University of Cape Town

FACULTY OF COMMERCE

Analysis of the Cross-Section of Equity Returns on the JSE Securities Exchange Based on Linear and Nonlinear Modeling Techniques

BY

Kathleen E. Hodnett

Thesis prepared under the supervision of Professor Paul van Rensburg and submitted in full fulfillment of the requirements for the Degree of Doctor of Philosophy in Finance in the School of Management Studies of the Faculty of Commerce at the University of Cape Town

Cape Town, Republic of South Africa

March 2010

Declaration

Apart from the assistance which is acknowledged and the quotations which are specifically referenced in the text and bibliography section of the thesis, this thesis is entirely my own work and not being submitted for degree purposes at any other university.

Kathleen E. Hodnett

March 2010

Dedication

Dedicated to my parents, Mr. Stanley Hodnett and Mrs. Patricia Hodnett

Acknowledgement

I thank my supervisor, Prof. Paul van Rensburg, for his assistance and guidance through the completion of this thesis. His valuable comments and suggestions are humbly appreciated.

Thanks are also due to my former colleagues from the University of Natal (Durban), namely: Prof Anthony Lumby, Professor Merle Holden, Professor Nicola Viegi Professor Nicholas Wood, Professor Adrian Saville and Mr Andy Hsieh. Your endless support and enthusiasm over the years is acknowledged and appreciated. In addition to this, I would like to thank my current work colleagues Mr Ashwin Prayag and Mr Ryan Kruger from the Finance Department at the University of Cape Town for your assistance during the completion of the thesis. I am most grateful to my lecturers, Ms Elena Ilkova, Ms Marlene Hesketh and Prof Paul van Rensburg. Thank you for providing me with an excellent foundation and for introducing me to the exciting world of finance.

All of the mistakes are mine.

Abstract

This research investigates the relationship between firm-specific style attributes and the cross-section of equity returns on the JSE Securities Exchange (JSE) over the period from 1 January 1997 to 31 December 2007. Both linear and nonlinear expected returns forecasting models are constructed based on the cross-section of equity returns. A blended approach combining a linear modeling technique with a nonlinear artificial neural network technique is developed to identify future potential top performing shares on the JSE.

1. Both linear and nonlinear models identify book-value-to-price and cash flow-to-price as significant styles attributes that distinguish near-term future share returns on the JSE.

2. This thesis found updating the identity of attributes is equally important as updating the factor payoffs of attributes in applying the stepwise regression approach.

3. Nonlinearity on the JSE equity returns is found to complement the forecasting power of linear factor models.

4. In terms of artificial neural network modeling, the extended Kalman filter learning rule introduced in the thesis is found to outperform the traditional backpropagation approach.

5. This thesis found that updating the identity of attributes via a genetic algorithm in the nonlinear forecasting models is superior to the static nonlinear forecasting models.

6. Both linear and nonlinear models are found to be more adequate in identifying future outperformers than identifying future underperformers on the JSE.

The results of the research provide for potential alpha generating stock selection techniques for active portfolio managers in the South African equity market using the blended linear-nonlinear approach.

Based on the argument of Daniel and Titman (1997) that attributes themselves, rather than factor loadings on the risk portfolios proposed by Fama and French (1993), explain the cross-section of equity returns, 38 style attributes are used to build expected return factor models to forecast equity returns on the JSE. The results obtained from numerous studies, for example Banz (1981), Fama and French (1992, 1993) and Van Rensburg and Robertson (2004), to name a few, suggest that there might exist nonlinearity in explaining the cross-sectional variation in equity returns. This motivates the research to develop nonlinear models in addition to linear models to forecast JSE equity returns.

Following the methodology of Haugen and Baker (1996), in conjunction with the stepwise permutation procedure of Van Rensburg and Robertson (2003), four linear expected return factor models are constructed from the style attributes that collectively maximise the in-sample scores of the four performance measurement criteria, namely the average information coefficient (IC), the slope *t*-statistic, the Grinold information ratio (Grinold IR) and the Qian and Hua information ratio (QH IR). The out-of-sample evaluation indicates that the model that maximises the in-sample score of QH IR, with its model inputs updated every 12 months in a rolling moving-window procedure, achieves the best out-of-sample scores with robust periodic performances. By including the regression constant in the expected return factor model of Haugen and Baker (1996), the out-of-sample performances of the model with the regression constant is found to consistently outperform the counterpart models without the regression constants. It is also found that the dynamic updating of the identities of the model inputs in the moving-window procedure is essential in maintaining the robustness of the models' forecasting power over time.

