Lai, Ping-fu (Brian) – Wong Chung Hang

Performance of Stock Market Prediction

A study on prediction accuracy and realised return

SUMMARY: Traditional portfolio theory stated that diversified portfolio is optimised regarding returns. It can generate the highest return with relatively lowest risk. Market risk cannot be diversified, so the most intelligent approach is to buy and hold assets in the long run. Therefore, the market return is treated as the benchmark return. Return added in advance to the benchmark return is alpha return. As such, the objective of trading actively is to beat the benchmark, i.e. throughout trading, alpha return is expected. Otherwise, active trading does not make any sense. Active trading is based on prediction, whether by fundamental analysis, by technical analysis or by applying the principles of Chinese feng shui. This study found that with a certain level of accuracy of prediction, it can achieve alpha return. It suggests a model for estimating the level of accuracy. Also, the model is enhanced by including transaction cost. Some implications or corollary can be concluded by this study. Firstly, a model is proposed which can be treated as a baseline for formulating the required accuracy of different computer aided prediction models such as SVN, Neural Network, and GARCH. In other words, to examine if one prediction model works, we can examine the required level of accuracy by the proposed model and then compare the required accuracy and prediction accuracy. Secondly, there are several corollaries on market behaviour the evidence of which supports the finding that most active traders are unable to exploit market opportunities, i.e. to buy and sell better and make more buys in a bull market.

KEYWORDS: stock market, prediction accuracy, realised return, portfolio theory
JEL-CODE: G1

Due to the development of financial engineering, econometrics and artificial intelligence, various stock market prediction methods are proposed and experimented with to predict stock prices. With the rapid growth of digital computing power, stock market prediction, combined with artificial and statistical analytics methods, has moved into the realm of technology. Artificial neural networks, genetic algorithms, and support vector machines are some common techniques used for modelling prediction. This study is mainly inspired by the study conducted by *P.M. Tsang, P. Kwok, S.O. Choy, R. Kwan, S.C. Ng, J. Mak, J. Tsang, K. Koong and T.L. Wong* (2007) which found that the prediction accuracy of Neural Network is relatively high (over 70%) on the Hong Kong Stock Market. However, after including transaction cost, the realised return dropped from profitable to loss-making. This triggered concern regarding the relationship between prediction accuracy, realised return and transaction cost.

Recent study has shown that high accuracy cannot result in high return after translating the prediction to trading signals. So the

Email address: blai8733@gmail.com

first problem is that we have to identify the relationship between prediction accuracy and return made by the prediction. It is expected that high accuracy should result in high return. The second problem is that we have to assess the trading strategy and find one trading strategy which can utilise the accuracy of the prediction method so that this trading strategy can give increasing return with increasing prediction accuracy. The third problem is the impact on the return made after taking account of transaction cost. Obviously, transaction cost can lower the return realised. However, we have to Chart out what transaction cost level is acceptable with regard to the prediction so that it can still generate return.

LITERATURE REVIEWS

Fama (1965) proposed the famous Efficient Market Hypothesis (EMH) that stock price movements are random and unpredictable if the market is fully efficient. Though there are some empirical studies supporting EMH, many market participants and professionals are still skeptical about whether EMH is valid in the stock market. The basic assumption of those skeptical people is that the market is not that efficient and that there is still room to make reasonable predictions of market prices. As such, three streams of prediction methods are devised:

- fundamental analysis,
- technical analysis
- and technological methods.

Fundamental Analysis (Ritchie, 1996) is mainly based on the assumption that the stock price should be reflected by its return on investment and intrinsic value. Fundamental analysts usually study the business model and financial statements to analyse the value and expected return of one corporation. Though most of time the projected value of the corporation is different from the market value on the stock exchange, most cases found that the stock price can reach the intrinsic value in the long run.

Technical analysis (Murphy, 1999) represents various techniques used to make prediction by observing the past stock price pattern. The assumption of technical analysis is that price movement follows a particular pattern so that future price can be determined by observing similar price movement patterns in the past. However, many techniques are subjective, have been proved statistically invalid and lack a rational explanation for their usage (Coulson, 1987). This creates difficulty for experimenting and determining the performance of various technical analysis prediction indicators because there is no unique usage of the technical indicators.

Technological methods are those analytical methods used to make predictions based on statistical or artificial intelligence models computed by digital computers. It started with the study of the application of time series forecasting on stock price prediction (Box and Jenkins, 1976). For these types of methods, autoregressive integrated moving-average (ARIMA) or multivariate regression are some sample time-series forecasting techniques used. *Pesaran* and *Timmermann* (1994) presented a good example using multivariate regression to predict the Dow Jones Industrial Average (DJIA) and S&P500 index. Later, Pesaran and Timmermann (1995) conducted another study on the predictability of stock return for the US stock market.

On the other hand, together with the development of artificial intelligence and machine learning, the application of these prediction methods to the stock market was introduced. Support Vector Machines (SVM) (Cortes, C. & Vapnik, V. 1995) and Artificial Neural Network (ANN) (Haykin 1998) are some popular techniques used for predicting the stock market. SVM is a concept for a set of learning methods analysing data and recognising patterns by classifying the data set. A standard SVM is a non-probabilistic binary linear classifier which mathematically classifies input into two possible classes. An ANN is a computational model consisting of an interconnected set of artificial neurons simulating the biological neural networks. Different information as input is then passed through these artificial neurons and changes the weighting of different input during the learning phase. As such, the complex relationships between input and output can be modeled.

Many studies on SVM application to time-series stock market prediction have been conducted by researchers recently. *Cao* and *Tay* (2001) used SVM to predict the movement of the S&P 500 Daily Index in the Chicago Mercantile. The degree of accuracy of this study is measured by the estimates' deviation from the observed value (Cao and Tay 2001). Apart from actual value, *Kim* (2003) made predictions on the index movement direction of the South Korea Stock Market with SVM. The best prediction accuracy for the holdout data is 57.8%. *Tony Van Gestel* (2001) implemented the LS-SVM time series model to predict the stock price movement of German stocks.

In recent years, many studies on stock market predictions using ANN have been conducted (Kimoto et al., 1990; *Y*oon and Swales, 1991; Freisleben, 1992; White, 1993; Baestaens and van den Bergh, 1995; *Y*ao and Poh, 1996; Lawrence R, 1997; *Y*ao et al., 1999). All these studies proved that ANN can predict the stock price with relatively higher accuracy and generate positive return among different stock markets. Though the results were encouraging, these studies did not take into consideration the transaction costs. P.M. Tsang, P. Kwok, S.O. Choy, R. Kwan, S.C. Ng, J. Mak, J. Tsang, K. Koong and T.L. Wong (2007) extended these studies and implemented NN5 for stock price prediction on the Hong Kong Exchange. They found that ANN can also attain a high accuracy level exceeding 70%, but the return generated can be negative after transaction costs are included. This inspires the study on the relationship between prediction accuracy, return and transaction costs.

