An empirical methodology for developing stockmarket trading systems using artificial neural networks

Vanstone, Bruce; Finnie, Gavin

Published in:
Expert Systems with Applications

DOI:
10.1016/j.eswa.2008.08.019

1-1-2009

An empirical methodology for developing stockmarket trading systems using artificial neural networks

Bruce J. Vanstone
Bond University, bruce_vanstone@bond.edu.au

Gavin Finnie
Bond University, Gavin_Finnie@bond.edu.au

An Emperical Methodology for Developing Stockmarket Trading Systems using Artificial Neural Networks

Bruce Vanstone
Bond University
Gold Coast, Queensland, Australia
bruce_vanstone@bond.edu.au

Gavin Finnie
Bond University
Gold Coast, Queensland, Australia
gavin_finnie@bond.edu.au

ABSTRACT

A great deal of work has been published over the past decade on the application of neural networks to stockmarket trading. Individual researchers have developed their own techniques for designing and testing these neural networks, and this presents a difficulty when trying to learn lessons and compare results. This paper aims to present a methodology for designing robust mechanical trading systems using soft computing technologies, such as artificial neural networks. This methodology describes the key steps involved in creating a neural network for use in stockmarket trading, and places particular emphasis on designing these steps to suit the real-world constraints the neural network will eventually operate in. Such a common methodology brings with it a transparency and clarity that should ensure that previously published results are both reliable and reusable.

1. INTRODUCTION

Artificial Neural Networks (ANNs) have earned themselves an excellent reputation as non-linear approximators. According to Tan [1], of all the AI techniques available, ANNs are the technique which deals best with uncertainty. Like other forms of soft-computing, ANNs exhibit a high tolerance to imprecision and perform well in noisy data environments.

These characteristics of ANNs make them particularly suited to financial problem domains, particularly the arena of financial trading. The stockmarket represents a datasource with an abundance of uncertainty and noise. Indeed, it is these characteristics which make financial trading both particularly challenging and particularly rewarding.

There have been repeated attempts by a number of researchers to apply ANNs to the problem of developing mechanical trading systems. In this context, a mechanical trading system is one which operates according to a fixed set of rules, and contains no discretionary components. Unfortunately, there are a number of concerns with the manner in which this field appears to have progressed.

Principal amongst these concerns is a lack of a formal methodology describing the procedure to follow to create mechanical trading systems using soft computing, indeed, a lack of a formal methodology describing how to create any mechanical trading systems. The reasons for this are crystal clear. Successful systems traders jealously guard their intellectual capital regarding their development methodologies. This inhibits many researchers from developing systems which would actually work when real-world constraints were applied.

Apart from the lack of understanding of the methods of creating such trading systems, there is a further lack of understanding about the correct way to benchmark and test such systems. Again, this is understandable. Unfortunately, this appears to have led to a situation where published results may be the end combination of a vast number of prior experiments, leading to poorly developed systems. In practice, such systems usually fail rapidly.

Further complications in the development process come from a lack of understanding of the constraints of real-world trading, such as accounting for transaction costs, and the implementation of money management algorithms.

Developing economically feasible trading systems using ANNs is possible, although it is clearly a complex task. In many ways, ANNs themselves compound the difficulty. As noted by White [2], neural methods of back-propogation are designed to reduce network training error. Perhaps a better objective would be to optimize profit subject to risk, an opportunity not directly afforded by using neural networks.

The objective in using ANNs to design trading systems is to attempt to resolve the mismatch between traders' behavioural expectations, and ANN training methods, through careful selection of the ANNs inputs and outputs.

Conducting the development process within a well defined methodology, subject to real-world trading constraints and effective benchmarking is the key to designing successful mechanical trading systems.

This paper will aim to lay down a methodology for developing a mechanical trading system using ANNs. From an ANN viewpoint, it will cover the selection of inputs, the ANN architecture, and the ANN outputs. From a trading system viewpoint, it will focus on real-world constraints, and the technique of benchmarking.

The starting point for the creation of a trading ANN is the selection of variables likely to influence the desired outcome. There are a number of methods used by practitioners to find such variables, and they broadly fall within the remit of either Fundamental Analysis, or Technical Analysis. It is essential to have an understanding of these two complementary forms of analysis, so that an intelligent choice of inputs can be made, and these can be matched to likely outputs in terms of their possible effect and duration.

This paper will review research in the areas of fundamental and technical analysis, to arm the reader with the required knowledge to confidently select neural inputs and likely outputs. The mechanics of trading system structure and design are also reviewed. The paper will then continue with a study of ANNs themselves, paying particular regard to architecture. At that point, the paper will progress to a study of real-world trading constraints, such as slippage, accounting for costs, and money management. Finally, the paper will describe appropriate techniques to benchmark a mechanical stockmarket trading system.

The methodology presented here is distilled from a combination of trading systems research and experience. This paper does not attempt to claim that following this methodology will lead to guaranteed success, nor that there is no other way to develop a successful ANN trading system.

It simply attempts to add an element of process to a research area whose results are often as chaotic as the underlying processes studied.

2. LITERATURE REVIEW

Both the fields of Fundamental Analysis and Technical Analysis have long histories of using their variables as predictors of future returns. This Literature Review highlights some of the key papers which support this approach, and, in the case of Technical Analysis, addresses some concerns of academic credibility.

When searching the domain of possible fundamental or technical variables to include in an ANN, it is generally advised that all variables selected be supported by published research.

2.1. FUNDAMENTAL ANALYSIS

There is a long established tradition of attempting to use the financial ratios produced from fundamental analysis as predictors of a company's future share price (or return, or price direction).

The process of using fundamental variables to make stock trading decisions begins with Benjamin Graham, as early as 1928. Publication of Grahams work dates back to his first book, Security Analysis, in 1934. This book is still in print and is now in its 5th edition. Graham produced a number of books distilling investment wisdom, including the Intelligent Investor, initially published in 1949. This book, last published in 1973, made detailed comments on the building of portfolios and selecting stocks. Graham urged investors to pay attention to three fundamental variables, namely size of firm, capitalization, and price-earnings ratio. The book provided detailed information of how to select companies using these variables. Research by Oppenheimer and Schlarbaum [3], tested Grahams approach to determine its effectiveness. They extracted the rules provided to investors in each of the four editions of The Intelligent Investor, and using publically available information, found that positive risk-adjusted rates of return were delivered to investors that followed the approach between 1956 and 1975. Rates of return were 3% to 3.5% higher than a naïve buy-and-hold stategy (in a frictionless market). When the various market frictions (costs) were taken into account, rates of return were 2% to 2.5% higher than the buy-and-hold strategy. Oppenheimer and Schlarbaum state that '…it is reasonable to conclude that our evidence contradicts the semi-strong form of the efficient market hypothesis'.

According to Lowe [4], Graham also published a list of ten attributes of an undervalued stock, which could be used by investors seeking excess return. These ten attributes were:

- Earnings-to-price yield >= double the AAA bond yield,
- P/E <= four-tenths highest average P/E in most recent 5 years
- Dividend yield >= two-thirds the AAA bond yield,
- Price <= two-thirds tangible book value per share
- Price <= two thirds net current asset value
- Total debt less than tangible book value
- Current ratio >= 2
- Total debt <= net quick liquidation value
- Earnings doubled in most recent 10 years, and
- No more than two declines in earnings of 5 percent or more in the past 10 years

It was noted that few companies could meet all 10 criteria.

Grahams work inspired a number of researchers to focus on detecting security price return anomalies that could be ascribed to fundamental variables.

Graham's value investment philosophy is well entrenched, in part due to the success of Warren Buffett, who is widely recognized as the world's greatest 20th century investor. Buffett credits his success to Graham, however, according to Bierig [5], rather than just seeing a balance sheet as a frozen snapshot of a company, Buffett broadens his definition of value, and investigates the dynamics of the company. In this sense, Buffet has become subjective rather than objective.

Basu [6] investigated whether stocks with low P/E ratios earned excess returns when compared to stocks with high P/E ratios. It was found that during the study period (April 1957 – March 1971), portfolios built from low P/E stocks earned higher returns than those portfolios built from higher P/E stocks, even after adjusting returns for risk. The study concluded that there is an information content present in publicly available P/E ratios, which could offer opportunities for investors, and that this was inconsistent with the semi-strong form of the EMH. There are some clear parallels with the first two guidelines of Graham's 10-point list here. The first guideline suggested earnings-to-price yield be double the AAA bond yield. The earnings-to-price yield is the inverse of the P/E ratio, and ensuring it is greater than the AAA bond yield effectively capped the P/E ratio. In this manner, it steered investors away from high P/E stocks. The second guideline required P/E be four-tenths highest average P/E in most recent 5 years, again, effectively steering the investor away from high P/E stocks.

Banz [7] focused on the 'size effect'. Essentially, the size effect concerns the relationship between the market capitalization of a firm, and its return. Banz reports that during the study period (1936 – 1975), common stock of small firms had higher returns than the common stock of large firms, even after adjusting for risk. Banz also raises the issue that the size effect may just be a proxy for one or more other factors, which are correlated with size, an interpretation he also applies to Basu's findings concerning the P/E effect.