The nonlinear expected return factor models are constructed using the cascade correlation artificial neural network (ANN) architecture, with or without direct connection from the input layer to the output layer in the neural network architecture. The inputs of the ANN models are selected using a genetic algorithm with or without the cascade variable pre-selection technique. In addition, the models are trained using either the embedded backpropagation learning rule or the embedded extended Kalman filter (EKF) learning rule within the cascade correlation architecture. Results of the study indicate that while the cascade variable pre-selection technique is essential in improving the forecasting power of the backpropagation models, no significant relationship is established between the pre-selection technique and the forecasting power of the Kalman models. The

consolidated results support the view of Qi and Maddala (1998) that removing the direct connection constraint can improve model performances for the backpropagation models. However, the empirical relationship between the direct connection constraint and the performances of the Kalman models remains inconclusive. The Kalman models in general outperform their counterpart backpropagation models over the examination period.

The best linear model (QH IR model) and the best nonlinear model (the ANN Kalman model) are subsequently employed to develop a blended stock selection model over the out-of-sample period from 1 January 2002 to 31 December 2007. The blended model assigns weights to the expected returns of sample shares forecasted by the respective best linear and nonlinear models in an optimisation procedure that maximises the in-sample QH IR (computed as average IC divided by the standard deviation of ICs) of the blended model in the moving-window procedure. The blended model is found to allocate higher weights to the ANN Kalman model in 4 of the 5 sub-periods, and it achieves higher average IC score with median-to-low standard deviation of ICs compared to the constituent models over the out-of-sample period. When the performances of the stocks selected by the blended model are compared to that of the constituent models, it is found that the top quintile of stocks selected by the blended model outperform the constituent models on a risk-adjusted basis in the out-of-sample period. Although all three models successfully identify true winning stocks for the top quintile of stocks, the bottom quintile of stocks chosen by all three models do not underperform the market proxy statistically. Overall, the blended model effectively diversifies the forecasting errors of the constituent models, and hence is the best stock-selection model identified in this research over the examination period.

Table of Contents

List of Figures and Tables

LIST OF FIGURES

LIST OF TABLES

1

Introduction

"Actually, Fama and French use the logarithm of size and the logarithm of the book-to-market ratio as independent variables in their cross-sectional regression models, thus assuming a nonlinear relationship between the expected rate of return and the untransformed values of the independent variables. They use this specific form of a nonlinear relationship without giving a specific reason and without discussing alternative nonlinear functions".

- Stehle, Bunke and Sommerfeld (1997: 2)

1.1 Research Background

Modern portfolio theory of Markowitz (1952) and the separation theorem of Tobin (1958) describe the asset allocation decision of rational investors in an efficient capital market. The asset pricing relationship depicted by the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), along with the efficient market hypothesis (EMH) of Fama (1970, 1991), form the joint hypothesis for tests of market efficiency. The hypothesis states that firm-specific information is reflected in the share price in an efficient capital market, and firm-specific risk is diversifiable in a large portfolio. As a result, the only relevant risk is market risk, measured by an asset's beta coefficient against the movements of the market portfolio.

Tests of the joint hypothesis reveal several anomalies relating to the use of firm-specific information to earn abnormal returns above the risk-adjusted returns suggested by the CAPM. The CAPM-related anomalies include the value effect of Basu (1977), the small firm effect of Banz (1981), the long-term price reversals of De Bondt and Thaler (1985, 1987) and the momentum effect of Jegadeesh and Titman (1993). Possible explanations for these anomalies include methodological bias in empirical research, misspecification of the CAPM pricing relationship and the argument of investor irrationality from the behavioural perspective.

Further to this, the existence of nonlinearity in the pricing relationship based on firm-specific attributes is observed in studies conducted by Banz (1981), Fama and French (1992, 1993), Van Rensburg and Robertson (2004), Eakins, Stansell and Buck (1998, 2003), Cao, Leggio and Schniederjans (2005) and Cao, Parry and Leggio (2009).