RESEARCH METHODOLOGY

As the main objective of this study is to assess the relationship between prediction accuracy and realised return, a trading strategic decision making system should be designed for simulation purposes. This trading strategic decision making system can then be used to test the realised return under a different prediction accuracy level so that the relationship between prediction accuracy and realised return can be reviewed according to the result of the simulation *(See Chart 1).*

Trading strategic decision making system

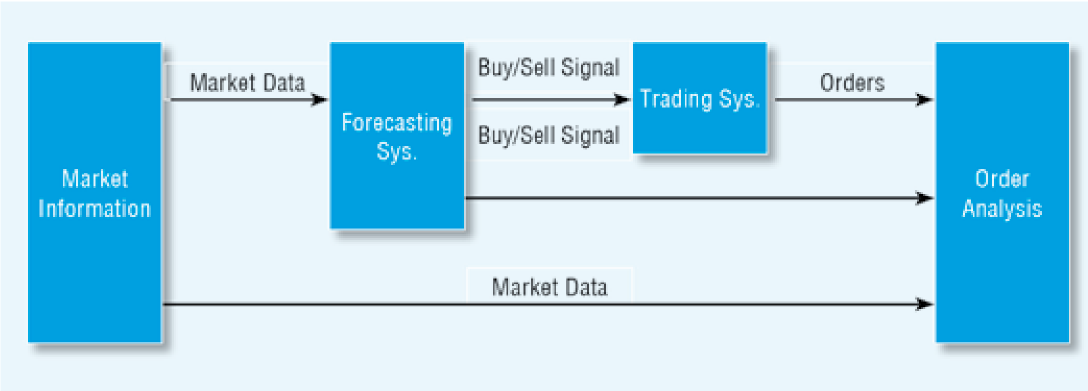
The trading strategic decision making system should involve three components:

- a Market Information System,
- a Forecasting System
- and a Trading System.

The order analysis component is used to consolidate and analyse the orders made by the Trading system. Market Information System is used to consolidate data from different market data sources and feed into the Prediction System. The prediction system is the component applying the forecasting logic including all prediction methods such as Fundamental Analysis, Technical Analysis and

Chart 1

TRADING STRATEGIC DECISION MAKING SYSTEM



Artificial Intelligent Methods like ANN and SVM. No matter which prediction method is used, the prediction system is a unit accepting market information as input and generating buy/sell signals based on the prediction result. The third component is the trading system which is used to make transactions based on the buy/sell signals from prediction. The SPDR Dow Jones Industrial Average ETF (NYSEARCA:DIA) will be used as the proxy for the Dow Jones Industrial Average Index. The exchange traded fund ETF is an investment tool which tracks the performance of other assets or indexes. In this case, the SPDR Dow Jones Industrial Average ETF tracks the Dow Jones Industrial Average Index whereby when the Dow Jones Industrial Average Index rises 5%, the DIA will also rise 5%. The volatility, i.e. the market risk, is also the same as the Dow Jones Industrial Average Index. As long as the Dow Jones Industrial Average Index can be a typical representation of the US stock market, we can treat the Dow Jones Industrial Average Index as an indicator of market performance. Therefore, we can treat the ETF DIA as a tradable market portfolio which attributes the equivalent market return and market risk corresponding to the US stock market. The market data is downloaded from Yahoo! Finance the ETF NYSE:DIA monthly data from 1998 to 2012 of which is collected and listed in a table.

Prediction system

This prediction system is the heart of the whole Trading Strategic Decision Making System that, without changing other components, the prediction system can be plugged with different models, from technical analysis to fundamental analysis, from ARIMA to ANN. Obviously, different models have different levels of prediction accuracy. As the purpose of this research is to study prediction accuracy, a random buy/hold/sell signal generator will be designed and implemented in this part so that for each run it can generate prediction signals in different levels of accuracy relying on randomness. The output signal of the prediction system is chosen as two values only: buy/sell signals because this is the best fit for the input of trading or the order system. In case multiple buy signals occur consecutively, the system is able to purchase stock shares based on a ratio. For example, if the ratio is set as 100% for each buy signal, all cash will be used to purchase shares. On the other hand, for each sell signal, all holding shares should be sold and returned to cash. If three consecutive buy signals occur, the first signal indicates a purchase with all cash. For the subsequent second and third buy signals, it indicates to purchase more shares.

Compared to the value based signal which predicts the exact value of the forecasted stock price, the two value signals can definitely help reduce the complexity of the system. Also, a buy/sell order is closer than the output of a value which cannot determine which trading action follows. Therefore, the chosen buy/sell output can then be easily fed into the trading system and simulate the buy/sell actions. In addition, a different trading strategy can be determined and compared in this component. The traditional buy-and-hold strategy will be chosen as one test case and simulate the result by generating one buy signal at the start of the period and one sell signal at the end of the period. The day trade strategy will be another test case with the buy signal occurring only at a daily open price and the sell signal occurring at the daily closing price of the stock. Irregular trade strategy can also be applied in which buy and sell signals are not required to be in the same period every time, e.g. the sell signal can occur after one day, two days and three days for each buy signal that occurs.

In this regard, to simulate the accuracy of different prediction systems, random prediction signals will be generated with multiple iterations. For each period or duration in this study, monthly data, buy/sell signals are generated randomly with a chance of 50% each. Throughout the years examined, a number of buy/sell signals will be generated and result in a realised return with a specific accuracy. So for one iteration, it generates a number of buy/sell signals, with the realised return of the portfolio at the end of the years examined. If one iteration represents the performance of a particular prediction method, the systems run multiple times and can generate the results of multiple prediction methods with varying prediction accuracy and return. As a random generating signal with a probability of 50% is chosen, it is expected that the accuracy and return are distributed as binomial. With sufficient sample sizes, the distribution can be near normal distribution. A high (99%) and low degree (1%) of accuracy can be achieved after several iterations. This is the key part of the study that consolidates market information, prediction information and order information from Trading Strategic Decision Making System. This component is for consolidating and analysing the orders made by the trading system. The cumulative return will be calculated after consolidating orders, and the accuracy or hit rate will be calculated by counting the correct buy/sell orders against the total number of orders. Then the relationship between Prediction Accuracy and Return can be plotted and reviewed. This will be followed by asserting the transaction costs for each trade and further reviewing the result.

RESULT AND FINDINGS

Model building

A component is the prediction system which includes a random signal generator so that different levels of accuracy resulted after several iterations. In the prediction system, it includes a field of "% buy signal" which indicates the ratio of buy signal occurrences. For example, if this is set as 50%, it means that the system generates buy and sell signals with an equal chance of 50% each. On the other hand, if this is set as 10%, it means that the system generates buy signals less than sell signals, which 9 out of 10 on average are sell signals. Obviously, if 0% is set, it will never buy and this system is not a valid model. Random signals can only be either 1 or 2, with 1 meaning bull and 2 meaning bear. In other words, assume in reality someone predicts (by whatever means, it can be technical or fundamental analysis) the price can be propelled to a higher level or it can sink to a lower level. Also, it can be configured to set the percentage of cash or assets involved in each transaction/prediction. If we set "%Buy/Sell for each prediction" as 100%, it means each time a buy signal occurs, it will use all cash to purchase shares or sell all holdings in exchange for cash. However, if this value is set as 50%, it means each time a buy signal is received, it will use only 50% of current cash to purchase shares or sell 50% of holdings.

The trading order/purchase order system is simulated by another table with buy/sell instructions, transaction fees, transaction units, transaction amounts, asset holdings, asset values, cash, and portfolio values listed together with the market information component. The buy/sell instruction denotes the trading action based on the prediction signals generated in the prediction system. The transaction fee is a constant value which can be set at the initial stage of the simulation. The transaction unit is the number of shares to be purchased or sold for each transaction. To simplify the calculation, the transaction unit involves a decimal place which is calculated by dividing cash and the market price of the asset. Transaction amount is the cash flow involved in each transaction which negative value denotes a buying transaction with negative cash flow paying for the shares. Positive value denotes a selling transaction with a cash flow increase in the transaction. Asset holding is the total number of shares which the portfolio holds while asset value is the multiple of asset holding and market price. Cash denotes the cash component of the overall investment the cash and the assets of which form a portfolio. For each iteration, the initial cash holding is USD 1,000 where currency is not included in the spreadsheet but it is assumed that USD is used. Portfolio value denotes the sum of cash and asset value in the market. It treats both cash and assets as part of the portfolio, and performance is measured by counting both the cash holding and market value of the asset holding.