Also in 1981, Reinganum [8] described a misspecification of the simple one-period CAPM model, namely, that data on firm size can be used to create portfolios that earn abnormal returns. From studying small firms listed on the New York and American Stock Exchanges, during the period from 1963 to 1977, Reinganum discovered average rates of return for small firms to be nearly 20% per year greater than those of large firms.

Rosenberg et al. [9] presented two strategies aimed at exploiting fundamental information to increase returns. The first, the "book/price" strategy buys stocks with a high ratio of book value to market price, and sells stocks with the reverse. The second strategy, "specific return reversal" computes specific returns per stock, and relies on the observation that specific returns tend to reverse in the subsequent month. Thus, this strategy buys stocks with negative specific returns in the preceding month, exploiting this reversal. The study sourced data from Compustat, on 1400 of the largest companies, from 1980 to 1984, and stocks were priced mainly from the NYSE. The study demonstrated statistically significant results of abnormal performance for both strategies, and suggests that prices on the NYSE are inefficient. Here, the first strategy provides support for Graham's fourth guideline, namely that price be two-thirds tangible book value per share, effectively steering the investor toward stocks with a higher book value than price.

DeBondt and Thayler [10] present evidence that investors tend to overreact when considering recent data. This overreaction led to a reversal effect, with stocks that had been prior 'losers' likely to become future 'winners'. The researchers also investigate seasonality patterns in returns data. They demonstrate that the winner-loser effect is not primarily a size effect, and the earnings of 'winner' firms and 'loser' firms show reversal patterns consistent with overreaction. In terms of seasonal influence, DeBondt and Thayler report that excess returns for 'losers' are negatively related to both long-term and short-term formation performance, particularly in January. For 'winners', they find that January excess returns are negatively related to the excess returns for the prior December.

Detailed research from Fama and French [11] surveys the above style of anomaly detection, and conclude that if asset-pricing is rational, then size and the ratio of book value of a stock to its market value must be proxies for risk, as opposed to reflecting market inefficiency.

Lakonishok et al [12] find that a wide range of value strategies (based on sales growth, Book-to-market, Cash flow, earnings, etc) have produced higher returns, and refute Fama and French's claims that these value strategies are fundamentally riskier. Using data from end-April 1963 to end-April 1990, for the NYSE and AMEX, Lakonishok et al find evidence that the market appears to have consistently overestimated future growth rates for glamour stocks relative to value stocks, and that the reward for fundamental risk does not explain the 10% - 11% higher average returns on value stocks. This study lends further support to the fourth guideline, again effectively steering the investor toward stocks with a higher book value than price.

Fama and French [13] respond to Lakonishok et al by focusing on size and book-to-value, and form portfolios of stocks partitioned by these variables from the NYSE, AMEX and NASDAQ, from 1963 to 1992. Their results demonstrate that both size and BE/ME (book-to-market equity) are related to profitability, but find no evidence that returns respond to the book-to-market factor in earnings. They conclude that size and BE/ME are proxies for sensitivity to risk factors in returns. Their results also suggest that there is a size factor in fundamentals that might lead to a size-related factor in returns.

Later, Fama and French [14] study returns on market, value and growth portfolios for the US and twelve major EAFE countries (Europe, Australia, and the Far East). They recognize that value stocks tend to have higher returns than growth stocks, finding a difference between low B/M (Book-to-market) stocks and high B/M stocks of 7.68% per year on average. They find similar value premiums when investigating earnings/price, cash flow/price and dividend/price. They find that value stocks outperform growth stocks in twelve of thirteen major markets during 1975 – 1995. They also find a value premium in emerging markets. Fama and French conclude that these results are explained by a one-state-variable ICAPM (or a two-factor APT) that explains returns with the global market return and a risk factor for relative distress.

Frankel and Lee [15] estimate firms fundamental values (V) using I/B/E/S concensus forecasts and a residual income model. They find that V is highly correlated with stock price, and that the V/P ratio is a good predictor of long-term returns. They state that this effect is not explained by a firm's market beta, B/P ratio, or total market capitalization (size). They also find evidence that errors in consensus analysts forecasts are predictable, and these prediction errors can be exploited by incorporating the error with V/P. They conclude that the evidence suggests that firm's value estimates may well provide a better forecast ability than simply using ratios, and that prices converge to value estimates gradually over greater than 12 month horizons. They also state that the predictability of long-term forecast errors in consensus forecasts is consistent with a long-term mispricing hypothesis. Finally, the authors also acknowledge that the results may demonstrate yet another proxy for cross-sectional risk differences, but state that this is an unlikely conclusion.

Piotroski [16] investigates whether fundamental analysis can be used to provide abnormal returns, and right shift the returns spectrum earned by a value investor. In anomaly terms, Piotroski focused on high book-to-market securities, and shows that the mean return earned by a high book-to-market investor can be shifted to the right by at least 7.5% annually, and a simple investment strategy based on high book-to-market securities generates a 23% annual return between 1976 and 1996. The research is stimulated by the observation that portfolios of high book-to-market firms normally contain several strong performing firms (achieving strong returns), and many deteriorating ones (achieving poor returns). Piotroski defines three different classes of financial performance signals, namely:

- Profitability,
- Leverage, Liquidity and source of funds, and,
- Operating Efficiency.

From these three classes of signals, nine simple signals are defined, and an aggregate score of the nine signals is used to rank the constituents. The nine signals involve seven fundamental variables, namely:

- net income before extraordinary items,
- cash flow from operations, (both scaled by the beginning of year total assets),
- leverage,
- liquidity,
- whether external financing has been raised recently,
- current gross margin scaled by total sales, and
- current year asset turnover ratio.

Within the portfolios constructed from the higher aggregates, Piotroski notes that the returns are concentrated in small and medium sized companies, companies with low share turnover, and firms with low analyst following. It is also noted that superior performance is not dependant on initial low share prices. Again, support is found for Graham's fourth guideline in this study. Of further interest is the determination that one-sixth of the annual return difference between the ex-ante strong and weak firms is earned over the four three-day periods surrounding earning announcements. This information is of obvious interest to those advocating market timing approaches.

Kanas [17] finds a non-linear relation between stock returns and the fundamental variables of dividends and trading volume.

Aby et al. [18] focus on combining fundamental variables to screen stocks for value. This is a reasonably common approach, with some authors reporting outstanding results. Aby et al. developed portfolios based on four fundamental conditions, namely: Single Valued P/E (P/E<10), Market Price < Book Value, established track record of return on Shareholder Equity (ROE > 12%), and dividends paid out less than 25% of earnings. They conclude that when the four criteria are used to screen stocks, quality investments seem to result, again providing support for Graham's fourth guideline. The authors state that higher yields do not seem to provide good long term returns, possibly due to the use of retained earnings to enhance equity per share. Overall, the main contribution of their work is to establish a relationship between ROE (> 12), and share price performance. The research alludes to the fact that Buffett believes 12 is an appropriate value for ROE in domestic (US) markets. The authors find that the value of 12 for ROE provides a clear line of demarcation between performance and non-performance is share price terms. The authors tested the filter criteria against the Value Line database between August 31, 1989 to August 31, 1999. The filter conditions described cut the database down from 6000 possible stocks to just 14. These 14 yielded an average return of 30.55% per year for the ten years. It is interesting to note that in earlier work Aby et al. [19], the same authors had focused on shares with simply a low P/E and a market price below book value, and had concluded that this filter method did not produce satisfactory returns.

2.2. TECHNICAL ANALYSIS

Modern Technical Analysis dates from the work of Charles Dow, who in 1884 drew up an average of the daily closing prices of 11 important stocks. Between 1900 and 1902, Dow wrote a series of articles in the Wall Street Journal documenting stock price patterns and movements he observed in the average. These articles were the first to describe systematic phenomena in the stock markets.

Although Dow's work represents the beginning of modern technical analysis, it is worthy of note that markets and analysis techniques existed centuries before this, notably in Japan since 1730, where the first futures contracts (in rice) were traded. For a fascinating description of the operation of these markets, refer to Nison [20]. Tvede [21] reports that interest in the future prices of these rice 'futures' ran high, with the Japanese government suspending the forward market in 1869 due to excessive volatility.

Today, a manual of technical analysis is likely to be composed of techniques which fall into one of three primary classifications, namely:

- Charting (typically pattern matching),
- Indicators (and oscillators),
- Esoteric approaches

This paper will focus on the technical indicators and oscillators, as these are easily reproduced according to their mathematical definitions. In contrast, Charting and pattern matching is usually highly subjective and without rigorous mathematical definition. Esoteric approaches are excluded from this study, as they have no scientific justification. Warnecke [22] provides examples of the criticisms often leveled at these techniques.