The two main branches of studies that attempt to explain the CAPM-related anomalies using factor models include the use of factor mimicking portfolios to represent priced risks proposed by Fama and French (1993) and the characteristic approach of Daniel and Titman (1997). While Fama and French (1993) use the size and value factor mimicking portfolios to explain the residual returns of the CAPM, Daniel and Titman (1997) argue that it is the style attributes themselves that explain the cross-section of equity returns. The argument of Daniel and Titman (1997) is supported by local studies conducted by Van Rensburg and Robertson (2003a) and Van Rensburg and Robertson (2004) on the JSE Securities Exchange (JSE).

1.2 Overview

This research undertakes to examine the pricing relationship between firm-specific style attributes and the cross-section of equity returns on the JSE over the period from 01 January 1997 to 31 December 2007. The research attempts to determine whether firm-specific style attributes can be successfully used as inputs to identify superior performers on the JSE by integrating linear and nonlinear modeling techniques based on apparent market anomalies. Based on the equity pricing relationship observed in the research, expected return linear and nonlinear factor models are constructed to forecast forward equity returns. Selecting and ranking shares based on their forecasted returns enable direct comparison to be drawn between the stock-selection abilities of the alternative models in a fractile analysis. Based on mean-variance efficient optimisation using the forecasting accuracy and the volatility of accurate forecasts from the best linear model and the best nonlinear model, the final outcome of this research is a blended stock-selection model that maintains satisfactory forecasting accuracy with diversified forecasting errors.

Chapter 2 presents the theoretical framework of the research, which includes modern portfolio theory of Markowitz (1952), separation theory of Tobin (1958), the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965), the arbitrage pricing theory (APT) of Ross (1976), the efficient market hypothesis (EMH) of Fama (1970, 1991) and behavioural finance. With regard to nonlinear modeling techniques, this chapter provides a discussion of artificial neural networks (ANN). An artificial neural network, which mimics the functioning of a biological neural network and the human brain, consists of interconnected artificial neurons operating in parallel. ANNs have the ability to model complex, nonlinear relationships in data. The traditional fixed-architecture ANN model is presented. Supervised learning algorithms, namely, the backpropagation algorithm (Werbos (1974) and Rumelhart and McClelland (1986)) and the extended Kalman Filter (EKF), an extension of Kalman filter theory (Kalman, 1960), is introduced in this chapter. These learning algorithms form the basis for empirical work in Chapter 7 and Chapter 8.

Empirical tests of the joint hypothesis and studies that document capital market anomalies such as Basu (1977), Banz (1981), De Bondt and Thaler (1985, 1987), Jegadeesh and Titman

(1993) and Lakonishok, Shleifer and Vishny (1994) are presented in Chapter 3. In addition, studies that aim at explaining empirical anomalies using linear factor models such as Fama and French (1992, 1993, 1995, 1996 and 1998) and Jegadeesh and Titman (1993) and studies that test the robustness of linear factor models constructed from firm-specific style attributes such as Haugen and Baker (1996, 2009) and Hanna and Ready (2005) are also discussed in this chapter. With regard to literature regarding nonlinearity, Eakins, Stansell and Buck (1998, 2003), Cao, Leggio and Schniederjans (2005) and Cao, Parry and Leggio (2009) employ artificial neural networks (ANN) to forecast equity returns from a set of firm-specific attributes and macroeconomic variables. Significant studies conducted on the JSE such as Page and Way (1992/1993), Muller (1999), Achour, Harvey, Hopkins and Lang (1999), Fraser and Page (2000), Van Rensburg (2001), Van Rensburg and Robertson (2003a, 2003b, 2004) and Frankish (2004) are included in Chapter 3. Significant prior literature motivates this research to develop linear and nonlinear expected return factor models using firm-specific attributes to explore the apparent market anomalies on the JSE. Chapter 4 presents the problem statement, research objectives and methodological overview of the research. Potential research biases are also discussed in Chapter 4.

There are in total 38 candidate firm-specific style attributes selected from the fundamental values relative to share price category, solvency and liquidity category, fundamental growth category, operating performance category, size and return momentum category and consensus analyst forecast category. Univariate tests that determine the significance of the mean and the sign of the factor payoff to each attribute is demonstrated in Chapter 5. The linear factor models are constructed based on the expected return factor model of Haugen and Baker (1996), in conjunction with the stepwise permutation procedure of Van Rensburg and Robertson (2003a). The details of the in- and out-of-sample results of the linear factor models developed in this research are discussed in Chapter 6. On the other hand, the nonlinear expected return factor models are constructed within the cascade-correlation artificial neural network (ANN) architecture. Models are trained using either the embedded backpropagation learning rule or the embedded extended Kalman filter (EKF) learning rule within this cascaded-correlation architecture. The results of the ANN backpropagation models and the results of the ANN Kalman models are analysed in Chapter 7 and Chapter 8 respectively.