Subsequently, other columns are used to count the performance of each trade. This is used to check the accuracy and realised return on each trade. For each buy signal, we can buy by using 50% of cash or sell 50% of current stocks holdings. However, it is difficult to tell a successful buy/sell pair. Therefore, it is chosen to use the concept of "Average buy price" which means that for each buy transaction, it can dilute the average price of the asset holding. For example, if we first buy 20 shares at $80 each, the average buy price is $80 for each share. Later on, when we buy 20 more shares at $60 each, the average buy price becomes $70. We used $160 for the first trade and $120 for the second trade. We used $280 to buy 40 shares in total and therefore on average we bought 40 shares at $70 each. In this case, we can see that if a buy signal results in purchasing stocks at a lower price, we can say that a successful sell can be made by selling the stock at a price higher than the average buy price. For example, recalling the above case, if a sell signal is generated and the market is valuing the stock price at $75, it is a successful sell as it can result in profit. As such, two columns are used for measuring performance: Average Buy Price and Profit of Sell.

By generating prediction signal and calculating the return of each trade (sell instruction), it can be treated as one iteration by continuing this process from the starting date of market data to ending date of market data. The result is summarised and final portfolio value, accuracy, benchmark value and alpha return against benchmark are recorded. Accordingly, each iteration can be treated as the realised result of one prediction method or prediction trial. So we have to run multiple iterations so as to simulate the result of different prediction methods. As such, the accuracy rate will vary from low to high so that it simulates the accuracy distribution among different simulation methods. After running the above specified process multiple times (1,000 times in this study), two series of data are consolidated for further study: Accuracy and Alpha Return against Benchmark %.

In this sense, Alpha Return and Accuracy can be formulated as the following relationship by linear regression: $A = \beta \cdot \alpha + \varepsilon$ where A is the accuracy, α is the alpha return, β is the coefficient, ε is the residual.

So the research is run based on the above process by first running the test with 1,000 iterations and plotting the linear regression of the resulting accuracy and alpha return. Based on this, several scenarios are simulated accordingly with these varying parameters: transaction fee, Bull % per Signal and %Buy/Sell for each prediction.

Scenario 1

From 100% Buy/Sell for each prediction to 10% Buy/Sell for each prediction

This scenario studies the behaviour of the accuracy and alpha return relationship by varying the percentage of buy/sell ratio for each prediction. In this case, 100% Buy/Sell for each prediction means that for each buy/sell signal, it uses 100% cash to buy shares or sell all holdings for a sell signal. In contrast, for a 10% Buy/Sell for each prediction it means that for each buy/sell signal, only 10% cash is used to buy shares while reserving 90%. If there is another consecutive buy signal, it then buys another sum of shares using 10% cash. In other words, for each trade, it does not buy everything or sell everything but just a proportion because it is expected that later on it can buy at a lower price or sell at a higher price. This is a typical trading strategy which is referred to or known as Balanced Investing. Under this strategy, only partial change is applied to the portfolio so that the risk is considered to be reduced. This can help investors or traders to control the loss or the exposure of their portfolio. Fifty per cent Bull per Signal and 0% transaction fee are set as constants for all cases of this scenario.

100% Buy/Sell for each prediction

Distribution

With 100% Buy/Sell for each prediction, it is found that the distribution on the alpha return is not that normally distributed. The mean alpha return is –17.32%, and by observing the *Chart 2* below, it can be seen that most of the trials are right skewed. Therefore, it means that the number of trials below the mean value is higher than the number of trials above the mean value. It follows from this that the number of trials with positive return is lower than those with negative return; 75% percentile is at around –0.03 and alpha return starts to become positive at around 78%.

Regression summary report

Referring to the *Table 1* and *Chart 3* above, accuracy and alpha return can be plotted by a linear regression line. The attributes of linear regression include *Y*-Intercept and coefficient beta. It found that the *y*-intercept is at 0.60381, i.e. around 60% and that the coefficient beta is 0.127547. This means that when alpha return is 0, the accuracy is at a level of around 60%.

90% Buy/Sell for each prediction

Distribution

With 90% Buy/Sell for each prediction, it is found that the distribution on the alpha return is not that normally distributed. The mean

Chart 2

HISTOGRAM

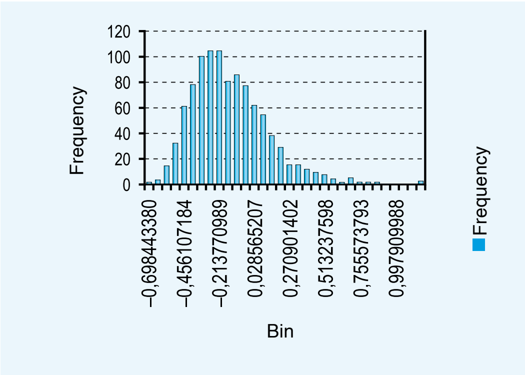


Chart 3

REGRESSION MODEL

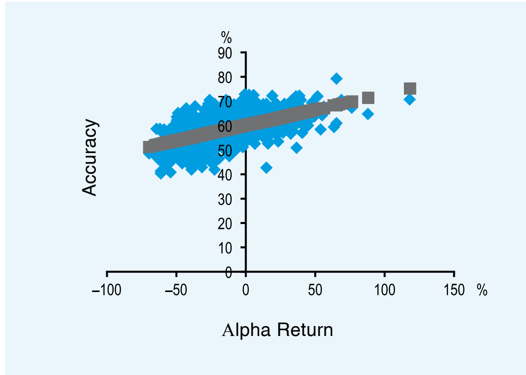


Table 1

INTERCEPT

	Coefficients	Standard Error	*t Stat*	*P-value*	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.60381	0.001842	327.8862	0	0.600196	0.607424	0.600196	0.607424
X Variable (1)	0.127547	0.005943	21.46112	2.61E-84	0.115884	0.13921	0.115884	0.13921

alpha return is –18.43%, and by observing the graph below, it can be seen that most of the trials are right skewed. Therefore, it means that the number of trials below the mean value is higher than the number of trials above the mean value. It follows from this that the number of trials with positive return is lower than those with negative return. Seventy-five percentile is at around –0.05 and alpha return starts to become positive at around 80%.

Regression summary report

Referring to the *Table 2* and *Chart 5* above, accuracy and alpha return can be plotted by a linear regression line. The attributes of linear regression include *Y*-Intercept and coefficient beta. It found that the *y*-intercept is at 0.619658, i.e. around 62% and the coefficient beta is 0.091617. This means that when alpha return is 0, the accuracy is at a level of around 62%.

80% Buy/Sell for each prediction

Distribution

With 80% Buy/Sell for each prediction, it is found that the distribution on the alpha return is not that normally distributed. The mean alpha return is –18.1% and by observing the graph below, it can be seen that most of the trials are right skewed. Therefore, it means that the number of trials below the mean value is higher than the number of trials above the mean value. It follows from this that the number of trials with positive return is lower than those with negative return. Seventy-five percentile is at around –0.04 and alpha return starts to become positive at around 80%.