Technical Analysis is enjoying a recent resurgence in academia after having been out of favour for several decades. The main reason for this lack of favour concerns the Efficient Market Hypothesis (EMH), which supports the random-walk theory. In essence, since the early work of Fama [23], the random-walk theory has held sway. This theory states that successive price changes in stock prices are independent, identically distributed random variables. The important implication of this hypothesis is that it implies that a series of price changes has no memory, which further implies that the study of past prices cannot provide a useful contribution to predicting future prices. As the majority of technical analysis techniques focus on probability based on past price behavior, the natural conclusion is that Technical Analysis cannot work.

Regardless of the random walk theory, a large number of market participants use technical analysis as their main method of stock selection. Indeed, Taylor and Allen [24] conducted a UK survey of forex dealers on behalf of the Bank of England, and found that at least 90% of the respondents placed some weight on Technical Analysis for decision making. It has been suggested that due to its high usage, technical analysis may, in fact, be becoming a self-fulfilling methodology. In other words, if enough people follow the principles, then those principles can be expected to become manifest in the character of price time series.

To complete the discussion regarding technical analysis, it is occasionally stated that as technical rules become more widely known, the abnormal returns they attempt to identify will be reduced, and the usefulness of the technical rules themselves will be destroyed. Silber [25] finds against this conclusion, instead concluding that ‘the continued success of simple technical trading rules is possible as long as there are price smoothing participants in the market’. In this context, Silber’s example of price smoothing participants refers to the central banks.

Rather than focus on Technical Analysis as a discipline, the remainder of this literature review will focus on the research support for the use of Technical Analysis variables, such as Moving Averages, Indicators and Oscillators.

The majority of the academic literature concerning technical analysis concerns the testing of simple technical rules, such as moving averages. Truly effective technical rules are not published in academic journals, and are usually kept secret.

According to Pring [26], there are three basic principles of Technical Analysis, namely:

- Prices move in trends,
- Volume goes with the trend,
- A trend, once established tends to persist

The moving average and its derivatives are designed to expose when a security has begun trending, and as such, deal with the first and third principles listed above. The idea of observing (and profiting from) trends has a long history, and is one of the early systematic phenomena described by Dow.

Academic research in the area of moving averages dates from the work of Neftci and Policano [27], who studied moving averages, and the slope of price movements on the chart (named trendlines by technical analysts). They studied closing prices of gold and T-bills, and created buy-and-sell rules based on trendlines and moving averages. Although they described their results from the study of trendlines as inconclusive, they reported a significant relationship between moving average signals and prices. Of particular interest was the fact that a set of significant parameters for one commodity were often insignificant for another commodity. This difference in significant parameters is often termed a markets ‘personality’.

Murphy [28] demonstrated that different sectors of the market move in relationships with other sectors, a field of study now known as Intermarket Analysis.

Neftci [29] examined the relationship of the 150 day moving average rule to the Dow-Jones Index. This research concluded that the moving average rule generated Markov times (no dependence on future information) and has predictive value.

Two popular technical trading rules were tested by Brock et al. [30], namely, moving averages and trading range breaks (known by technical analysts as Support and Resistance trading). Using data from the start of the DJIA in 1897 to the last trading day of 1986, the authors test a variety of combinations of moving averages, using a 1% band around predictions to eliminate whipsaws. They find support for the use of moving averages, and report that the differences in utility are not readily explained by risk. They conclude their results are consistent with the technical rules having predictive power.

Inspired by Brock et al [30] above, Mills [31] tests the same two trading rules in the London Stock Exchange, using FT30 data from 1935 – 1994. Mills’ results are remarkably similar to Brocks, with Mills concluding that the trading rules could predict stock prices, and are thus profitable in periods when the market is inefficient.

Levich and Thomas [32] test currency futures contracts in five currencies over the period 1976 to 1990. They report persistent trading profits over the 15 year period using a variety of commonly researched moving average rules. Levich and Thomas concluded ‘the profitability of trend following rules strongly suggest some form of serial dependency in the data, but the nature of the dependency remains unclear’.

LeBaron [33] provided more support for the moving average, by using moving average rules as specification tests for foreign exchange rates. He concluded that exchange rates do not follow the random walk, and that the deviations are detected by simple moving average rules.

Lehmann [34] considers evidence supporting variation in equity returns, attempting to decide whether the evidence is indicative of predictable changes in expected return, or market inefficiency. Lehmann finds that 'winners' and 'losers' one week often experience reversals of fortune in the following week. The costless portfolio constructed by Lehmann (difference between 'winner' and 'loser' portfolios) showed profit in 90% of weeks. Lehmann concludes that the reversals of fortune are probably reflections of the imbalances in the market for short-term liquidity, and states that 'it is difficult to account for these results within the efficient markets framework'. Lehmann's work is often quoted by practitioners as supporting Technical Analysis, as it supports the idea that price trends occur frequently enough to create profit opportunities for technical traders. Lehmann does not specifically make this statement.

Jegadeesh [35] examines the predictability of monthly returns on individual securities. Ten portfolios were formed based on the predicted returns using estimates of the regression parameters. The difference between abnormal returns on the extreme decile portfolios was 2.49 percent per month over the period 1934 to 1987. Slightly different values are provided when comparing extreme decile portfolios excluding January results (2.20% per month), and when January was considered separately (4.37% per month). Jegadeesh rejects the random walk hypothesis, and concludes that returns predictability is due to either market inefficiency, or systematic changes in expected stock returns. This paper is often used to support the principles of technical analysts, as it shows evidence that increases (and decreases) in prices during one month are often reversed out the following month. Patterns of that nature would suggest that investors could profit from technical trading strategies, and would also be a breach of market efficiency.

Very little academic research exists supporting the use of specific technical indicators and oscillators. The main academic work above relates to Moving Average rules and Momentum based rules. To allow the neural network to have access to the same types of indicators and oscillators being used by practitioners, a survey of the main practitioners journal, 'The Technical Analysis of Stocks and Commodities' was conducted. For the sake of brevity, detailed reviews are not provided for the articles studied, rather, a list of the most 'popular' (i.e. most frequently referenced) technical variables is provided below. The assumption is that these variables are most in use due to the fact that they are useful.

Technical Variables most frequently cited in 'The Technical Analysis of Stocks and Commodities'
Moving Averages (including a variety of derivatives built from basic moving averages)
Volatility based variables, such as ATR (Average True Range)
Volume based variables, such as Volume directly, or OBV (On Balance Volume)
ADX (Average Directional Index – See Wilder [36])
Stochastics (based on the work of George Lane)
Momentum (both price and volume)
RSI (Relative Strength Index – See Wilder [36])
Variety of miscellaneous indicators and oscillators (eg MACD, Intermarket indicators, Money Flow, TRIN (Traders Index), etc)

To sum up the position regarding technical analysis, it is reasonable to state that after a long absence from academia, technical analysis is beginning to enjoy a return to mainstream investment finance. It is becoming more common to see universities promote subjects with titles such as 'Computational Finance', and even Siegel [37] 'cautiously' supports the use of Moving Averages.

It should be noted that Computational Finance courses are not specifically devoted to Technical Analysis, but also cover other topics, such as Behavioural Finance and Intelligent Finance.

Intelligent Finance seeks to develop a comprehensive undetstanding of financial markets by the combination of fundamental, technical and strategic analysis. To pursue this field further, the reader is referred to Pan [38].

2.3. TRADING SYSTEMS

According to Chande [39], a trading system consists of three major functions, namely:

- Rules to enter and exit trades,
- Risk Control, and,
- Money Management

2.3.1. RULES TO ENTER AND EXIT TRADES

In the approach used in this paper, the ANN will be the signal generator, and its output will be used to either initiate or exit trades. Rules to enter and exit trades are based on the strength of the ANN output signal. This approach is discussed in detail in Section 3 Neural Network Creation.

2.3.2. RISK CONTROL

In the context of stock market trading, a trader is typically concerned with downside risk, which describes how much money is at risk on an individual trade-by-trade basis. This method of approaching risk leads to traders placing orders to sell/buy securities to cover open long/short positions when losses cross pre-determined thresholds. These are known as stop-loss orders.

As investors are typically preoccupied with return, it is also appropriate to consider risk to be appropriately controlled by trade risk within the confines of a trading system. After all, this is the entire purpose of a trading system. This method of considering risk is growing in popularity, see for example Kaufman [40], Longo [41], and Pocini [42].

A general framework for considering the issue of risk control is the TOPS COLA approach described by Chande [39]. TOPS COLA is an acronym for "take our profits slowly, cut off losses at once". In effect, it describes the traders approach to risk.

Trend following systems will typically have more losing trades than winning trades. In financial terms, this still leads to a viable system, as long as the value of losing trades is quite low, and/or the value of winning trades is high. Typically, according to Chande, about 5% of the trades made by a trend following system are the 'big ones'. In light of this information, it is easy to see how the TOPS COLA approach can work.

Risk control may therefore be defined as the process for managing open trades using predefined exit orders. Traders enter an initial stop-loss when they first enter a trade. If the price of the security falls, then a stop-loss is triggered and a loss is taken. Clearly, the closer the stop is to the actual price of the security, the less money will be lost if the price falls, but the more likely it is that a smaller price fluctuation or random noise will trigger the stop. If the stop price is further away from the current price, then there is potentially more money at risk, however, the chance is lower that the stop will be triggered by noise. Chande provides evidence to suggest that the use of tight stops may well be degrading long term portfolio performance.