Chapter 9 consolidates the performance scores of the best linear expected return factor model identified in Chapter 6, the best ANN backpropagation model identified in Chapter 7 and the best ANN Kalman model identified in Chapter 8. The analysis of the consolidated results enables the research to develop a blended stock-selection model from the best linear and the best nonlinear model. The performance statistics of the blended model and its constituent models are compared and evaluated in Chapter 9. Chapter 10 summarises significant findings of the research and provides solutions and recommendations to the research problem.

1.3 Contributions

The extensive examination period from 01 January 1997 to 31 December 2007, to author's knowledge at the time of writing, is one of the lengthiest and latest studies in the area of asset pricing on the JSE to date. The factor models are built from a total of 38 firm-specific attributes, which is the largest number of firm-specific attributes that has been tested in the South African equity market. Major contributions by the research are as follows:

1. In building the linear expected return factor models, Haugen and Baker (1996) place a zero constraint on the regression constant . Incorporating the constant term in the expected return forecasting process can potentially improve the robustness of the expected return factor models. The linear factor models are constructed using both methods (with and without using the regression constant). No such comparison has been done in empirical research.

2. Frankish (2004) adopts the stepwise permutation approach of Van Rensburg and Robertson (2003a) to select inputs for the expected return factor model of Haugen and Baker (1996). Contrast has been drawn by Frankish (2004) to demonstrate the advantage of using the optimal set of significant attributes over using the entire theoretically justified set of attributes. However, the identities of the inputs employed by Frankish (2004) are determined within a static in-sample period. This method does not account for the possibility of time-varying attributes.

   This research rectifies this problem by training and updating the identities of model inputs every 12 months using the dynamic moving-window (rolling) approach for both linear and ANN models. The results of the rolling models are deemed to be more robust relative to the results of the static models. The results of the rolling models are compared to that of the static models, which provide useful insights into the practical development of expected return factor models.

3. International research on nonlinear asset pricing using ANN modeling is limited to the fixed-architecture network which requires a pre-determined number of hidden neurons

and layers. This research represents the first intensive study in the area of nonlinear asset pricing in the local equity market (JSE). In addition, this research represents the first research that adopts the cascade-correlation ANN rather than the fixed architecture in building expected return nonlinear factor models.

4. The backpropagation learning rule is the most widely employed learning algorithm in ANN modeling to date. This research employs backpropagation as an embedded learning algorithm, within the cascade-correlation architecture. This research also employs the extended Kalman filter (EKF) learning algorithm as an alternative to backpropagation. The ANN models trained under the extended Kalman filter learning algorithm outperform the ANN models trained under the backpropagation learning algorithm, on a risk-adjusted basis. This provides a motivation for the use of the EKF learning algorithm in training financial data. To the author's knowledge, this study is the first in the area of asset pricing to test the use of the extended Kalman filter to train neural networks.

5. Past use of genetic algorithms includes the employment of genetic algorithms as an alternative to backpropagation for adjusting the network weights (training) of the ANN models. This research employs a genetic algorithm to select the best variables to be used as inputs in the neural network. To the author's knowledge, this study is the first to employ genetic algorithms as a variable selection technique in neural network training on the JSE.

6. This research presents the first blended model that performs asset allocations based on the stocks ranked and selected by two separate expected return factor models, using the concept of mean-variance optimisation of the forecasts by the constituent models. This is also the first attempt to investigate the possibility of integrating the linear expected return factor model with the ANN expected return factor model.

The outcomes of this research contribute to the existing body of knowledge in the field of efficient asset pricing and stock selection techniques within the South African equity market.

