.50%	71.4%	4.87%	2.6%	1.83	291	34.4
LSTM							
Q5	1.56%	63.8%	17.26%	23.6%	1.07	148	1.9
Q1	0.97%	63.8%	17.06%	25.9%	0.42	209	50.8
Q5 - Q1	0.59%	56.9%	5.24%	9.3%	0.59	356	26.6

Table 8: Performance Statistics for Quintile Sorted Portfolios (from 1998 to 2022)

Performance stats on the period of ‘98 to ‘22 for all strategies applied with the 10-decile sorted portfolio are listed below:

	AMR	M>0	Vol	MDD	SR	T/M	Mkt Cap (bn)
OLS							
D10	2.23%	70.3%	17.18%	12.9%	1.66	134	3.2
D1	1.62%	64.5%	19.22%	19.1%	0.78	128	51.2
D10 - D1	0.61%	62.1%	6.53%	23.6%	0.51	262	28.5
RF							
D10	3.22%	74.8%	15.08%	6.8%	2.94	109	0.5
D1	0.98%	62.1%	17.05%	25.8%	0.43	54	116.1
D10 - D1	2.24%	76.2%	5.54%	2.4%	2.48	164	62.1
XGB							
D10	2.95%	72.1%	15.25%	6.6%	2.61	134	1.8
D1	1.19%	62.4%	19.04%	32.1%	0.49	90	72.9
D10 - D1	1.76%	74.5%	6.07%	3.8%	1.76	223	39.2
NN2							
D10	3.10%	74.1%	16.12%	8.1%	2.63	160	1.6
D1	0.98%	64.1%	17.08%	27.3%	0.48	167	91.1
D10 - D1	2.12%	75.5%	5.73%	2.4%	2.25	326	49.0
NN3							
D10	3.25%	75.2%	15.66%	6.2%	2.77	138	1.1
D1	1.03%	60.7%	17.74%	30.6%	0.37	158	91.9
D10 - D1	2.22%	76.2%	5.85%	2.9%	2.34	296	49.4
NN4							
D10	3.24%	74.8%	15.40%	6.9%	2.90	114	0.6
D1	1.03%	60.7%	18.01%	27.1%	0.42	111	112.5
D10 - D1	2.21%	74.5%	6.15%	3.4%	2.21	225	60.2
NN5							
D10	3.26%	75.9%	15.48%	6.8%	2.89	111	0.5
D1	0.95%	61.4%	16.81%	25.6%	0.43	100	125.7
D10 - D1	2.31%	77.2%	5.73%	2.5%	2.48	211	67.3
NN6							
D10	3.24%	74.1%	15.41%	6.2%	2.92	118	0.7
D1	1.04%	61.0%	19.01%	32.7%	0.39	86	108.3
D10 - D1	2.20%	74.1%	6.28%	2.7%	2.16	204	58.1
LSTM							
D10	1.91%	65.2%	17.01%	11.5%	1.39	128	1.5
D1	1.00%	63.8%	18.70%	29.1%	0.37	94	79.4
D10 - D1	0.91%	60.0%	6.88%	9.9%	0.75	220	41.1

Table 9: Performance Statistics for Decile Sorted Portfolios (from 1998 to 2022)

7.5 Performance Statistics (1998 - 2012 vs 2013 - 2022)

Performance stats on the period of ‘98 to ‘12 for all strategies applied with the 5-quintile sorted portfolio are listed below:

’98 – ‘12	AMR	M>0	Vol	MDD	SR	T/M	Mkt Cap (bn)
OLS							
Q5	2.08%	65.9%	17.39%	19.1%	1.50	208	3.9
Q1	1.57%	61.4%	21.11%	37.4%	0.64	205	38.8
Q5 - Q1	0.51%	60.2%	7.59%	13.8%	0.38	413	22.0
RF							
Q5	2.84%	68.8%	16.49%	17.8%	2.31	145	0.6
Q1	0.90%	59.1%	17.92%	46.7%	0.33	74	63.9
Q5 - Q1	1.94%	77.8%	4.80%	2.7%	2.48	218	33.7
XGB							
Q5	2.71%	68.8%	16.47%	17.8%	2.18	201	1.7
Q1	1.01%	58.5%	18.91%	49.1%	0.37	143	55.4
Q5 - Q1	1.70%	77.3%	5.00%	3.0%	2.08	344	29.7
NN2							
Q5	2.81%	67.6%	18.00%	18.7%	2.08	218	1.4
Q1	0.99%	61.4%	17.80%	47.9%	0.40	224	56.0
Q5 - Q1	1.82%	73.3%	5.34%	1.6%	2.09	442	29.9
NN3							
Q5	2.89%	68.2%	17.35%	17.3%	2.24	200	1.1
Q1	0.90%	57.4%	18.13%	47.4%	0.33	208	58.6
Q5 - Q1	1.99%	77.8%	5.02%	2.9%	2.45	408	31.1
NN4							
Q5	2.89%	68.2%	17.17%	17.6%	2.26	150	0.6
Q1	0.93%	58.5%	18.57%	47.4%	0.33	149	62.5
Q5 - Q1	1.96%	76.1%	5.35%	3.4%	2.25	299	32.9
NN5							
Q5	2.84%	67.6%	17.14%	17.2%	2.22	142	0.5
Q1	0.92%	59.7%	17.66%	49.3%	0.36	113	66.4
Q5 - Q1	1.92%	76.1%	5.01%	2.7%	2.35	255	35.0
NN6							
Q5	2.85%	67.0%	17.15%	17.0%	2.23	157	0.8
Q1	0.99%	57.4%	18.96%	47.5%	0.36	125	60.3
Q5 - Q1	1.86%	75.6%	5.10%	3.4%	2.25	282	31.9
LSTM							
Q5	1.56%	57.1%	16.82%	19.5%	1.11	135	1.6
Q1	0.82%	62.1%	17.95%	46.5%	0.28	179	37.1
Q5 - Q1	0.74%	54.9%	5.50%	18.3%	0.74	306	19.6

Table 10: Performance Statistics for Quintile Sorted Portfolios (from 1998 to 2012)

Performance stats on the period of ‘13 to ‘22 for all strategies applied with the 5-quintile sorted portfolio are listed below:

’13 – ‘22	AMR	M>0	Vol	MDD	SR	T/M	Mkt Cap (bn)
OLS							
Q5	1.66%	71.1%	16.34%	11.8%	1.23	211	4.2
Q1	1.41%	71.9%	16.00%	24.8%	0.84	210	49.5
Q5 - Q1	0.25%	57.0%	3.95%	11.6%	0.23	421	27.4
RF							
Q5	2.03%	73.7%	16.28%	12.4%	1.56	130	0.6
Q1	1.04%	65.8%	14.20%	21.9%	0.64	81	80.0
Q5 - Q1	0.99%	65.8%	4.38%	2.1%	1.22	211	41.7
XGB							
Q5	2.05%	74.6%	16.07%	10.1%	1.60	169	1.8
Q1	1.09%	66.7%	15.14%	28.9%	0.63	144	62.9
Q5 - Q1	0.96%	68.4%	4.46%	3.0%	1.17	313	33.4
NN2							
Q5	1.94%	73.7%	15.96%	12.8%	1.51	234	1.9
Q1	1.18%	66.7%	15.24%	24.6%	0.70	289	65.8
Q5 - Q1	0.76%	64.0%	4.13%	3.2%	0.95	523	35.0
NN3							
Q5	2.04%	72.8%	16.02%	10.5%	1.59	180	1.3
Q1	1.05%	64.0%	14.97%	26.0%	0.60	252	68.5
Q5 - Q1	0.99%	67.5%	4.45%	2.6%	1.20	432	36.1
NN4							
Q5	2.02%	73.7%	16.32%	12.1%	1.55	150	0.9
Q1	1.09%	64.0%	14.70%	25.0%	0.65	194	76.3
Q5 - Q1	0.93%	67.5%	4.37%	2.7%	1.15	344	39.9
NN5							
Q5	2.05%	72.8%	16.28%	12.1%	1.57	136	0.7
Q1	1.07%	66.7%	14.26%	22.2%	0.67	143	84.8
Q5 - Q1	0.98%	65.8%	4.37%	2.6%	1.20	279	44.2
NN6							
Q5	2.02%	72.8%	16.24%	10.8%	1.55	161	1.1
Q1	1.13%	65.8%	15.17%	28.2%	0.66	160	72.6
Q5 - Q1	0.89%	64.0%	4.48%	2.8%	1.06	321	38.1
LSTM							
Q5	1.39%	66.7%	17.91%	23.6%	0.90	171	2.2
Q1	1.08%	66.7%	15.45%	25.9%	0.60	259	72.3
Q5 - Q1	0.31%	53.5%	4.73%	9.3%	0.27	430	37.6