Regression summary report

Referring to the *Table 3* and *Chart 7* above, accuracy and alpha return can be plotted by a linear regression line. The attributes of

Chart 4

HISTOGRAM

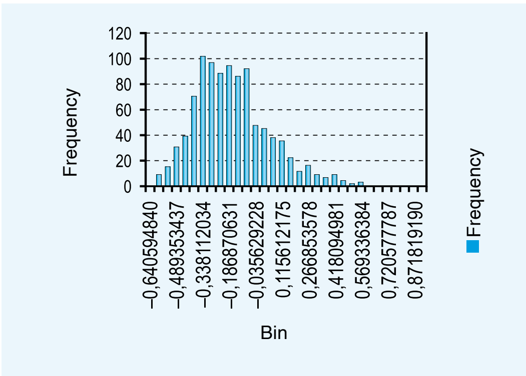


Chart 5

REGRESSION MODEL

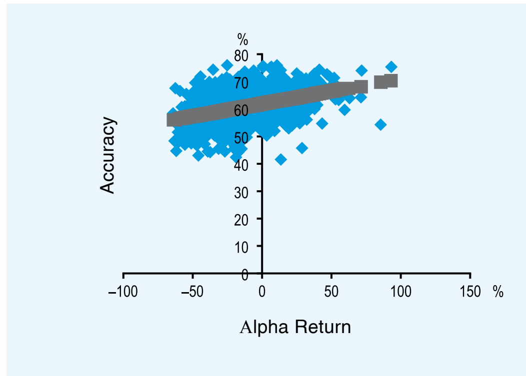


Table 2

INTERCEPT

	Coeffi-cients	Standard Error	*t Stat*	*P-value*	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.619658	0.002067	299.7856	0	0.615602	0.623714	0.615602	0.623714
X Variable (1)	0.091617	0.007017	13.05608	4.41E-36	0.077846	0.105387	0.077846	0.105387

Chart 6

HISTOGRAM

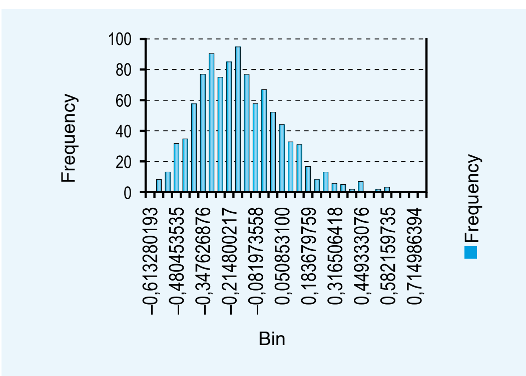


Chart 7

REGRESSION MODEL

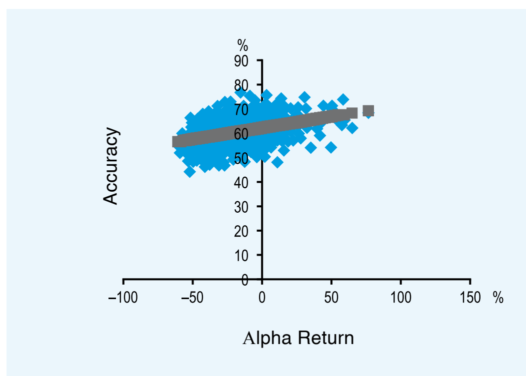


Table 3

INTERCEPT

	Coeffi-cients	Standard Error	*t Stat*	*P-value*	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.624953	0.002058	303.6856	0	0.620915	0.628991	0.620915	0.628991
X Variable (1)	0.09225	0.007414	12.44249	3.83E-33	0.077701	0.106799	0.077701	0.106799

linear regression include *Y*-Intercept and coefficient beta. It found that the *y*-intercept is at 0.624953, i.e. around 63% and that the coefficient beta is 0.09225. This means that when alpha return is 0, the accuracy is at a level of around 63%.

Scenario 2

From 90% Bull per Signal to 10% Bull per Signal

This scenario studies the behaviour of the accuracy and alpha return relationship by varying the percentage of Bull per signal. In this case, a 90% bull per signal means that at each random buy/sell signal, there is a 90% chance it will generate a buy signal and a 10% chance it will generate a sell signal. On the other hand, for a 10% bull per signal, there is a 10% chance it will generate a buy signal and a 90% chance it will generate a sell signal. So the chances of generating bull and bear signals are unequal. This is particularly important to see if any impact occurred for a rising market that if there are more predictions for bull, it can buy more and result in better performance on both prediction accuracy and alpha return. A 0% Transaction Fee and 100% Buy/Sell per prediction are set as constant for all cases of this scenario.

90% Bull per Signal

Distribution

With a 90% Bull per Signal, it is found that the distribution on the alpha return is not that normally distributed. The mean alpha return is –3.87% and by observing the graph below, it can be seen that most of the trials are right skewed. Therefore, it means that the number of trials below the mean value is higher than the number of trials above the mean value. It follows from this that the number of trials with positive return is lower than those with negative return. Seventy-five percentile is at around 0.08 and alpha return starts to become positive at around 73%.

Regression summary report

Referring to the *Table 4* and *Chart 9* above, accuracy and alpha return can be plotted by a linear regression line. The attributes of linear regression include *Y*-Intercept and coefficient beta. It found that the *y*-intercept is at 0.635094, i.e. around 63% and that the coefficient beta is 0.139865. This means that when alpha return is 0, the accuracy is at a level of around 63%.

70% Bull per Signal

Distribution

With 70% Bull per Signal, it is found that the distribution on the alpha return is not that normally distributed. The mean alpha return is –10.61% and by observing the *Chart 10* below, it can be seen that most of the trials are right skewed. Therefore, it means that the number of trials below the mean value is higher than the number of trials above the mean value. It follows from this that the number of trials with positive return is lower than those with negative return. Seventy-five percentile is at around 0.08 and alpha return starts to become positive at around 70%.

Regression summary report

Referring to the *Table 5* and *Chart 11* above, the accuracy and alpha return can be plotted by a linear regression line. The attributes of linear regression include *Y*-Intercept and coefficient beta. It found that the *y*-intercept is at 0.607449, i.e. around 60% and that the coefficient beta is 0.098654. This means that when alpha return is 0, the accuracy is at a level of around 60% level.

50% Bull per Signal

Distribution

With 50% Bull per Signal, it is found that the distribution on the alpha return is not that normally distributed. The mean alpha return is –18.97% and by observing the graph below,

Chart 8

HISTOGRAM

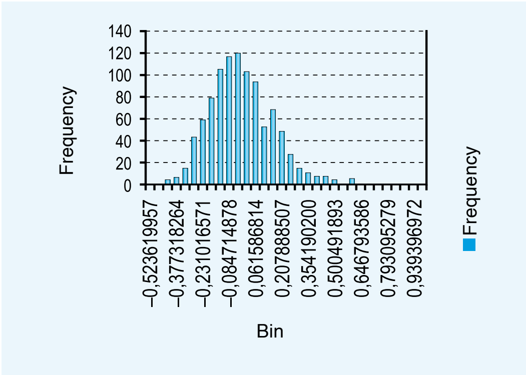


Chart 9

REGRESSION MODEL

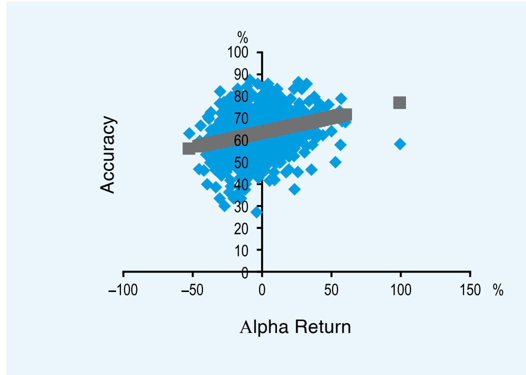


Table 4

INTERCEPT

	Coeffi-cients	Standard Error	*t Stat*	*P-value*	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.635094	0.002987	212.6175	0	0.629232	0.640956	0.629232	0.640956
X Variable (1)	0.139865	0.01624	8.612268	2.76E-17	0.107996	0.171734	0.107996	0.171734