10

Conclusion

The efficient market hypothesis (EMH) states that share prices reflect all firm-specific information relevant for price formation. Sharpe (1964) and Lintner (1965) separately develop the capital asset pricing model (CAPM) that linearly depicts the asset pricing relationships using the security market line (SML). In support of the EMH, the pricing relationship of the CAPM suggests that firm-specific information of individual shares is irrelevant for price formation through diversification in a large portfolio. As a result, the EMH and the CAPM jointly imply that asset returns are proportional to their covariance with the market portfolio, which is systematic in nature and measured by their respective beta coefficients. Empirical evidence on capital market pricing anomalies suggest that investors prefer to hold shares that possess strong tendencies towards certain firm-specific attributes, which results in the differences in the cross-section of equity returns. In addition, results of empirical studies (Banz (1981), Fama and French (1992, 1993) and Van Rensburg and Robertson (2004), to name a few) suggest that the influences of these anomalies on the variations of equity returns might not be linear. While Fama and French (1993) claim that the market anomalies are captured by the size and value factor mimicking portfolios, Daniel and Titman (1997) propose the use of style attributes themselves to explain the residual returns of the market model. Empirical investigation of the two approaches on the JSE by Van Rensburg and Robertson (2003a) and Van Rensburg and Robertson (2004) reveal that market capitalisation and price-to-earnings ratio rather than the return covariance of the factor mimicking portfolios better explain the cross-section of equity returns on the JSE.

This research investigates candidate factors that contribute to the variations in the cross-section of equity returns on the JSE over the period from 01 January 1997 to 31 December 2007. By employing linear and nonlinear modeling techniques, the relative abilities of various combinations of firm-specific style attributes in forecasting equity returns are evaluated.

The 38 style attributes examined by this research are divided into six categories, namely the fundamental values relative to share price, solvency and liquidity, fundamental growth,

operating performance, size and return momentum and consensus analyst forecast categories. The results of the univariate test indicate that the majority of the attributes in the fundamental value relative to share price category, the fundamental growth category, the size and return momentum category and the consensus analyst forecast category are able to explain the cross-sectional equity returns in their individual capacities over the period from 01 January 1997 to 31 December 2007. Style attributes that have significant factor payoffs are also associated with consistent signs with their factor payoffs over the examination period. Firms that have relatively high fundamental values relative to their share prices, lower financial leverage, higher fundamental growth, higher short-term return momentum, lower share prices, smaller market capitalisation and higher consensus forecast regarding their future earnings are rewarded positively.

Following the methodology of Haugen and Baker (1996), in conjunction with the stepwise permutation procedure of Van Rensburg and Robertson (2003), four linear expected return factor models are constructed from the style attributes that collectively maximise the in-sample scores of the four performance measurement criteria, namely the average information coefficient (IC), the slope *t*-statistic, the Grinold information ratio (Grinold IR) and the Qian and Hua information ratio (QH IR). Examining the identities of the model inputs being included by the respective models in the stepwise permutation procedure reveals that most of the selected style attributes have significant *t*-statistics for the time-series means of their univariate factor payoffs. This result supports the use of the *t*-statistic for the mean factor payoff as the initial screening process in identifying potential significant model inputs for the expected return factor models. Study results also indicate that problems of overfitting can be mitigated by eliminating the attributes included in the latter iteration when two or more attributes have correlation coefficient of greater than an absolute value of 0.70. In doing so, the robustness of the models improves with their respective out-of-sample performance scores. It is also found that models built with the use of the regression constant are more robust compared to the original Haugen and Baker (1996) model that places a zero constant constraint in modeling equity returns.

Comparing the performance statistics of the rolling linear factor models built within the dynamic rolling window procedure that updates model inputs every 12 months to the

performance statistics of their static counterparts, the results indicate that the rolling linear factor models are more robust and maintain their out of sample scores over time. The static linear factor models' out-of-sample scores deteriorate when the out-of-sample period moves further away from the training period, with the exception of the slope *t*-statistic model. The observed time-varying nature of the identities of the model inputs indicates that updating the estimates of the factor payoffs alone is not sufficient in providing accurate model forecasts. In addition, more-than-one attribute from each category are frequently included in the factor models at times. This observation is in support of the characteristic approach of Daniel and Titman (1997) that attributes themselves better explain the variations of the equity returns rather than the use of factor mimicking portfolios that represent the main risk factors in the market. It is also found that some models are more robust than others in terms of the consistency of their in- and out-of-sample performance scores. Thus, portfolio managers should focus on identifying the best expected return factor model that outperforms the competing models over time.