Table 11: Performance Statistics for Quintile Sorted Portfolios (from 2013 to 2022)

7.6 Features Importance

The figures illustrate the feature importance of the 8 models. The feature importance is based on the average SHAP values or impurity-based importance computed over the sample period.

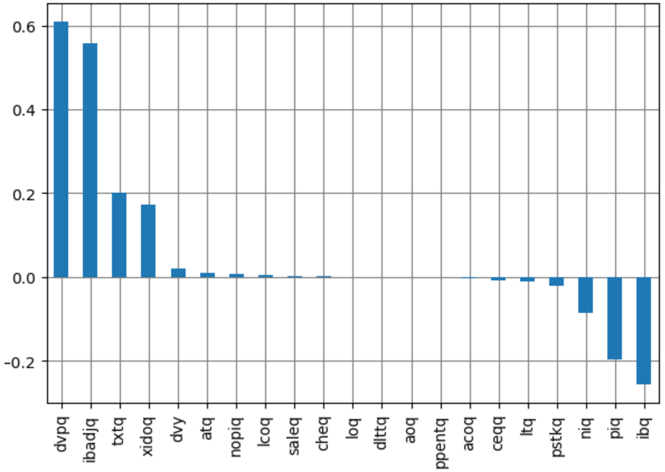

Figure 3: OLS Regression Model

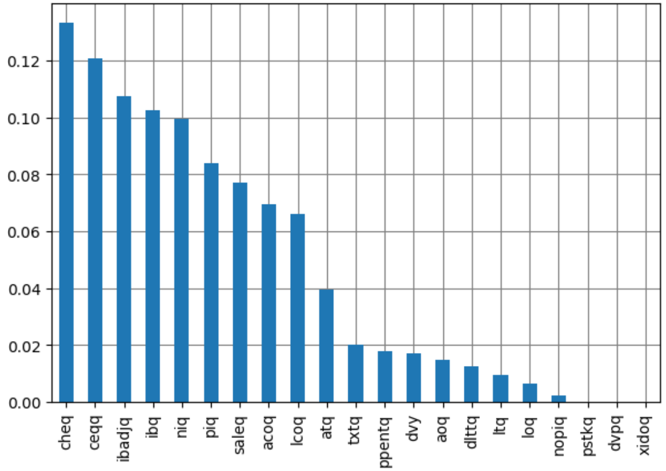

Figure 4: Random Forest Regression Model