If the price of a security rises, the trader may well adjust the stop loss to a break even stop, or a trailing stop. A break-even stop will cancel out the trade if the price falls, at an amount equal to the trade cost. A trailing stop will increase in value as the security price increase, either proportionately or otherwise. The trailing stop will never move down, it is only ever moved up under the security price, so once the price of the security falls, the stop will be hit.

A variety of methods are available, such as Volatility stops (refer to Overholser [43]), Fixed Dollar / Fixed Percent stops, Support and Resistance stops, Time delay stops, and Dev-Stops (refer to Kase [44, 45]).

There are a number of variants of the stops described above, and stops may also be set relative to volatility, or period of time elapsed, or combinations of the above. Also, the use of profit targets is not uncommon. These effectively exit a trade once a predefined amount of profit has been achieved. There are variants of profit exits also, but, in general, they run contra to Chande's principle of "take our profits slowly".

When trading using longer term classical bar charting in the market, it is not uncommon for traders to use very simple stops, typically setting a basic money management stop on the initial purchase, and leaving that stop in for the duration of the trade. The trade entry is determined by the detection of a trend, and the trade exit is controlled by the evidence that a trend has reversed, or ended. This is entirely consistent with the definition provided by Pring [26], and the use of stops is reserved for exiting out of trades where a trend appeared to begin, but very quickly ended.

It should be noted that there are a variety of other techniques in common usage, a brief summary of these is provided by Tharp [46]. Also, the use of stops within a given strategy, particularly if it is a long-term strategy, may not always be appropriate. Kaufman [40] demonstrates how the performance of a longer-term trending strategy without stops is most consistent, and concludes that the use of fixed value stops may even conflict with the strategy's objectives.

As can be seen, risk control for a trader concerns protecting open trades where money is at risk through the use of stops. In essence, although there are many styles of stop described above, the use of stops in this paper is simply to be able to site the neural networks developed within a realistic trading environment. For this purpose, an initial stop-loss will suffice.

The stop-loss threshold is selected by the study of the in-sample MAE as described by Sweeney [47], and later by Tharp [46]. The MAE studies the Maximum Adverse Excursion (MAE) of a set of trades, in an effort to determine the extent to which favorable (profitable) trades range into unprofitable territory before closing out profitably. This method of risk management allows traders to study the MAE characteristics of a set of trades, to identify preferred stop-loss points.

As the choice of stop strategy is closely related to the method used to signal entries, a discussion on which stop thresholds to implement is delayed until Section 3 Neural Network Creation.

2.3.3. MONEY MANAGEMENT

Money management, aka position sizing, refers to the actual size of the trade to be initiated, taking into consideration the account equity and potential trade risk.

Like risk control, the style of money management is closely related to the trading system, as it is influenced by many variables which are constrained by the specific system. As every trade carries a potential for loss, there is a need to determine the maximum amount of capital to expose at each trade, given a finite capital base. A number of specific approaches exist, and the reader is encouraged to pursue the following references:

- Kelly system: well described by Balsara [48]
- Optimal *f*: refer to Vince [49]
- Percent of equity: refer to Elder [50], and Pring [26]

To simplify the complexities of Money Management, this paper suggests using a fixed percentage of equity per trade (as suggested by Elder) for testing and benchmarking. Not only is this simple to implement, but it also avoids having to determine how much of any profit effect observed is attributable to the neural network developed, and how much is attributable strictly to money management. Given the goal of this paper, this choice seems appropriate.

3. NEURAL NETWORK CREATION

This section of the paper describes the methodology to produce a neural network which behaves well in in-sample testing. When such a network is identified, the parameters which support this network are effectively frozen for out-of-sample testing. It is at that point that the extent to which the ANN has been curve-fit will be revealed.

3.1.1. SELECTING INPUTS

Specific choices need to be made regarding the inputs to be selected. At this point, it is important to have some expectation of the timeframe that the system will be used to trade. Developing a long-term system may call for a greater reliance on fundamental variables, due to the timeframe of their production, and therefore, their likely influence. Developing a short-term system may call for a greater reliance on technical variables.

It is at this stage that the user must decide on their expected trading timeframe. Typically, a distinction is made between investors and traders, with investors associated with longer-term holding periods, and traders associated with shorter-term holding periods. The choice between trader and investor will influence the decision on which variables, fundamental or technical, will likely be the main neural network inputs.

Choices can be made by studying the published research reviewed above, by experience, or by experimentation. Ideally, variables should be selected in line with published research, however, due to the lack of detailed research regarding technical variables, it is suggested that function profiles be used. Function profiles effectively iterate a function through every possible realistic value, and measure the corresponding change in a different variable for every value of the function. Typically, they display their results graphically for easy interpretation.

When selecting technical variables it is important not to let the ANNs have visibility of actual prices or volume, only ratios and indicators built from the prices and volume. This will prevent the ANNs focusing on whether a stock has a high price (or volume) or a low price (or volume), and allow more for generalization about the ratios and indicators, and their relative relationships. As the goal with neural networks is to encourage generalization, supplying ratios to the neural network is an ideal way to accomplish this. This is because the same ratio can be built from any number of numerators and denominators, and it is unlikely (although not inconceivable) that the exact values of the numerator and denominator are relevant. Technical analysis generally focuses on ideas concerning price and volume behaviour, not the exact price and volume themselves.

Note that when using fundamental variables as inputs, it is advisable to delay their input to the network by some fixed time period. Typically, fundamental data is delayed by about 6 months before being used as an input to a neural network. This ensures that the variables being acted on are available to the market at the time they are made available to the neural network. Whilst some fundamental data is realized much more quickly than 6 months, the long lag is generally used to ensure that the neural network does not have access to data which would not have been available in the real-world. This is an essential element of backtesting.

3.1.2. SELECTING OUTPUTS

Again the selection of output variables is dependant on the expectation and timeframe of the system being developed. From the research presented in the literature review, there are clearly a great many choices for output variables.

While it may seem appropriate to attempt to predict price, a great deal of research shows that this is a particularly difficult task, possibly due to the fact that price changes do not tend to be smooth. Predicting price direction appears easier, and more likely to be successful, but then the trader has no real way of gauging the strength of the move in that direction. For example, a high degree of directional accuracy may not translate to high returns after costs if the movement in the direction forecast is small, as noted by Azoff [51] and Thawornwong and Enke [52]. Ruggiero [53] makes a number of suggestions, such as predicting a smoothed output function, such as a forward shifted technical indicator which is constructed from underlying price data. However, there are a number of inherent advantages and disadvantages in all technical indicators, and whilst they may be smoother than price action, they are typically only suitable for trading whilst market action is favorable to that actual indicator. For example, as Bauer [54], Pring [26] and a host of other technical analysts explain, trend based indicators perform well whilst the market is trending, but perform poorly at other times. Oscillators perform well when the market is not trending, but perform poorly otherwise. The temptation to create two technical neural networks, one for each main type of market activity is easily avoided, as then a further methodology would be required to tell which network to use at which point in time. In any event, a number of academics believe that the market actually goes through three phases, trending, non-trending, and chaotic, making the selection of which network to trust at which point in time much more complex. For further information on training neural networks with chaotic constraints, see Slim and Trabelsi [55]. Finally, according to Ruggiero [56], the decision on which target to predict should be based on three factors, namely, the desired trade frequency, risk / reward criteria, and expertise available.

In consideration of all of the above, it is proposed that the most useful ANN output is an indication of the relevant strength of any movement expected over the forecast period. This should give rise to a highly tradeable indicator, which can be expected to perform during both trending and non-trending (and any other!) phases of the market.

The method suggested here is to build an output variable which effectively captures the strength of expected return movement. Such a variable is easily built by measuring the maximum price change over some timeframe after the variable is sampled. This leads to a neural network which can give a strength forecast based on the inputs provided. This is particularly useful when choosing between a number of possible positions, and in setting thresholds for entry and exit rules.

A simple 200 day strength variable may be computed as:

$$\text{Target} = \left(\frac{\left(\max(close_{i+200}, ..., close_{i+1}) - close_i\right)}{close_i} \right) \times 100$$

For long-term trading systems, 1 year is a typical look-ahead period, see for example, Longo [41] and Reinganum [57].

Specific timeframe lengths could be determined by searching for cycles, but as already stated earlier, it is generally accepted that the market dynamics are changing over time. Thus, there appears little benefit in determining the optimal timeframe at any point in time, as it could not be relied on to hold, thus reducing the expected lifetime of a trading strategy, and adding additional risk (see for example, Balsara et al [58]). Published work exists which describes cycles already found in the Australian stockmarket, such as a 6-day cycle discovered by Pan et al. [59], however, this was discovered in the AORD index, and there is no reason to expect it would hold for individual stocks, each with its own individual characteristics. For example, it could be expected that certain stocks had radically different cycle lengths, such as bank stocks following the economic cycle, resource stocks following the strength of the industrial cycle in other countries, etc. Other published works use PCA (Principal Component Analysis) (e.g. Raposo et al. [60]), or Self Organizing Maps (e.g. Chi et al. [61]).