In order to explore nonlinearities in the cross-section of equity returns, nonlinear expected return factor models are constructed using the cascade-correlation artificial neural network (ANN) architecture. Models are trained using either the embedded backpropagation learning rule or the embedded extended Kalman filter (EKF) learning rule within the cascade correlation architecture. The nonlinear models use a genetic algorithm (used as the standard variable selection technique) to select the best subset of input variables to be included by the respective nonlinear models for training the artificial neural network. In addition to this standard variable selection technique, the ability of the cascade variable selection (which employs a pre-selection of input variables before the standard variable selection) is explored in this research. The study results reveal that book value-to-price (BVTP) and cash flow-to-price (CFTP) are the most important attributes selected by both the backpropagation models and the Kalman models. When the performance results of the ANN models built using the cascade pre-selection technique are compared to the performance results of the ANN models built using the standard variable selection technique, it is found that the cascade pre-selection technique effectively narrows down the number of model inputs for the backpropagation models, but not for the Kalman models. The trimming of the model inputs for the backpropagation models effectively improve the robustness of the backpropagation models

built under the cascaded pre-selection procedure over their standard variable selection counterparts. Due to the fact that a smaller set of variables are selected for the Kalman models in general, the cascade pre-selection technique does not provide additional benefits for the Kalman models. The dominant attributes for both the backpropagation models and the Kalman models are the book value-to-price ratio, and the cash flow-to-price ratio.

When the pool of the variables from which the ANN models select their inputs is narrowed down to those variables with greater than 1.0 for the *t*-statistics of their mean factor payoffs, both the backpropagation models and the Kalman models that select inputs from the reduced number of variables underperform their counterpart models that select inputs from the entire pool of all 38 attributes. The underperformance of the Kalman models with reduced pool of variables, are however, less significant than the backpropagation models with reduced pool of variables.

With regard to the direct connection constraint from the input layer to the output layer of the ANN models, the periodic performance results do not show that the ANN models (both backpropagation models and Kalman models) with their direct connection constraint removed improve their forecasting abilities. Although the consolidated in- and out-of-sample results is in support of Qi and (1998) that removing the direct connection constraint can effectively improve the forecasting accuracy of the backpropagation models, the empirical relationship between the direct connection constraint and the performances of the Kalman models remains inconclusive over the examination period. Contrary to the findings of the linear expected return factor models, there is no evidence suggesting that the rolling ANN models are more robust than their static counterparts.

The analysis of the consolidated in- and out-of-sample performance scores of the linear factor models reveals that the rolling QH IR model is the best linear expected return factor model for forecasting the JSE equity returns in this research. On the other hand, the best backpropagation model is the rolling backpropagation model that uses the cascade pre-selection technique in selecting variables with no direct connection between its input and the output layers. With regard to the Kalman model, the best Kalman model is the rolling Kalman model that uses the standard variable selection technique with direct connection between the

input and the output layers. Although the best backpropagation model achieves the highest in-sample scores for all performance measurement criteria except the QH IR score, its out-of-sample scores are the lowest amongst the best models. Overall, the best linear model achieves the best out-of-sample scores for all criteria and hence is the best expected return factor model constructed in this research.

The best linear model (rolling QH IR) and the best nonlinear model (the best Kalman model) are used as the constituent models to build the blended model that performs asset allocations based on the stocks ranked and selected by two separate expected return factor models. Based on the concept of mean-variance efficiency, the blended model would achieve an average IC score that approximates the weighted average IC score of the constituent models with its standard deviation of ICs less than the weighted average of the standard deviations of the constituent models. The QH IR optimisation procedure allocates greater weight to the best linear factor model relative to the weight allocated to the best nonlinear model. The examination of the periodic QH IR scores achieved by the blended model obtains higher average IC score with median-to-low level of standard deviation of ICs. The consolidated out-of-sample performance scores also show that the blended model has the highest out-of-sample forecasting ability per unit of the volatility of forecasting error compared to the constituent models. In terms of the fractile analysis over the out-of-sample period, the blended model and the constituent models successfully identify the future top performers for their top quintiles over the out-of-sample period from 01 January 2002 to 31 December 2007. By contrast, the bottom quintiles of the three models fail to underperform the market proxy after 2005. As a result, the log cumulative spreads between the top quintile returns and the bottom quintile returns for all three models are distinctively flatter over the second half of the out-of-sample period. Overall, the blended model effectively diversifies the volatility of the forecasting errors of its constituent models while maintaining satisfactory returns, and hence is the best stock-selection model constructed in this research.