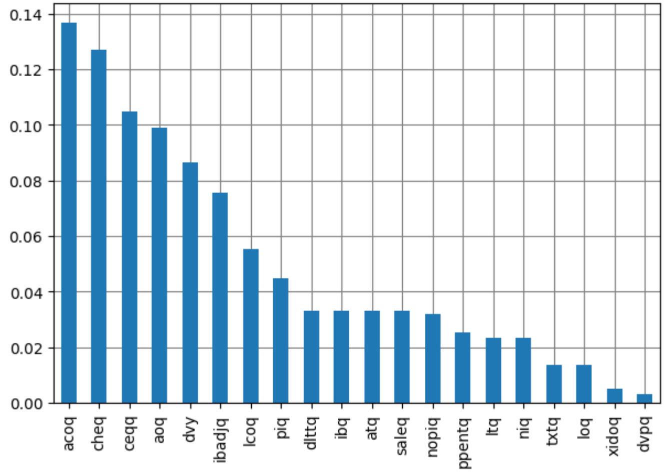

Figure 5: XGBoost Regression Model

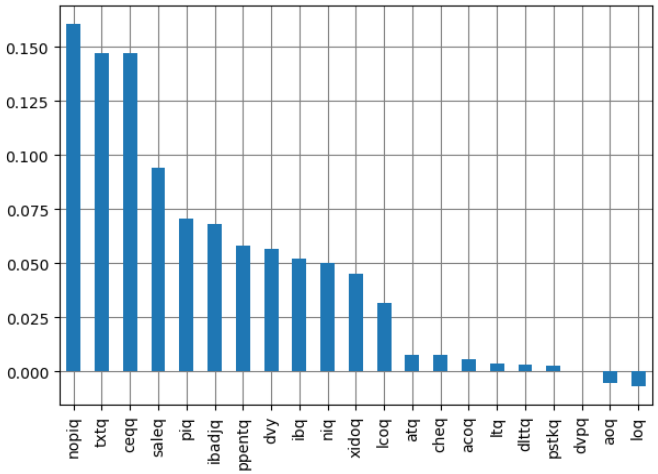

Figure 6: NN2 Model

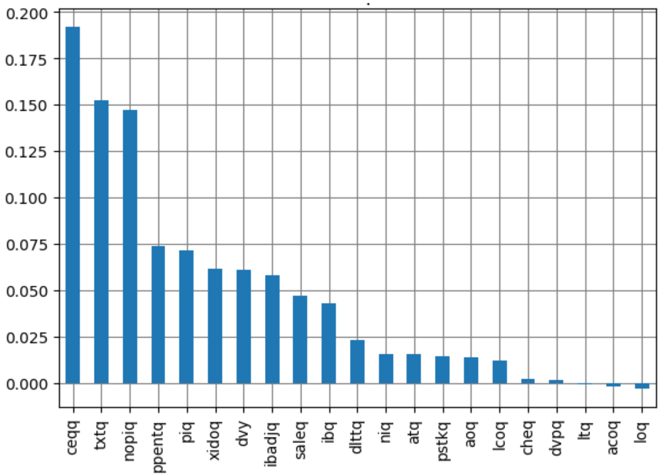

Figure 7: NN3 Model

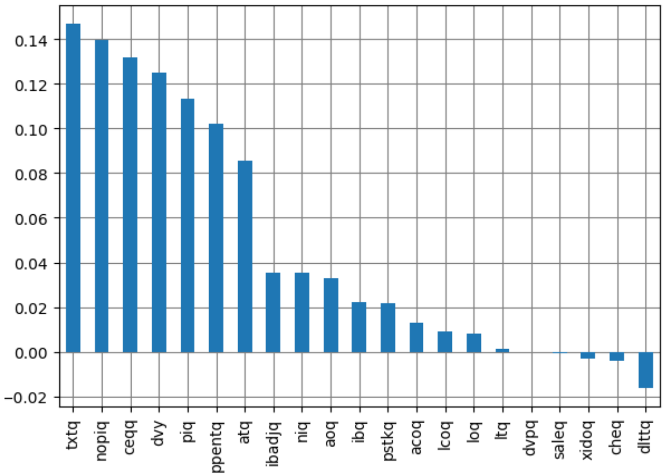

Figure 8: NN4 Model

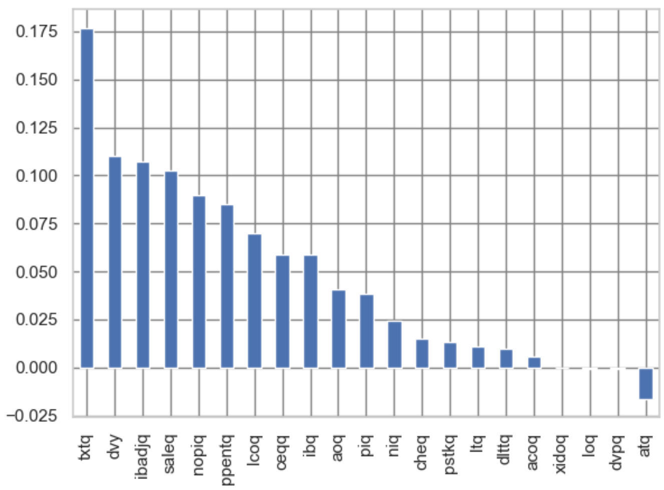

Figure 9: NN5 Model

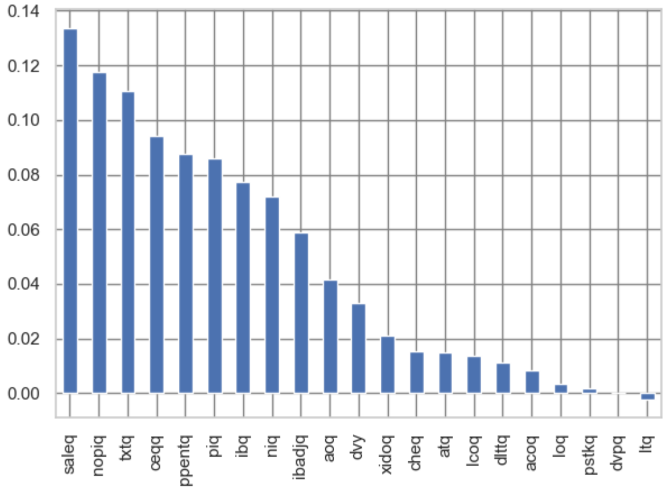

Figure 10: NN6 Model

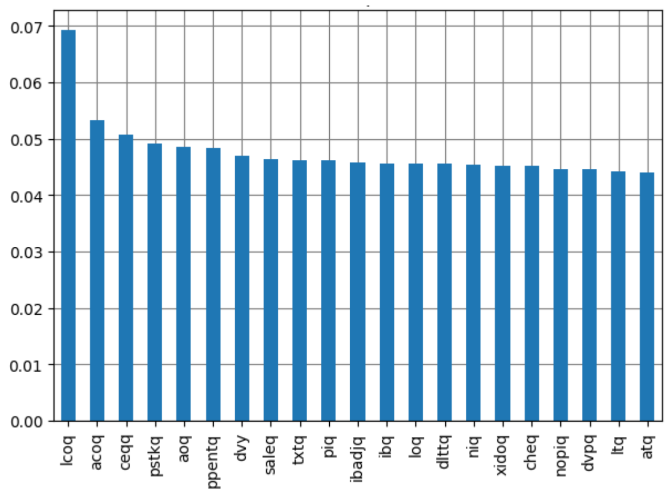

Figure 11: LSTM Model

7.7 6 Factor Time-Series Regressions

The table describes the intercept, slope coefficients, t-statistics, R^2, and the number of observations from time-series regression of the Q1, Q5 and Q5-Q1 positions' monthly return on 6 factors for OLS, RF, NN5, and LSTM strategy separately. The numbers in parentheses are t-statistics. *, **, *** represent the statistical significance at the 10%, 5%,and 1% levels, respectively.