Chart 10

HISTOGRAM

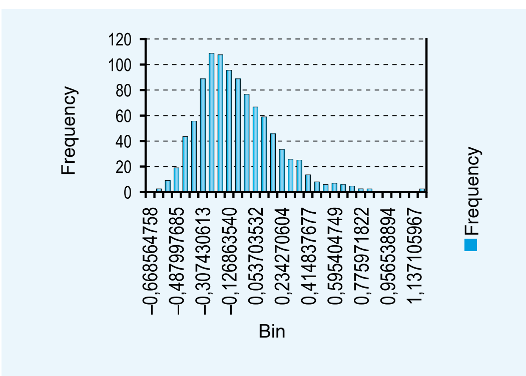


Chart 11

REGRESSION MODEL

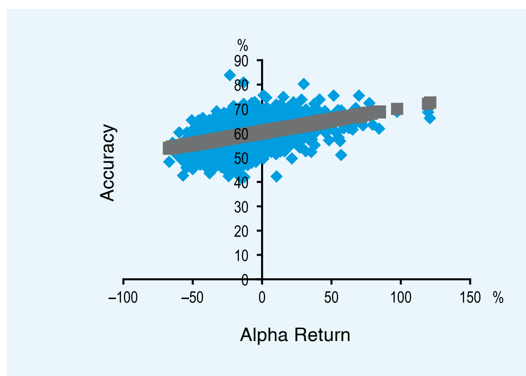


Table 5

INTERCEPT

	Coeffi-cients	Standard Error	*t Stat*	*P-value*	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.607449	0.001893	320.8967	0	0.603735	0.611164	0.603735	0.611164
X Variable (1)	0.098654	0.006783	14.54391	1.32E-43	0.085343	0.111965	0.085343	0.111965

it can be seen that most of the trials are right skewed. Therefore, it means that the number of trials below the mean value is higher than the number of trials above the mean value. It follows from this that the number of trials with positive return is lower than those with negative return. Seventy-five percentile is at around 0.0007 and alpha return starts to become positive at around 75%.

Regression summary report

Referring to the *Table 6* and *Chart 13* above, the accuracy and alpha return can be plotted by a linear regression line. The attributes of linear regression include *Y*-Intercept and coefficient beta. It found that the *y*-intercept is at 0.605371, i.e. around 60% and that the coefficient beta is 0.132869. This means that when alpha return is 0, the accuracy is at a level of around 60%.

Scenario 3

From 0% Transaction Fee to 2% Transaction Fee

This scenario studies the behaviour of the accuracy and alpha return relationship by varying the percentage of transaction fee. In this case, it ranges from no transaction fee to a transaction fee of 5% which occurred as a loss for each buy/sell trade. So as the number of trade transactions increases, the loss of each trade will be accumulated and result in a reduction in overall performance of active trading. A 50% Bull per Signal and 100% Buy/Sell per prediction are set as constant for all cases of this scenario.

1% Transaction Fee

Distribution

With a 1% Transaction Fee, it is found that the distribution on the alpha return is still

Chart 12

HISTOGRAM

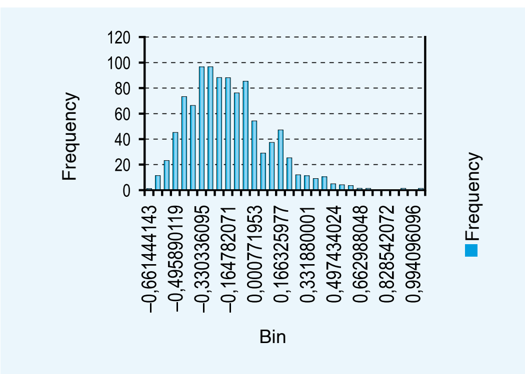


Chart 13

REGRESSION MODEL

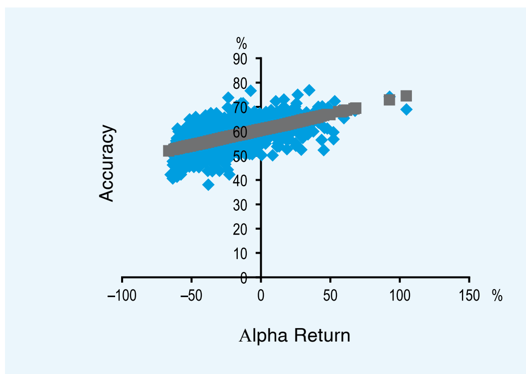


Table 6

INTERCEPT

	Coefficients	Standard Error	*t Stat*	*P-value*	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.605371	0.00199	304.1932	0	0.601466	0.609276	0.601466	0.609276
X Variable (1)	0.132869	0.006396	20.7741	6.04E-80	0.120318	0.14542	0.120318	0.14542

not that normally distributed. The mean alpha return is –19.47% and by observing the graph below, it can be seen that most of the trials are right skewed. Therefore, it means that the number of trials below the mean value is higher than the number of trials above the mean value. It follows from this that the number of trials with positive return is lower than those with negative return. Seventy-five percentile is at around 0.0052278 and alpha return starts to become positive at around 75%.

Regression summary report

Referring to the *Table 7* and *Chart 15* above, the accuracy and alpha return can be plotted by a linear regression line. The attributes of linear regression include *Y*-Intercept and coefficient beta. It found that the *y*-intercept is at 0.60558, i.e. around 60% and that the coefficient beta is 0.132127. This means that when alpha return is 0, the accuracy is at a level of around 60%.

SUMMARY OF THE RESULT

A summary of the results is presented in *Appendices 1 to 3.*

ANALYSIS AND DISCUSSION

With the simulation result, it shows that the alpha return is skewed to negative which implies that most active traders or investors are unable to take advantage of market opportunities. Transaction fees have a negative effect on performance, though it affects only alpha return and not the required level of accuracy. However, it shows that positive alpha return can still be achieved with the presence of Transaction Fee.

Required level of accuracy

Referring to Scenario 1, this linear relationship attributes the required level of accuracy that normally with around 60% level of accuracy a prediction model can achieve positive alpha return. The value of 60% accuracy is the *y*-intercept of the linear equation by substituting alpha return with zero. So it means if accuracy of prediction can be achieved higher than 60%, it can result in positive alpha return, i.e. realised return can be higher than benchmark return. Recall that there are multiple prediction methods or forecasting techniques using machine calculation or computing. We also found that most techniques can achieve around 60–70% accuracy with positive return, for example, the study conducted by P. M. Tsang, P. Kwok, S. O. Choy, R. Kwan, S. C. Ng, J. Mak, J. Tsang, K. Koong and T. L. Wong (2007) which found that the prediction accuracy of Neural Network is already over 70% on Hong Kong Stock Market. For future study or development of stock market prediction we can expect that, according to the linear regression result, with 80% accuracy it can achieve over 100% return higher than the benchmark.

Positive Alpha Return exists if required level of accuracy is met

There is a long history of discussion on alpha return since *Jensen* (1968) first proposed the idea. Alpha return is the additional return on top of benchmark return. One side is based on the traditional theory that market return and market risk are non-diversifiable and a wise strategy is to adopt a hold-and-buy strategy. Another side is based on Jensen's finding that it insists on the existence of alpha