Essentially, the practical solution, and indeed, the objective, is to select timeframes which are consistent with the traders trading expectations. This idea is consistently presented in literature describing the techniques involved in building trading systems, for example, Ruggiero [56].

Finally, Azoff [51] suggests the network have only one output, to avoid the effect of conflicting outputs pulling the weights in opposing directions during training. In this way the network is effectively focused on one task only.

3.1.3. PARTITIONING AVAILABLE DATA

Any study involving optimization, or neural networks must logically (or better still, physically) separate data which will be used for training, from data that will be used for testing.

There is acceptance within the academic community that the relationship between security prices (and returns), and the variables that constitute that price (return), changes over time as described by Refenes et al. [62] and also by Thawornwong and Enke [52]. In other words, the structural mechanics of the market are changing over time, and their effects on prices are also changing. For this reason, it is necessary to partition data vertically rather than horizontally. A vertical partition of a dataset will divide each securities dataset into two partitions, one for training, and one for testing. Typically, the training dataset is larger, and covers a significant date range of the overall data, whilst the testing dataset is smaller, and used to provide out-of-sample confidence. These two partitions are typically known as in-sample (training), and out-of-sample (testing) partitions. Using this approach, every securities data will be partitioned into training and testing subsets.

The horizontal approach to partitioning splits the entire range of data into either a training or a testing block. For example, horizontally partitioning 10 datasets, with 60% in training, and 40% as testing would yield 6 entire datasets used for training, and 4 entire datasets used for testing. This approach is invalid when it is recognized that the structural mechanics change over time, due to the fact that a neural network may well learn correlations that could not have been known in chronological time, and later, exploit these during the testing phase. This may well lead to higher quality predictions, but is clearly unrealistic.

In summary, it is advised to split all available security data vertically, with the final date of all training subsets aligned. The actual ratio for the split is generally chosen dependant on how much data is available, and is often arbitrarily chosen. However, some general guidelines can be distilled. For example, Ruggiero [56] suggests that the data sets used for training should cover a period at least 5 times longer than the desired life of the model to be produced, and suggests using 80% of the data for training, and 20% for testing. Azoff [51] takes a typical approach, which suggests that the training period should be long enough to cover typical market behaviour, including bull, bear, and sideways moving markets. Kaufman [40] suggests a 70:30 split between training and testing, Kim and Lee [63] suggest an 80:20 split, Gately [64] suggests saving only 10% of the available data for testing, thus a 90:10 split. From an optimization point of view, Pardo [65] suggests choosing a period long enough to cover a variety of market activity, and advises choosing a size large enough to generate at least 30 trades for statistical validity. Pardo also notes that the size of the models test window will affect trading shelf life, specifying that the life of the model will be between one-eight and one-quarter of the test window. Pardo further suggests a rule-of-thumb that the walk-forward window (in optimization) should be approximately 10% - 20% of the optimization window.

There are a wide variety of other competing and complementary guidelines available. In essence, the main principle is to capture as much diverse market activity as possible (with a long training window), whilst keeping as long a testing window as possible (to increase shelf life and model confidence). This study recommends sourcing at least 10 years data for each security, and then performing an 80:20 split. In this way, 80% of the data will be used for training, and 20% will be reserved for testing. This split provides a reasonable compromise, and takes the above guidelines into consideration.

Neural networks must be trained on data which includes delisted securities, to enable the neural network access to data which describes the real world environment. Not including data for securities which have delisted introduces survivorship bias. From a trading point of view, the trader cannot benefit from survivorship bias, so including the data for delisted securities will represent the worst case results, and these will be directly applicable to the trader.

3.1.4. In-Sample Benchmarks

As neural networks are developed in this methodology to be signal generators, and in-sample testing is initially outside of the context of a trading system, the test metric comparing different neural network architectures must be on the basis of their in-sample training. These in-sample metrics are used to assess the quality of the neural network architectures tests (see section 3.1.5 Determining Architecture). It is not appropriate to test each neural network architecture on the out of sample results and select the best performer.

For this reason, it is necessary to define metrics which can be used to test different neural network architectures, and the metrics presented below are focused on identifying how well the neural network has learnt its objective. Different metrics are provided for cases when the neural network is focused on long-term timeframes, or short term timeframes.

In the case of a long-term timeframe, a useful in-sample metric is the networks selectivity. This is because long-term trading networks usually serve as stock screening strategies. That is, they are designed to filter the entire market of stocks, and identify those which have the greatest chance of the highest appreciation within the expected trading time period. The signal generated by these neural networks is effectively a prediction of the likely strength of price increase over the next 1 year period.

To determine how to evaluate a screening strategy, it is necessary to review the purpose of such a strategy. Specifically, a screening strategy is used to reduce (refine) the number of securities that are competing for capital. A traders' requirement is to increase the likelihood of selecting stocks that will significantly increase in value. Thus, a suitable measure of success is to determine the percentage of stocks selected that achieve a pre-specified increase in value. By measuring these values for the entire market, then for the ANN predictions, it can easily be determined whether the neural network is effective. Finally, care must be taken to ensure that a reasonable number of predictions are output by the ANN for this process. Clearly, it would not make sense to select a network with a 100% success rating if there was only 1 trade predicted.

The metric used for this style of longer-term screening strategy measurement, termed Filter Selectivity, is defined as:

$$FilterSelectivity = \frac{(ClosedTrades * 100)}{TotalTrades}$$

where

ClosedTrades is the number of trades closed due to meeting the predefined increase in value
TotalTrades is the total number of trades selected by the screening strategy

For shorter term timeframes, such as those relying solely on technical variables, an objective measure of a accuracy is the measure of expectancy. The idea of expectancy in trading was first raised by Tharp [46], who proposed it as a useful method to compare trading systems. Expectancy is a measure of the expected profit per dollar risked on a fixed position size basis. It is used without money management settings enabled, which is appropriate for the in-sample testing. There are a number of variant formulas for calculating expectancy, this version presented is more conservative than Tharp's; it uses the average loss as the standard of risk (rather than the minimum loss as used by Tharp).

$$EXPECTANCY = \frac{((AW \times PW) + (AL \times PL))}{|AL|}$$

where
AW = Average amount won on profitable trade
PW = Probability of winning
AL = Average amount lost on losing trade (-ve)
PL = Probability of losing

Secondly, to assess the quality of the ANN architecture chosen, it is also appropriate to consider the 'Average Profit/Loss %', which is a standard trading system metric.

3.1.5. Determining Architecture

There are no standard rules available for determining the appropriate number of hidden layers and hidden neurons per layer, although for greater generalization, the smaller the number of hidden nodes and hidden layers the better. General rules of thumb have been proposed by a number of researchers. For example, Shih [66] suggests constructing nets to have a pyramidical topology, which can be used to infer approximate numbers of hidden layers and hidden neurons. Azoff [51] quotes a theorem due to Komolgorov that suggests a network with one hidden layer and 2N + 1 hidden neurons is sufficient for N inputs. Azoff concludes that the optimum number of hidden neurons and hidden layers is highly problem dependant, and is a matter for experimentation. Gately [64] suggests setting the number of hidden nodes to be equal to the total of the number of inputs and outputs. As another alternative, some researchers, for example Kim at al. [67] use a brute force approach, and train a great number of ANNs with different configurations, and then select that configuration that performed best. Yet another approach, such as that used by Jaruszewicz and Mandziuk [68] is to train networks for a fixed number of epochs, or fixing the number of hidden nodes to some arbitrary value, as in Kim and Lee [63]. Zirilli [69] proposes a formula based on prior knowledge of the number of unique patterns the net is expected to learn, but concedes that if you know your feature space well enough, you can usually determine this number of hidden nodes better yourself. Finally, another reasonably popular method is used by some researchers such as Kim & Lee [63] and Versace et al [70], whereby genetic algorithms are used to select between the combinatorial explosion of possible networks given choices such as network type, architecture, activation functions, input selection and preprocessing.

An alternative approach described by Tan [1], is to start with a small number of hidden neurons and increase the number of hidden neurons gradually. Tan's procedure begins with 1 hidden layer, containing the square root of N hidden nodes, where N is the number of inputs. Training the network takes place until a pre-determined number of epochs have taken place without achieving a new low in the error function. For example, ANNs can be trained until no new low had been achieved for at least 2000 epochs. At this point the network would be tested against the in-sample set, and benchmarked using the appropriate metrics described above. A new neural network is now created with the number of hidden nodes increased by 1, and the training and in-sample testing is repeated. After each test, the metric being used for benchmarking is assessed, to see if the new network configuration is superior. This process continues while the networks being produced are superior, that is, it terminates at the first network produced which shows inferior in-sample results.

This approach to training is an implementation of the early stopping method, which aims to preserve the generalization capabilities of neural networks. It is based on the observation that validation error normally decreases at the beginning of the training process, and begins to increase when the network starts to over-fit. Lack of generalization is caused by over-fitting. In an over-fit (over-trained, over-learned) situation, the network begins to memorize training examples, and loses the ability to generalize to new situations.