	OLS			RF			NN5			LSTM		
	Q1	**Q5**	**Q5-Q1**	**Q1**	**Q5**	**Q5-Q1**	**Q1**	**Q5**	**Q5-Q1**	**Q1**	**Q5**	**Q5-Q1**
Alpha	0.0074*** (6.703)	0.0151*** (12.622)	0.0039*** (4.089)	0.0024*** (3.546)	0.0203*** (20.500)	0.0090*** (15.660)	0.0030*** (4.474)	0.0203*** (20.398)	0.0087*** (15.053)	0.0028*** (3.699)	0.0113*** (10.302)	0.0043*** (6.759)
Mkt_RF	0.9699*** (35.796)	0.7824*** (26.515)	-0.0941*** (-4.033)	0.9862*** (59.445)	0.7224*** (29.111)	-0.1322*** (-9.203)	0.9957*** (59.481)	0.7218*** (28.948)	-0.1372*** (-9.526)	1.0197*** (54.550)	0.7703*** (27.878)	-0.1248*** (-7.846)
SMB	0.4914*** (12.247)	0.4416*** (10.106)	-0.0251 (-0.726)	0.1789*** (7.317)	0.6781*** (18.545)	0.2495*** (11.792)	0.1808*** (7.332)	0.7251*** (19.735)	0.2720*** (12.816)	0.1918*** (6.943)	0.5188*** (12.705)	0.1632*** (6.943)
HML	-0.1461*** (-3.478)	0.3112*** (6.800)	0.2284*** (6.316)	-0.0842*** (-3.330)	0.2002*** (5.293)	0.1421*** (6.489)	-0.0582** (-2.281)	0.1673*** (4.402)	0.1126*** (5.128)	-0.1259*** (-4.427)	0.3301*** (7.851)	0.2280*** (9.419)
Mom	0.0136 (0.574)	-0.0604** (-2.340)	-0.0370* (-1.815)	0.0197 (1.387)	-0.0270 (-1.270)	-0.0234* (-1.899)	0.0156 (1.088)	-0.0095 (-0.444)	-0.0125 (-1.016)	0.0474*** (2.964)	-0.1389*** (-5.875)	-0.0929*** (-6.823)
ST_Rev	-0.0581* (-1.923)	0.0882*** (2.682)	0.0732*** (2.814)	-0.0233 (-1.285)	0.0218 (0.801)	0.0226 (1.435)	-0.0569** (-3.103)	0.0344 (1.262)	0.0457** (2.896)	-0.0315 (-1.554)	0.1010*** (3.374)	0.0662*** (3.843)
LT_Rev	0.1028** (2.036)	-0.0267 (-0.485)	-0.0646 (-1.487)	-0.0405 (-1.321)	0.0175 (0.382)	0.0291 (1.096)	-0.0477 (-1.541)	0.0421 (0.914)	0.0450 (1.690)	-0.0679* (-1.971)	-0.0812 (-1.596)	-0.0066 (-0.226)
R^2	0.892	0.836	0.280	0.948	0.883	0.521	0.947	0.887	0.544	0.940	0.872	0.539
Obs	290	290	290	284	284	284	284	284	284	278	278	278

Table 12: 6 Factor Time-Series Regressions on OLS, RF, NN5, and LSTM Strategies

7.8 Correlations Between 9 Strategies & Graphs for Cumulative Returns

The table describes the correlation analysis of the Q5-Q1 spread returns of the 9 strategies

	OLS	RF	XGB	NN2	NN3	NN4	NN5	NN6	LSTM
OLS	1	0.3653	0.5443	0.1403	0.4005	0.3966	0.3319	0.4388	0.5377
RF	0.3653	1	0.8580	0.8398	0.9249	0.9600	0.9643	0.9537	0.7639
XGB	0.5443	0.8580	1	0.6314	0.8247	0.8314	0.7738	0.8640	0.8725
NN2	0.1403	0.8398	0.6314	1	0.8567	0.8849	0.8965	0.8651	0.5330
NN3	0.4005	0.9249	0.8247	0.8567	1	0.9262	0.9248	0.9418	0.7065
NN4	0.3966	0.9600	0.8314	0.8849	0.9262	1	0.9675	0.9782	0.7382
NN5	0.3319	0.9643	0.7738	0.8965	0.9248	0.9675	1	0.9496	0.6899
NN6	0.4388	0.9537	0.8640	0.8651	0.9418	0.9782	0.9496	1	0.7701
LSTM	0.5377	0.7639	0.8725	0.5330	0.7065	0.7382	0.6899	0.7701	1