Chart 14

HISTOGRAM

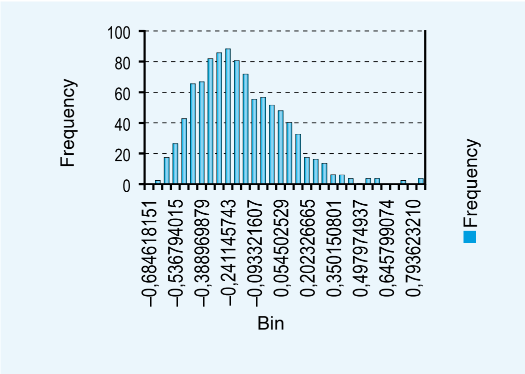


Chart 15

REGRESSION MODEL

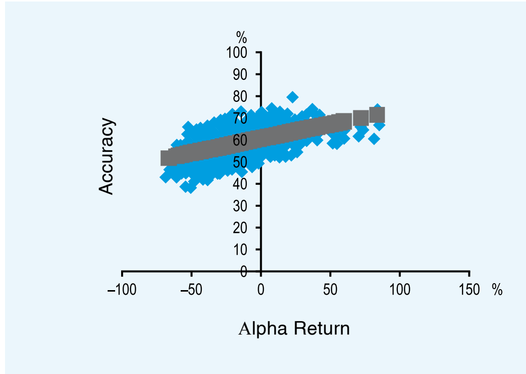


Table 7

INTERCEPT

	Coefficients	Standard Error	*t Stat*	*P-value*	Lower 95%	Upper 95%	Lower 95.0%	Upper 95.0%
Intercept	0.60558	0.002034	297.6823	0	0.601588	0.609572	0.601588	0.609572
X Variable (1)	0.132127	0.006487	20.36696	2.19E-77	0.119396	0.144857	0.119396	0.144857

return on top of market return. Generally, the controversial point on the existence of alpha return is mainly whether market return really reflects all market information and behaves as all investors expected. In other words, if alpha return exists, it either means that the market does not efficiently indicate market return or the market return with alpha return indicates that a more optimised market portfolio should be adopted. However, this study shows that positive alpha return is achievable if a particular level of prediction accuracy can be achieved. This also explains why in the investment world there are some investors who can always win, e.g. *Warren Buffett*. One corollary is that if Buffet can keep winning and making a profit, his prediction result should be able to achieve a level of at least 60% accuracy.

Positive Alpha Return can still be Achieved with Presence of Transaction Fee

The study conducted by P. M. Tsang, P. Kwok, S. O. Choy, R. Kwan, S. C. Ng, J. Mak, J. Tsang, K. Koong and T. L. Wong (2007), concludes that with presence of transaction fee, the performance of the prediction is severely degraded. Referring to the above findings, it holds the same conclusion partially that the average alpha return is decreased as transaction fees increase. However, the magnitude is different than in the study by P. M. Tsang, P. Kwok, S. O. Choy, R. Kwan, S. C. Ng, J. Mak, J. Tsang, K. Koong and T. L. Wong,(2007); it turns from gain to loss significantly with a transaction cost difference of only 0.15%. However, in this study, the alpha return difference is slight compared to the study by P. M. Tsang, P. Kwok, S. O. Choy, R. Kwan, S. C.

Ng, J. Mak, J. Tsang, K. Koong and T. L. Wong (2007). Also, positive alpha return can still be achieved even if there is a transaction fee of 2%.

The corollary of this finding is that by using this simulation, we can observe the fair transaction fee level. If we imagine all market participants predict the market randomly with some of them predicting more accurately and others less so. In this sense, as mentioned above, most active investors or traders cannot make profits though there are still some who can realise a positive gain. So by using this simulation, maybe using more samples and iterations, we can generate the model which can tell us which level of transaction fee is fair to all market participants. At the very least, it is unacceptable if no market participants can make profit with a particular transaction fee level. This is one area for future study in which we can use this model or process to define the maximum transaction fee level which market participants of one market can afford.

Active trading can be superior to passive holding

Traditional Modern Portfolio Theory tells us that due to the fact that the expected return relative to expected risk level can be optimised by forming a portfolio, the market portfolio representing all assets of the market can be treated as the optimised portfolio in terms of return against risk. Therefore, the most intelligent or wise strategy is to hold the market portfolio passively without active trading because all other risks related to asset price can be diversified except the market risk of the market portfolio. However, this study created some evidence that positive alpha return can be achieved under a certain accuracy level of active trading. Although as above mentioned, most active traders or investors do not manage to do this in the stock market. We can still see cases that, with a particular level of accuracy, can result in positive alpha return being achieved. It means that if we add the criteria as the finding of this study, i.e. the level of prediction accuracy, active trading can become meaningful as contrasted to the discussion on the topic in the past.

Most active traders or investors are unable to take advantage

For Scenario 1, 100% buy//sell for each prediction case, the total number of positive alpha return is 224. It means that there are only 224 predictions out of 1,000 iterations that can produce positive alpha return. As each iteration represents the random prediction process throughout the study duration (from 1998 to 2012), the alpha return performance should be randomised and distributed under a certain pattern. It can be observed that no matter which parameters have changed, from 100% buy/sell for each iteration to a 5% transaction fee, all the histograms displaying the distribution generated are right skewed. As each iteration simulates one prediction method which can be sometimes accurate but sometimes inaccurate, we can find that most of time the alpha return of each iteration is negative.

Buy all sell all better than buy little sell little each time

For Scenario 1, the objective is to observe the behaviour if the proportion of each active trading affects the alpha return. By varying the % Buy/Sell for each prediction from 100% to 10%, we can observe that the mean alpha return decreases as % Buy/Sell for each prediction decreases. The number of positive alpha return also decreases from 224 to 0. Though the mean accuracy is increased from 58% to 72.79%, it cannot offset the negative effect of 0 positive alpha return resulted for 10% Buy/Sell for each prediction. The fact is that with 100% buy/sell for each prediction, there can still be 224 trials with positive alpha return with a required accuracy of 58% only. This implies that better performance is to be expected using a buy all sell all active trading strategy. One can take advantage if he or she chooses to make predictions actively and buy shares with an all cash for buy signal and sell all holdings for a sell signal. By simulating the buy/sell strategy this provides evidence for the controversial discussion on the effectiveness of buy all/sell all strategy on active trading. We can expect that if we buy stock shares using only 10% of cash and sell only 10% of holdings each time, though the level of accuracy is high, the final alpha return achieved can be low and the probability of getting positive alpha return is near 0.

Pick more buy in bull market

For Scenario 2, the objective is to observe the behaviours if the buy signals and sell signals are not evenly generated. By varying the % of Bull per Signal from 90% to 10%, we can observe that the mean alpha return decreases as % of Bull per Signal decreases. The number of positive alpha return also decreases from 374 to 11. This implies that if there are more buy signals, it can achieve a higher alpha return level. In other words, active traders buy more and are more likely to realise gains. One explanation is that during the observed duration of the market, from 1998 to 2012, the DIA rose from $79.23 to $128. We can treat the US stock market during that period as being in a long run bull market state. As such, we can expect that it is more likely to rise for each period (i.e. each month) of measured duration. Therefore, though the accuracy does not rise, the overall performance on alpha return can be enhanced.

CONCLUSION

In summary, this research uses a computer aided simulation to simulate the behaviour of various stock market prediction methods. It uses randomised signals with only buy/sell signals to simulate the signal output of various prediction methods. It works because whether the method involves computer aided prediction such as Artificial Neural Network, Support Vector Machine, or financial forecasting using fundamental analysis or technical analysis, or even feng shui or fortune teller on prediction, the outputs are the same in that only buy/sell signals will result. These methods can be accurate throughout the years or somewhat accurate but some do not depend on external factors. The result should be the same by listing the results of all prediction methods with different levels of accuracy. This study works because all predictions, though with some baseline rationales and different levels of accuracy, are still random in nature. The pattern is somehow random but not in cycles. Thus it should somehow be able to model the real situation against the real life trading actions.