3.1.6. Setting Signal Thresholds

Each neural network developed has fit itself to the characteristics of the market which the training data represents, within the constraints of its architecture. A simple way to observe this fit is with the use of a function profile. From inspection of the function profiles for each neural network, the threshold at which the neural network output signal begins to signal profitable trades can be easily established.

Therefore, for the in-sample testing, the buy signal should take account of the individual neural networks threshold, and also take account of whether the signal is increasing in strength, or decreasing in strength from its previous forecast. Naturally, the sell signal should also take account of the threshold, and also take account of whether the signal is increasing in strength, or decreasing in strength from its previous forecast. It is also considered a desirable property of a trading system if the rules for exiting a trade are the contra to the rules for entering it.

Therefore, a general buy and a general sell rule can be explicitly stated, and then applied to each trading system. Where x is the signal strength threshold chosen from the function profile, then the entry and exit rules become:

Buy: Buy tomorrow when neural signal output(today) > x, and neural signal output(today) > neural signal output(yesterday)

Sell: Sell tomorrow when neural signal output(today) <= x, and neural signal output(today) < neural signal output(yesterday)

These simple buy and sell rules take account of the threshold signal strengths, and using the same generic buy and sell rule for each network gives greater confidence of the generalization of the results.

This paper suggested that for each neural network, the output is a signal strength rating, scaled between 0 and 100. It is then to be expected that, in general, as the numeric value of the signal increases, so should the expected returns to this signal strength. This general principle should be seen by examining a function profile of the signal output of each neural network.

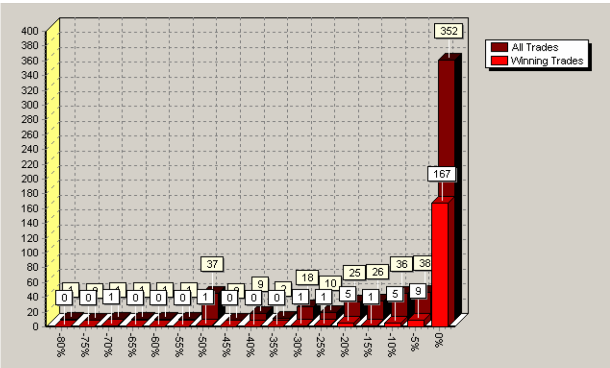
3.1.7. Setting Trading Stops

In this methodology, the MAE technique previously discussed is suggested. This technique can be used to identify an appropriate stop-loss percentage for the in-sample set of trades. This stop-loss percentage is then used to control trading risk for the out-of-sample trades.

By building a histogram of the actual (in-sample) trade data, split according to trades that were eventually won (were profitable), and trades that were eventually lost (were unprofitable), a visual inspection can be made of a useful stop threshold. This information is very valuable to a trader, as it also gives an indication of how the profit/loss percentages will be affected when the stop is introduced. In this approach, the stop percentage value determined from the in-sample data will be then used as the stop value in the out-of-sample testing data.

Typically, the exact value chosen is selected by "eyeballing"; what is required is to locate the point at which the number of 'winning' trades falls away very sharply, whilst typically the number of 'losing' trades does not. According to Chande's principles, if in doubt, we should err towards selecting a larger value of stop loss than a smaller one. This gives a trade plenty of room to develop a profitable outcome. The main reason we should err towards wide stops, is that if a stoploss is hit, it will always force us to incur a loss. There is never a profitable outcome for a stoploss being triggered. This implies that if the decision to take the trade is correct, then we should attempt to stay in the trade for as long as possible, and avoid being shaken out by noise, or short-term adverse behaviour.

The graph below represents the MAE/MFE observations for a trading system which doesn't use stops. These MAE/MFE observations are taken from the in-sample test of the trading system. By "eyeballing" the graph, it is clear that the majority of successful trades don't stray further than 5% into the red before rebounding to conclude profitably. In this case, the threshold would be set at 5%.



4. REAL-WORLD CONSTRAINTS

There are a great variety of tradable securities and trading mechanisms available. The trading systems considered in this paper are all 'long' end-of-day systems. That is, the systems implemented only trade long positions, and do not attempt to sell stocks short.

One of the main issues with short-selling is that it cannot just be assumed that stocks can be sold short. The reason for this is that in some markets only certain securities can be sold short, and these can only be sold short at certain times. For example, the method the ASX (Australian Stock Exchange) uses to determine which shares can be sold short can make short-side selling reasonably complex. In effect, the ASX determine the list of Approved Securities (for short selling), and allow not more than 1/10th of the total quantity in issue of eligible securities to be short sold. The shortsell list is updated every trading day, based on the shortsell list for the previous trading day, as described by the Australian Stock Exchange, in the Short Sales document [71]. For an individual trader, it is very difficult to know historically whether a short-sell transaction could actually have been executed.

All trades initiated from end-of-day data must be day+1 long market orders. This means that after a signal is given, then the trade takes place on the next day the market is open, at market open price. For example, after the market has closed on day *t*, the trading system would be run, and any buy (sell) signals generated are queued for opening positions (closing positions) for the start of the next days trading, day *t*+1. In this way, there is no possibility of acting on information which is not publicly available to all traders. In essence, this is similar to the issue of displacing fundamental data by at least 6 months, again, to ensure that the trading system is not being tested on data which was not available in the market.

All trading simulations must account for transaction costs, and it is advised that these be over-estimated for historical testing. Traditionally, the cost of brokerage for individuals has been falling, therefore, using todays transaction costs to simulate historical trading results as of 10 years ago is very misleading, particularly if the strategy being tested generates a large number of trades.

Another realistic simulation constraint is slippage. Although a trade may be initiated at market open, this does not mean the trade will be opened (closed) at market open price. There will inevitably be slippage due to the fact that at market open there may be a great many trades scheduled. Naturally, the price can move around quite considerably in the early part of trading, and slippage is the method to account for this cost. Slippage settings of 5% would be reasonable.

It is also important when developing and benchmarking systems of this type that simulations respect volume constraints. It is not realistic to assume that there is an unlimited amount of stock available for purchase. Historical technical data includes the volume data item. When training and testing, it is realistic to assume that the positions sizes acquired be some smallish factor of the overall trade volume available. A suitable factor might be 5%, or perhaps even less dependant on the market volatility.

Finally, it is unwise in historical simulations to refer directly to cut-off values for variables such as price. For example, it would be unrealistic to include a condition that price must be less than $5 to initiate a trade. Historic price data is adjusted for splits etc, therefore, historically a price may be shown as $5, but at the actual date that stock was traded in the market, it could well have been a different price.

5. BENCHMARKING

Once the appropriate in-sample architecture has been decided, the architecture and training must be frozen, and the network can proceed to out-of-sample benchmarking. At the same time, all the parameters of signal strength threshold, stop-loss threshold and money management threshold used in the in-sample testing must also be frozen.

A primary objective of a trading system is to produce (and capture) profit. However, in itself, the amount of profit obtained is an unsuitable benchmark for a variety of reasons. The desire to produce a profit must be tempered with such considerations as trading risk, equity curve management, amount of capital required, drawdown, and consistency. These factors determine how tradeable a system would be in practice.

Trading systems are typically assessed according to a variety of metrics. The metrics presented in Table 1 are sourced from Babcock Jr [72], Chande [39], Ruggiero [56], Pardo [65], Kaufman [40], Tharp [46], and Refenes [73]. Each metric is briefly discussed, and where appropriate, guidelines for desirable values are given.

It should be remembered that the factors which determine whether a system is acceptable or not are ultimately the choice of the trader. No system should be chosen if it displays undesirable characteristics; however, individual traders would differ on their choice of system, dependant on such issues as their tolerance to risk, their amount of starting capital, and their trading horizon.