Table 13: Correlation Matrix for All 9 Strategies

The graph describes the cumulative Q5-Q1 spread returns of the 9 strategies

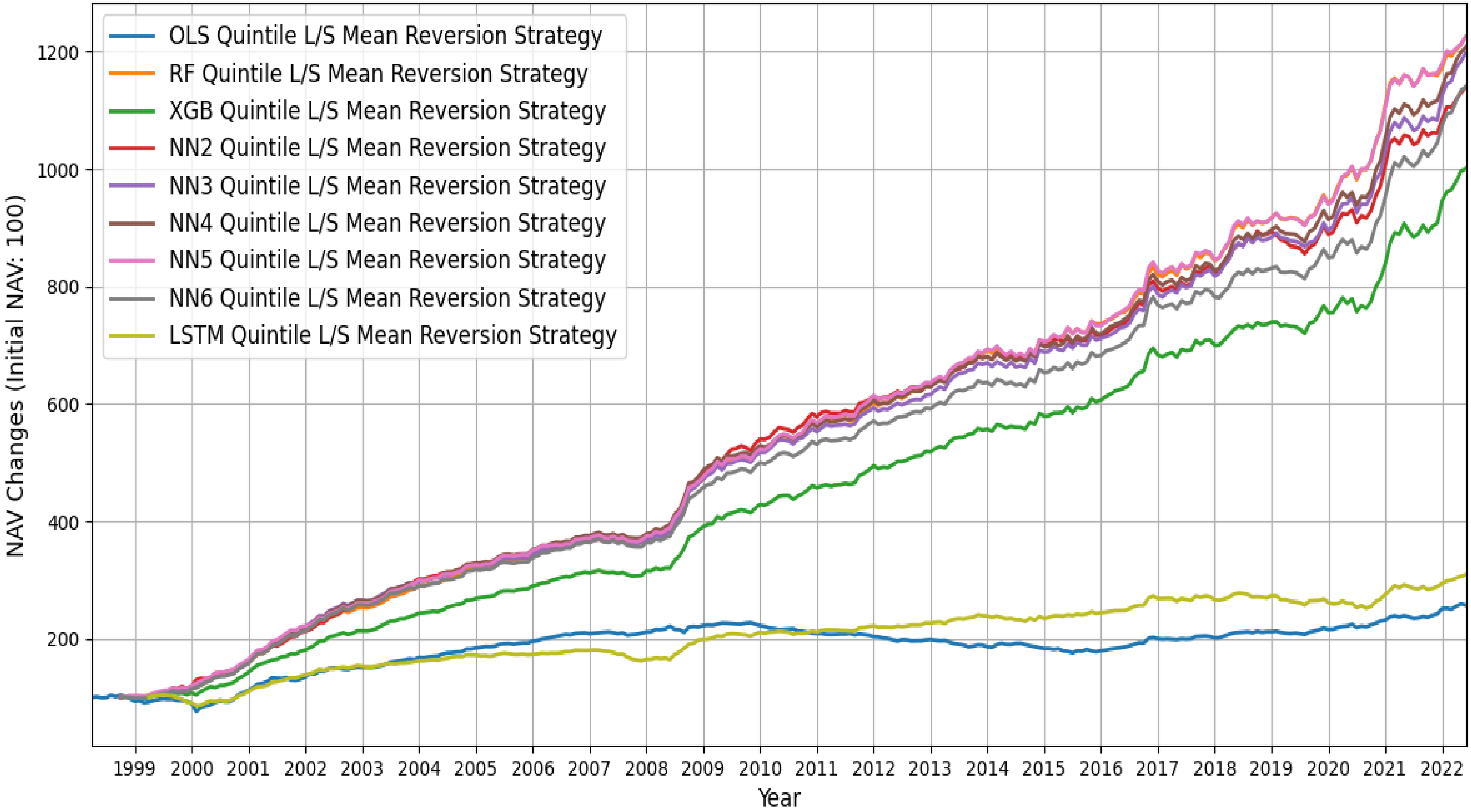

Figure 12: Cumulative Spread Returns on all 9 Quintile Strategies