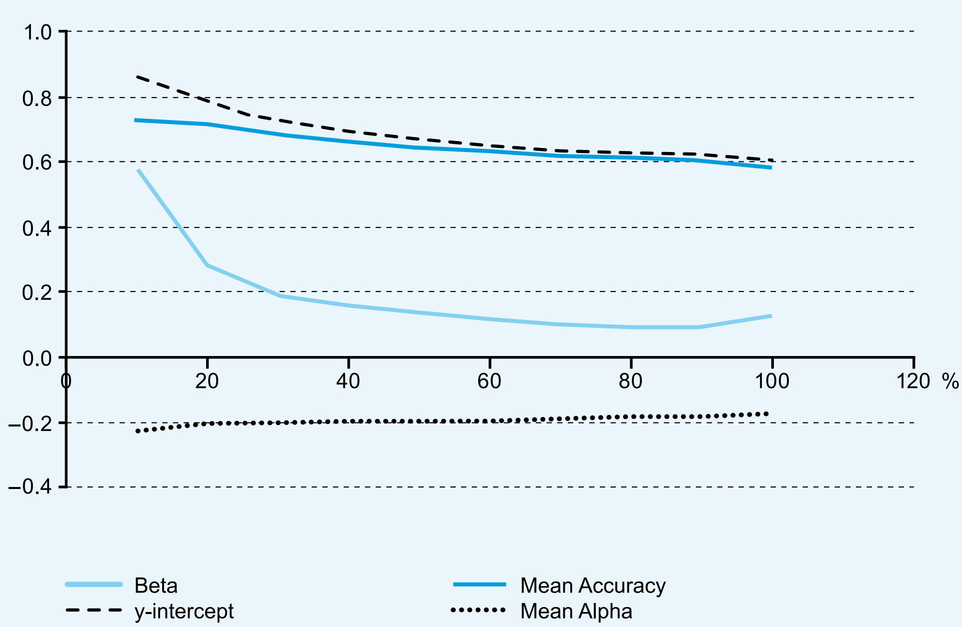
Also, three scenarios are studied with this simulation: from 100% Buy/Sell for each prediction to 10% Buy/Sell for each prediction, from 90% Bull per Signal to 10% Bull per Signal and from 0% transaction fee to 2 % transaction fee. With these scenarios, we can examine different active trading behaviours. The findings can provide evidence on controversial questions regarding active trading and passive holding investing strategies. We found that there is evidence to support that positive alpha return exists by fulfilling the condition that the required level of accuracy met for active trading. Our study shows that positive alpha return can still be achieved even if transaction fee is present. In such a sense, we can conclude that active trading can be superior to passive holding because alpha return can be achieved if the required level of accuracy is fulfilled. With recent development on computer aided stock market prediction, there are some studies on this area such as the study conducted by P.M. Tsang, P. Kwok, S.O. Choy, R. Kwan, S.C. Ng, J. Mak, J. Tsang, K. Koong and T.L. Wong, (2007) which can achieve a level of accuracy of around 70%. In other words, if we apply the neural network prediction of that study with 70% accuracy while the required level of accuracy is 60% according to our study for US market, it should be able to generate positive alpha return. There are three interesting corollaries based on the result of the research: most active traders cannot realise gains, Buy all Sell all and pick more buy are better strategies in a bull market. This study provides evidence for the above observation.

Appendix 1

Chart 16

SCENARIO 1

% Buy/Sell for each prediction	100%	90%	80%	70%	50%	30 %	20%	10%
Beta	0.127547	0.091617	0.09225	0.102529	0.136814	0.19422	0.291558	0.588334
y-intercept	60.38%	61.97%	62.50%	63.43%	66.85%	72.21%	77.73%	86.28%
Mean Accuracy	58.17%	60.28%	60.83%	61.49%	64.19%	68.30%	71.73%	72.79%
Mean Alpha	–17.32%	–18.43%	–18.10%	–18.86%	–19.46%	–20.14%	–20.61%	–22.93%
Number of Positive Alpha	224	186	181	135	110	40	12	0

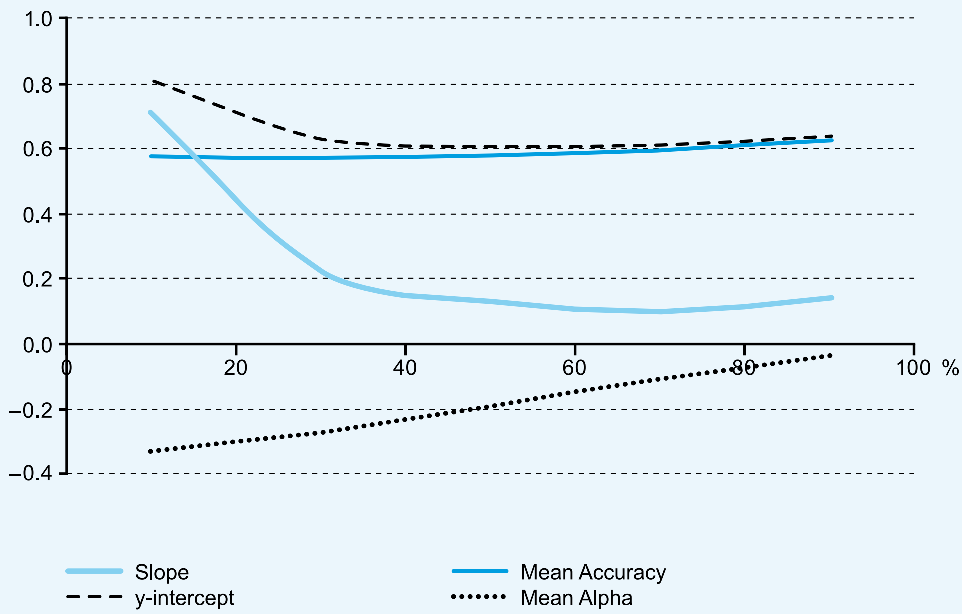


Appendix 2

Chart 17

SCENARIO 2

% Bull per Signal	90%	70%	50%	30%	10%
Slope	0.139237	0.098671	0.132884	0.224435	0.70336
y-intercept	0.635287	0.607447	0.605357	0.631555	0.808933
Mean Accuracy	0.629892	0.596977	0.580152	0.571238	0.576834
Mean Alpha	–0.03874	–0.10611	–0.18967	–0.26875	–0.32999
Number of Positive Alpha	374	296	196	100	11

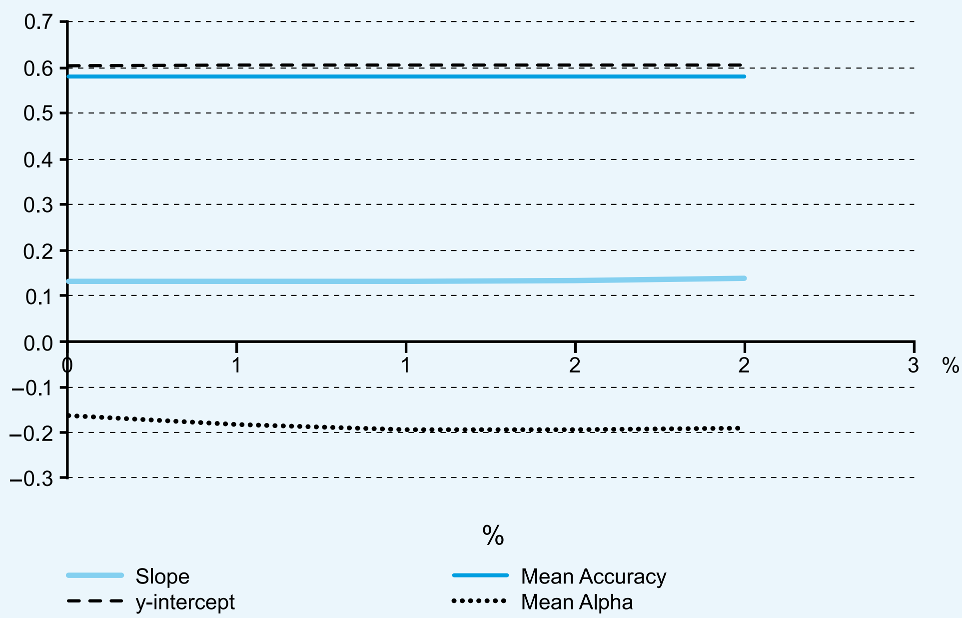


Appendix 3

Chart 18

SCENARIO 3

Transaction Fee	0%	1%	2%
Slope	0.134948	0.133	0.14009
y-intercept	0.606207	0.6058	0.60642
Mean Accuracy	0.584125	0.5799	0.579793
Mean Alpha	–0.16363	–0.19474	–0.19007