Metric	**Brief Description**
Net Profit	Ending Capital – Starting Capital
Annualized Gain (%)	Annualized Net Profit (Loss). aka Annual Percentage Return (APR)
Number of Trades	Total Trades initiated by strategy
Exposure (%)	Area of portfolio equity exposed to market, as calculated on a day-by-day basis.
Winning Trades (%)	Percentage of trades that were winners.
Average Profit (%)	Average profit per winning trade, expressed as a percentage. Includes effect of trading costs, and does not take open positions into account.
Losing Trades (%)	Percentage of trades that were losers.
Average Loss (%)	Average loss per losing trade, expressed as a percentage. Includes effect of trading costs, and does not take open positions into account.
Max. Drawdown (%)	Largest peak to valley decline in the equity curve, on a closing price basis, expressed as a percentage of open equity.
Profit Factor	Gross Profit divided by Gross Loss. (Desirable systems should display over 2 for this ratio).
Recovery Factor	Absolute value of Net Profit divided by Max Drawdown. (Desirable system must display over 1 for this ratio).
Payoff Ratio	Absolute value of average profit per trade divided by average loss per trade. (Desirable system must display over 2 for this ratio).
Sharpe ratio	Sharpe Ratio measures risk adjusted return. Specifically, the ratio indicates the historic average differential return per unit of historic variability of the differential return. Sharpe [74] provides a detailed discussion of the limitations and uses of the Sharpe Ratio. It is calculated by obtaining the average percentage return of the trades generated, as well as the standard deviation of returns. The average return and average standard deviation are annualized by using the average number of days held per trade as a baseline. The annualized average return is then divided by the annualized standard deviation of returns.
Ulcer Index	Ulcer Index measures overall volatility of a portfolio. It is the square root of the sum of the squared drawdowns.
Luck Coefficient	Shows how the largest (by profit) trade compares to the average trade.
Pessimistic Rate of Return	A statistical adjustment of the wins to losses ratio for the purpose of estimating the worst expected return based on previous results. Pessimistic Rate of Return is calculated by decreasing the number of winning trades by the square root of the total winners, and increasing the number of losing trades by the square root of the number of losers. The result is then computed by multiplying the new number of winners by the average amount won, and dividing this by the new number of losers multiplied by the average amount lost.
Equity Drop Ratio	Potential for loss expressed as a probability by computing the standard deviation of all drops in the equity curve measured from each equity low to the previous equity high and dividing the result into the annualized return. Only equity drops greater than 2% are

	considered. The equity drop ratio favors higher profits, favors short term fluctuations in the equity curve, and does not penalize a system for large gains (as only equity drops are used to measure risk).

Table 1 Trading System Metrics

Further, when benchmarking a trading system, it is appropriate to perform a students t-test to determine the likelihood that the observed profitability is due to chance. This is the method recommended by Katz [75], Katz and McCormick [76], Chande [39], Stakelum [77], and Kaufman [40].

The means of the strategies developed are tested against the mean of the distribution curve that a random trading strategy would produce, which is assumed to be zero under the null hypothesis of no excess returns.

The hypotheses for the t-tests would be:
H_0: $\mu_{profit} = 0$,
H_1: $\mu_{profit} > 0$

The use of the t-test relies on assumptions of normality and independence. Essentially, these assumptions are constraints upon the usefulness of the t-test in evaluating trading systems.

Typically, the assumption of normality is dealt with by reference to the Central Limit Theorem, which indicates that as the number of cases in the sample increases, the distribution of the sample mean approaches normal. Consequently, as long as the sample size is adequate (generally stated as at least 30 samples), the statistic can be applied with reasonable assurance.

The constraint of independence presents a more difficult issue when testing trading systems. Essentially, the violation is potentially one of serial dependence, which occurs when cases constituting a sample are not statistically independent of one another. One method of dealing with this issue is to perform a runs test, as described by Vince [49]. The runs test shows whether the sequence of wins and losses in the sample trades contains more or less streaks than would ordinarily be expected in a truly random sequence, which has no dependence between trials. Although a runs test does not prove or disprove dependency, it can be used to determine an acceptable confidence limit in order to accept or reject a hypothesis of dependency. Vince demonstrates the runs test is essentially a matter of obtaining the Z scores for the win and loss streaks of systems trades, as follows:

$$\text{Z Score} = \left(\frac{(N*(R-0.5))-X}{\sqrt{\frac{(X*(X-N))}{(N-1)}}} \right) \text{ where}$$

N = total number of trades,

Equation 1 Computing the Z-Score for a Runs test

It should be noted that serial dependency is not necessarily an issue for system tradeability, and for more advanced forms of money management, it may well be desirable.

Once out-of-sample benchmarking has been completed, the trader has a realistic model of the trading system, which can then be realistically assessed. From this model, the trader can make accurate judgements about whether this particular trading system meets the trader's specific individual trading requirements.

Consistency is one of the most important areas for a trader to focus on, and the level of system consistency can be determined by comparing the figures for the in-sample model to the figures for the out-of-sample model. Clearly, the smaller the amount of variation between the two models, the greater the likelihood that the neural network has captured the generalities of the profit-generating phenomena.

An accurate model also serves another purpose; it gives the trader guidelines within which to operate. Should the trader decide to trade using this model, then it will be clear going forward whether the model is operating within the expected guidelines, and, more importantly, it will give early warning if the model unexpectedly deviates from expectations. This could happen if some underlying characteristic of the market changed, and it is important for the trader to realize this as soon as possible.

6. CONCLUSION

This paper has presented a methodology for designing and testing stockmarket trading systems using soft computing technologies, specifically artificial neural networks.

A number of models have been developed using the methodology presented in this paper. For examples of this methodology in practice, the reader may wish to pursue other papers written by the authors in this area, for example Vanstone et al. [78, 79]. These papers step through the process of selecting input variables, designing artificial neural networks for trading, and benchmarking of the trading results.

This methodology presented clearly separates the in-sample process of training neural networks and selecting parameters from the out-of-sample benchmarking process. It also aims to ensure that if the neural models developed during the in-sample training process are curve-fit, then that is clearly exposed during the out-of-sample benchmarking. This process of breaking up the development into a number of discrete, testable steps provides another advantage – it allows the developer to focus on correcting a specific part of the process if and when things go wrong.

The objective of developing viable mechanical stockmarket trading systems based on technologies such as neural networks is achievable. The key is to conduct the development process within a well-defined methodology, and as close to real-world constraints as possible.

REFERENCES

1. Tan, C.N.W., *Artificial Neural Networks: Applications in Financial Distress Prediction and Foreign Exchange Trading* (Gold Coast, QLD: Wilberto Press, 2001).
2. White, H., Economic Prediction using Neural Networks: The case of IBM Daily Stock Returns. *Second Annual IEEE Conference on Neural Networks* 1988, 451-458.
3. Oppenheimer, H.R. and G.G. Schlarbaum, Investing with Ben Graham: An Ex Ante Test of the Efficient Markets Hypothesis, *Journal of Financial and Quantitative Analysis*, XVI(3), 1981, 341-360.
4. Lowe, J., *Benjamin Graham on Value Investing* (New York: Penguin Books, 1994).
5. Bierig, R.F., The evolution of the idea of "Value Investing": From Benjamin Graham to Warren Buffett, *Duke Journal of Economics*, 12, 2000.
6. Basu, S., Investment Performance of Common Stocks in Relation to their Price-Earnings Ratios: A Test of the Efficient Market Hypothesis, *Journal of Finance*, 32(3), 1977, 663-682.
7. Banz, R.W., The Relationship between Return and Market Value of Common Stocks, *Journal of Financial Economics*, (9), 1981, 3-18.
8. Reinganum, M., Abnormal Returns in small firm portfolios, *Financial Analysts Journal*, 37, 1981, 52-56.
9. Rosenberg, B., K. Reid, and R. Lanstein, Persuasive evidence of market inefficiency, *Journal of Portfolio Management*, 11, 1984, 9-17.
10. DeBondt, W. and R. Thayler, Further Evidence of Investor Overreaction and Stock Market Seasonality, *Journal of Finance*, XLII(3), 1987, 557-581.
11. Fama, E. and K. French, The cross-section of expected stock returns, *Journal of Finance*, 47(2), 1992, 427-465.
12. Lakonishok, J., A. Shleifer, and R. Vishny, Contrarian Investment, extrapolation and Risk, *Journal of Finance*, 49(5), 1994, 1541-78.
13. Fama, E.F. and K.R. French, Size and Book-to-Market Factors in Earnings and Returns, *Journal of Finance*, 1995, 131-155.
14. Fama, E.F. and K.R. French, Value versus Growth: The International Evidence, *Journal of Finance*, 1998, 1975-1999.
15. Frankel, R. and C. Lee, Accounting Valuation, Market Expectation and Cross-Sectional Stock Returns, *Journal of Accounting and Economics*, June, 1998, 283-319.
16. Piotroski, J.D., Value Investing: The use of Historical Financial Information to Seperate Winners from Losers, *The Journal of Accounting Research*, 38 (Supplement), 2000, 1-41.
17. Kanas, A., Neural Network Linear Forecasts for Stock Returns, *International Journal of Finance and Economics*, 6, 2001, 245-254.
18. Aby, C.D., et al., Value Stocks: A Look at Benchmark Fundamentals and Company Priorities, *Journal of Deferred Compensation*, 7(1), 2001, 20-31.
19. Aby, C.D., et al., Selection of Undervalued Stock Investments for Pension Plans and Deferred Compensation Programs, *Journal of Deferred Compensation*, 6(3), 2001.
20. Nison, S., *Japanese Candlestick Charting Techniques: A Contemporary Guide to the Ancient Investment Techniques Of The Far East* (New York2001).
21. Tvede, L., *The Psychology of Finance* (Chichester: Wiley, 1999).
22. Warnecke, S.J., Hear this, Bob Prechter! A Critic attacks the Elliot Wave Theory, *Barron's*, January 26, 1987, 1987, 13.
23. Fama, E., The Behaviour of Stock Market Prices, *Journal of Business*, 1965, 34-105.
24. Taylor, M.P. and H. Allen, The Use of Technical Analysis in the Foreign Exchange Market, *Journal of International Money and Finance*, 11, 1992, 304-314.
25. Silber, W.L., *Technical Trading: when it works and when it doesn't*. Working Paper Series: Saloman Brothers Centre for the Study of Financial Institutions, Leonard N. Stern School of Business, New York University, 1993).
26. Pring, M.J., *Martin Pring's Introduction to Technical Analysis* (Singapore: McGraw-Hill, 1999).
27. Neftci, S.N. and A.J. Policano, Can Chartists outperform the Market? Market Efficiency tests for 'Technical Analysis', *Journal of Futures Markets*, 4(4), 1984, 465-478.
28. Murphy, J.J., Intermarket Analysis, *Intermarket*, 1988, 20-21.
29. Neftci, S.N., Naive Trading Rules in Financial Markets and Wiener-Kolmogorov Prediction Theory: A Study of 'Technical Analysis', *Journal of Business*, 1991, 549-571.
30. Brock, W., J. Lakonishok, and B. LeBaron, Simple Technical Trading Rules and the Stochastic Properties of Stock Returns, *Journal of Finance*, 47(5), 1992, 1731-1764.
31. Mills, T.C., Technical Analysis and the London Stock Exchange: Testing trading rules using the FT30, *International Journal of Finance and Economics*, 2, 1997, 319-331.
32. Levich, R. and L. Thomas, The significance of Technical Trading Rule Profits in the Foreign Exchange Markets: A Bootstrap Approach, *Journal of International Money and Finance*, 12, 1993, 451-474.
33. LeBaron, B., *Technical Trading Rules and Regime shifts in Foreign Exchange*, in