The graph describes the cumulative D10-D1 spread returns of the 9 strategies

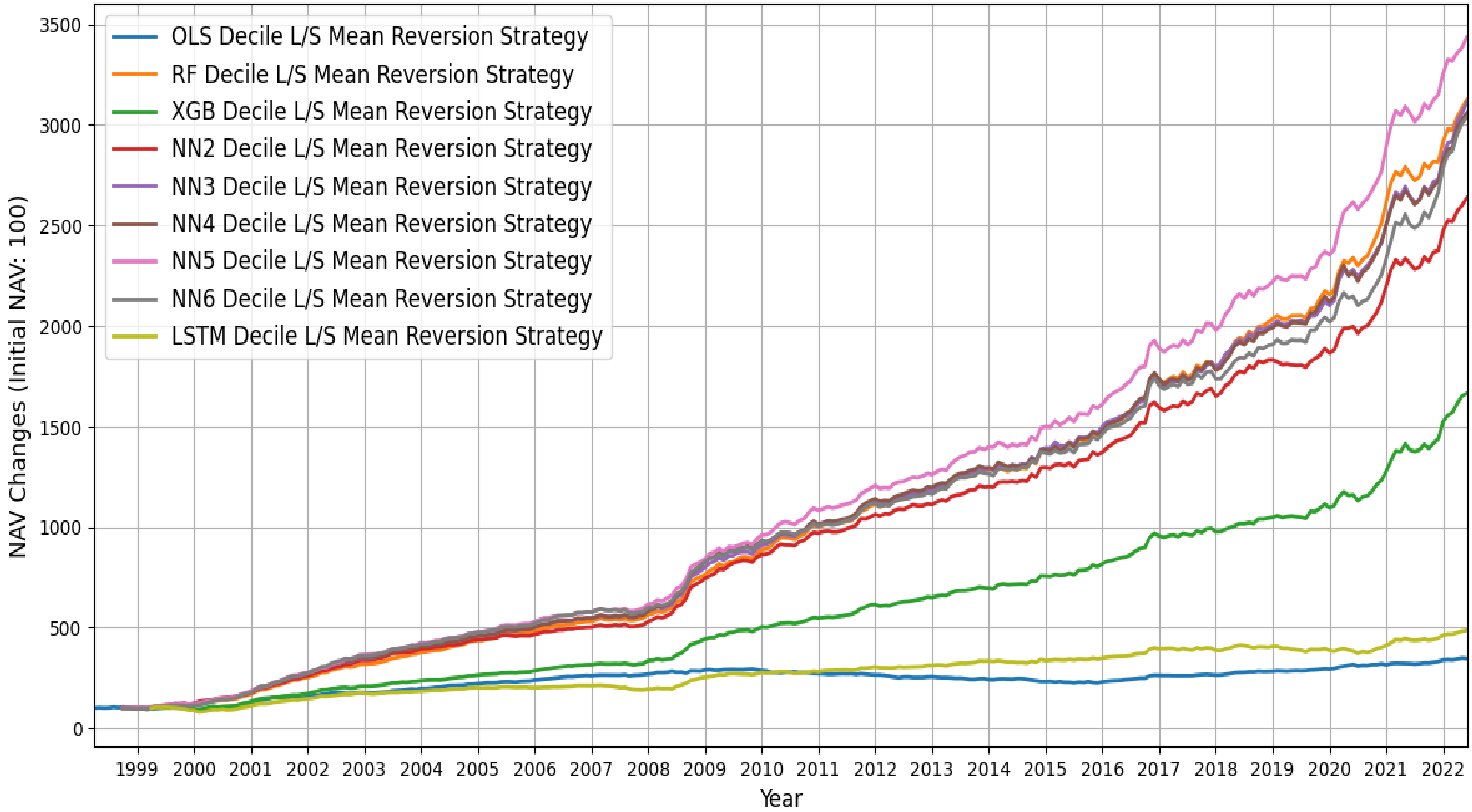

Figure 13: Cumulative Spread Returns on all 9 Decile Strategies