Notes

SAMPLE DATA SET WITH MODEL INFORMATION

Table 1

Market Information (DIA)				Prediction System	
Date	Close	Daily Return	Bull % per Signal	Bull/Bear Signal Bull=1 Bear=2	% Buy/Sell for each prediction
Initial			0.5	0.3636364	1
1/20/1998	79.23	0	0.5	2	1
2/2/1998	85.59	0.080273	0.5	2	1
3/2/1998	88.19	0.030377	0.5	2	1
4/1/1998	90.87	0.030389	0.5	2	1
5/1/1998	88.84	–10.02234	0.5	1	1
6/1/1998	89.69	0.009568	0.5	2	1
7/1/1998	88.53	–10.01293	0.5	2	1
8/3/1998	74.87	–10.1543	0.5	1	1
9/1/1998	78.19	0.044344	0.5	1	1
10/1/1998	86.03	0.100269	0.5	2	1
11/2/1998	90.81	0.055562	0.5	1	1

Table 2

Market Information (DIA)	Trading System						
	Date	Transaction Fee	Transaction unit	Transaction Amount	Asset Holding	Asset Value	Cash
Initial		0.02			0	0	1000
1/20/1998	Sell	0	0	0	0	0	1000
2/2/1998	Sell	0	0	0	0	0	1000
3/2/1998	Sell	0	0	0	0	0	1000
4/1/1998	Sell	0	0	0	0	0	1000
5/1/1998	Buy	0.02	11.03548128	–1000	11.0354813	980.3921569	0
6/1/1998	Sell	0.02	–111.03548128	1009.5678	0	0	1009.568
7/1/1998	Sell	0	0	0	0	0	1009.568
8/3/1998	Buy	0.02	13.21987867	–11009.568	13.2198787	989.772316	0
9/1/1998	Buy	0	0	0	13.2198787	1033.662313	0
10/1/1998	Sell	0.02	–113.21987867	1160.0523	0	0	1160.052
11/2/1998	Buy	0.02	12.52401896	–11160.052	12.524019	1137.306162	0

Table 3

Market Information (DIA)		Performance Measurement	
Date	Portfolio Value	Average Buy Price	Profit of Sell
Initial	1000		
1/20/1998	1000	0	0
2/2/1998	1000	0	0
3/2/1998	1000	0	0
4/1/1998	1000	0	0
5/1/1998	980.3922	88.84	0
6/1/1998	1009.568	0	9.380159087
7/1/1998	1009.568	0	0
8/3/1998	989.7723	74.87	0
9/1/1998	1033.662	74.87	0
10/1/1998	1160.052	0	147.5338459
11/2/1998	1137.306	90.81	0

Literature

Box, G., Jenkins, G. (1976): Time Series Analysis: Forecasting and Control. Holden Day

C. L. Wilson (1994), Self-organizing neural network system for trading common stocks. In Proc. ICNN'94, Int. Conf. on Neural Networks, pages 3651–3654, Piscataway, NJ, 1994. IEEE Service Center

C. Lee Giles, Steve Lawrence, Ah Chung Tsoi, (1997): Rule Inference for Financial Prediction using Recurrent Neural Networks, Proceedings of IEEE/IAFE Conference on Computational Intelligence for Financial Engineering (CIFEr), IEEE, Piscataway, NJ, 1997, pp. 253–259

Cao, L. J., Tay, F. E. H.(2001): Financial Forecasting Using Support Vector Machines. Neural Comput, (2001) pp. 184–192

Cortes, C. & Vapnik, V. (1995): Support-vector networks. Machine Learning 20, pp. 273–297

Coulson, D. R., (1987): The Intelligent Investor's Guide to Profiting from Stock Market Inefficiencies. Probus Publishing

Dirk Emma Baestaens and Willem Max van den Bergh(1995): Tracking the Amsterdam stock index using neural networks. In Neural Networks in the Capital Markets, chapter 10, pages 149–162. John Wiley and Sons, 1995

Edward Sek Khin Wong, (2009): Cognitive Pattern Analysis Employing Neural Networks: Evidence from the Australian Capital Markets, International Journal of Economics and Finance, Vol. 1, No. 1 Feb 2009

Fama, E. F., (1965): The behavior of stock market prices. Journal of Business 38, pp. 34–105

Freisleben, B., (1992): Stock market prediction with backpropagation networks. The Fifth International Conference Industrial and Engineering Applications of Artificial Intelligence and Expert System

Gestel, T.V., Suykens, J. A. K., Baestaens, D.E., Lambrechts, A., Lanckriet, G., Vandaele, B., Moor, B.D., Vandewalle, J. (2001): Financial Time Series Prediction Using Least Squares Support Vector Machines within the Evidence Framework. IEEE Transactions on Neural Networks, 12 (2001) pp. 809–821

H. White. (1993): Economic prediction using neural networks: The case of IBM daily stock returns. In Neural Networks in Finance and Investing, chapter 18, pages 315–328. Probus Publishing Company, 1993

Haykin, S., (1998): Neural Networks: a Comprehensive Foundation, seconded. Prentice-Hall, Englewood Cliffs, NJ

J. Yao, C. L. Tan, and H. L. Poh, (1999): Neural Networks for Technical Analysis: A Study on KLCI, International Journal of Theoretical and Applied Finance, Vol. 2, No.2, 1999, pp. 221–241

Jensen, M. C. (1968): "The Performance of Mutual Funds in the Period 1945–1964," Journal of Finance 23, 1968, pp. 389–416

Kim, Kyoung-jae (2003): Financial time series forecasting using support vector machines. Neurocomputing, (2003) pp. 307–319

Kimoto,T., Asakaya, K., Yoda, M. And Takeoka, M. (1990): Stock market prediction system with modular neural networks. Proc. IEEE International Joint Conference on Neural Networks, pp. 1–6

Lawrence R, (1997): Using Neural Networks to Forecast Stock Market Prices. University of Manitoba

Malkiel, B.G. (1996): A Random Walk Down Wall Street, 6th edn. W.W. Norton, New York

Murphy, J. J., (1999): Technical Analysis of the Financial Markets: a Comprehensive Guide to Trading Methods and Applications. New York Institute of Finance

P.M. Tsang, P. Kwok, S.O. Choy, R. Kwan, S.C. Ng, J. Mak, J. Tsang, K. Koong and T.L. Wong (2007): Design and Implementation of NN5 for Hong Kong Stock Price Forecasting, Engineering Applications of Artificial Intelligence, Vol. 20 , 2007, pp. 453–461

Paderborn, Germany, 451–460. Beltratti, A., Margarita, S. and Terna, P. (1996): Neural Networks for Economic and Financial Modelling. London, Thomson Computer Press

Pesaran, M. H., and A. Timmermann, (1994a): Forecasting stock returns. An examination of stock market trading in the presence of transaction costs, Journal of Forecasting 13, pp. 335–367

Pesaran, M. H., and A. Timmermann, (1995): The robustness and economic significance of predictability of stock returns, Journal of Finance, vol. 50, pp. 1201–28

Yoon, Y. and Swales, G. (1991): Predicting stock price performance: a neural network approach. Proc. IEEE 24th Annual International Conference of Systems Sciences, pp. 156–62