Advanced Trading Rules, E. Acar and S. Satchell, Editors. 1997, Butterworth Heinemann.

34. Lehmann, B.N., Fads, Martingales and Market Efficiency, *The Quarterly Journal of Economics*, CV(1), 1990, 1-28.
35. Jegadeesh, N., Evidence of Predictable Behaviour of Security Returns, *Journal of Finance*, XLV(3), 1990, 881-898.
36. Wilder, J.W., *New Concepts in Technical Trading* (Greensboro, NC: Trend Research, 1978).
37. Siegel, J.J., *Stocks for the Long Run*. 3rd ed: McGraw-Hill, 2002).
38. Pan, H., D. Sornette, and K. Kortanek, Intelligent Finance - An Emerging Direction, *Quantitative Finance*, 6(4), 2006, 273-277.
39. Chande, T.S., *Beyond Technical Analysis: how to develop and implement a winning trading system* (New York: Wiley, 1997).
40. Kaufman, P.J., *Trading Systems and Methods*. Wiley Trading Advantage (New York: Wiley, 1998).
41. Longo, J.M., *Selecting Superior Securities: Using Discriminant Analysis and Neural Networks to differentiate between 'Winner' and 'Loser' stocks*, in *UMI Dissertation Services Number 9601890*. 1996, Rutgers University. Graduate School - Newark.
42. Pocini, M., Momentum Strategies applied to Sector Indices, *Journal for the Colleagues of the International Federation of Technical Analysts*, 2004 Edition, 2004, 3-10.
43. Overholser, R., Designing a Money Management Strategy: Using Volatility-Based Stops and the Kelly Criterion, *Technical Analysis of Stocks and Commodities*, 18(5), 2000, 38-.
44. Kase, C.A., *Trading with the odds: using the power of probability to profit in the futures market* (Chicago: Irwin Professional Publishing, 1996).
45. Kase, C.A., *The Kase Dev-Stop(sm) - Accounting for Volatility, Variance and Skew in Managing Trade Risk*.
46. Tharp, V.K., *Trade your way to Financial Freedom* (NY: McGraw-Hill, 1998).
47. Sweeney, J., *Maximum Adverse Excursion: analyzing price fluctuations for trading management* (New York: J. Wiley, 1996).
48. Balsara, N.J., *Money management strategies for futures traders*. Wiley finance editions (New York: Wiley, 1992).
49. Vince, R., *Portfolio Management Formulas: mathematical trading methods for the futures, options and stock markets* (NY: John Wiley & Sons, 1990).
50. Elder, A., *Trading for a Living*: John Wiley & Sons, 1993).
51. Azoff, M.E., *Neural Network Time Series Forecasting of Financial Markets* (Chichester: Wiley, 1994).
52. Thawornwong, S. and D. Enke, The adaptive selection of financial and economic variables for use with artificial neural networks, *Neurocomputing*, 56, 2004, 205-232.
53. Ruggiero, M., Build a real neural net, *Futures*, June 1996, 1996.
54. Bauer, R.J. and J.R. Dahlquist, *Technical Market Indicators: Analysis & Performance* (New York: Wiley, 1998).
55. Slim, C. and A. Trabelsi, Neural Network for Modelling Financial Time Series: A New Approach, *Computational Science and its Applications*, 3, 2003, 236-245.
56. Ruggiero, M., *Cybernetic Trading Strategies: developing a profitable trading system with state-of-the-art technologies* (New York: Wiley, 1997).
57. Reinganum, M., *Selecting Superior Securities* (Charlottesville, Va: The Research Foundation of the Institute of Chartered Financial Analysts, 1988).
58. Balsara, N.J., K. Carlson, and N.V. Rao, Unsystematic Futures Profits with Technical Trading Rules: A Case for Flexibility, *Journal of Financial and Strategic Decisions*, 9(1), 1996, 57-66.
59. Pan, H., C. Tilakaratne, and J. Yearwood, Predicting Australian Stock Market Index Using Neural Networks Exploiting Dynamical Swings and Intermarket Influences, *Journal of Research and Practice in Information Technology*, 37(1), 2005, 43-55.
60. Raposo, R.d.C.T., A.J.d.O. Cruz, and S. Mendes, Using Frames and Multi-Agents for Stock Market Prediction Based on Fundamentalist Analysis with Fuzzy Neural Networks, *WSEAS Transactions on Systems*, 3(9), 2004, 2892-2897.
61. Chi, S.-C., et al., The Study of the Relationship among Technical Indicators and the Development of Stock Index Prediction System. *NAFIPS 2003: 22nd International Conference of the North American Fuzzy Information Processing Society* 2003, 291-296.
62. Refenes, A.-P.N., M. Azema-Barac, and A.D. Zapranis, Stock Ranking: Neural Networks Vs Multiple Linear Regression, *IEEE*, 1993.
63. Kim, K.-J. and W.B. Lee, Stock market prediction using artificial neural networks with optimal feature transformation, *Neural Computing and Applications*, 13(3), 2004, 255-260.
64. Gately, E., *Neural Networks for Financial Forecasting* (New York: Wiley, 1996).
65. Pardo, R., *Design, testing, and optimization of trading systems*. Wiley Traders Advantage Series (New York: Wiley, 1992).
66. Shih, Y., *Neuralyst Users Guide*: Cheshire Engineering Corporation, 1994).
67. Kim, J.-H., et al., Stock Price Prediction using Backpropagation Neural Network in KOSPI. *International Conference on Artificial Intelligence IC-AI'03* 2003, 200-203.
68. Jaruszewicz, M. and J. Mandziuk, One Day Prediction of NIKKEI Index Considering Information from Other Stock Markets.

International Conference on Artificial Intelligence and Soft Computing ICAISC 2004 2004.

69. Zirilli, J.S., *Financial Prediction using Neural Networks*: International Thomson Computer Press, 1997).
70. Versace, M., et al., Predicting the exchange traded fund DIA with a combination of genetic algorithms and neural networks, *Expert Systems with Applications*, 27(3), 2005, 417-425.
71. Short Sales,2004,http://www.asx.com.au/markets/l3/shortsales_am3.shtm
72. Babcock Jr., B., *The Dow Jones-Irwin Guide to Trading Systems* (Homewood, Illinois: Dow Jones-Irwin, 1989).
73. Refenes, A.-P.N., ed. *Neural Networks in the Capital Markets*. 1995, Wiley: Chichester.
74. Sharpe, W.F., The Sharpe Ratio, *The Journal of Portfolio Management*, 21(1), 1994, 49-58.
75. Katz, J.O. and D.L. McCormick, *The encyclopedia of trading strategies*. Irwin trader's edge series (New York: McGraw-Hill, 2000).
76. Katz, J.O. and D.L. McCormick, Using Statistics with Trading Systems, *Technical Analysis of Stocks and Commodities*, 15, 1997, 347-352.
77. Stakelum, J., Designing a Personal Neural Net Trading System, *Technical Analysis of Stocks and Commodities*, 13, 1995, 31-36.
78. Vanstone, B., G. Finnie, and C.N.W. Tan, Applying Fundamental Analysis and Neural Networks in the Australian Stockmarket. *International Conference on Artificial Intelligence in Science and Technology (AISAT 2004)*, Hobart, Tasmania, 2004.
79. Vanstone, B., G. Finnie, and C.N.W. Tan, Evaluating the Application of Neural Networks and Fundamental Analysis in the Australian Stockmarket. *IASTED International Conference on Computational Intelligence (CI 2005)*, Calgary, AB, Canada, 2005.