In conclusion, although the best linear model achieves the most consistent outperformance over the examination period in its individual capacity, the fact that the blended model is the most mean-variance efficient model over the examination period serves as evidence that there exists nonlinearity in the cross-section equity returns. Thus, the exploration of nonlinearity in

the pricing relationship between style attributes and equity returns aids to the construction of a more mean-variance efficient portfolio than using a pure linear expected return factor model. In an attempt to explore nonlinearity in the cross-sectional equity returns, this research has adopted the cascade-correlation ANN architecture rather than the traditional method that is based on a fixed-architecture neural network with a predetermined number of hidden neurons and layers. Limited to the intended purpose and the scope of this research, no comparison is drawn between the cascade correlation ANN and the fixed architecture ANN, which is a potential area of future research. Other specific details addressed by this research include the effectiveness of the cascade variable pre-selection technique and the relaxation of the direct connection constraint in enhancing the robustness of the ANN expected return factor models. Other methods of improving ANN modeling such as determining the optimal learning rule for the problem presented to the network, type of neural network (feed-forward, recurrent, probabilistic or decoupled neural network, to name a few), variable selection, learning rate adjustments, and advanced methods of avoiding ANN modeling overfitting demand attention for future research.

2

Theoretical Overview

2.1 Introduction

This chapter provides a theoretical overview of the research. This includes an overview of market efficiency and the different forms of the efficient market hypothesis, modern portfolio theory, the development of the capital asset pricing model (CAPM), the arbitrage pricing theory (APT), and behavioural finance. Artificial neural networks, a nonlinear modeling technique are discussed as an alternative to the linear modeling techniques.

According to the efficient market hypothesis the value of a company is fully reflected in its share price. Thus informational efficiency implies that a security's price instantaneously adjusts to reflect new information as it arrives in the market. Markowitz (1952) introduced modern portfolio theory (MPT), where rational investors choose optimal investments positioned on the efficient frontier, depending on their degree of risk-aversion. Logically extending the MPT concept, Sharpe (1964) and Lintner (1965) independently developed the capital asset pricing model (CAPM) which postulates a linear relationship between the expected rate of return on an investment and the level of systematic (non-diversifiable) risk. A major flaw of the CAPM is presented by Roll (1977) who argues that the CAPM is not a testable model due to the unobservability of the market portfolio. A multifactor model is presented by Ross (1976) who presents the arbitrage pricing theory as an alternative to the single-factor CAPM. Like the CAPM, the APT assumes a linear pricing relationship between returns and the explanatory variables.

Behavioural finance offers an alternative paradigm to the efficient market hypothesis. By relaxing the behavioural assumptions of the efficient market hypothesis, advocates of behavioural finance argue that, due to psychological biases, investors make irrational forecasts which impact on the pricing of securities in the market.

Relaxing the linear relationship between returns and the explanatory variables and allowing for a nonlinear relationship to exist, allows for the testing of the existence of nonlinearities in asset pricing to be explored. Artificial neural networks, which are abstractions of biological neural networks, are designed to imitate the functioning of the human brain and the peripheral nerves. ANNs are composed of processing elements operating in parallel, which have the ability to map any arbitrary nonlinear function. The backpropagation learning algorithm is a gradient descent mechanism of updating the weights of the network, until the network error is minimised. The extended Kalman filter, based on state-space theory, is an alternative learning algorithm designed to capture nonlinearities.

2.2 Efficient Market Hypothesis

In an efficient market, security prices fully reflect all available information concerning the market and security prices in an unbiased fashion. Thus, informational efficiency suggests that as new information about a particular security enters the market, that security's price will rapidly adjust to reflect this new information. An implication of this is that an investor will not earn a return in excess of the risk-adjusted rate of return and thus the only way to earn a higher return is to incur more risk. In studying the movement of security prices Kendall (1953) suggests that stock and commodity prices follow a random walk and concludes *"in a series of prices which are observed at fairly close intervals the random changes from one term to the next are so large as to swamp any systematic effect which may be present. The data behave almost like wandering series"* (Kendall, 1953:11). Thus, if the future path of a security is random, the past price of a security cannot be used to predict the future price. Fama (1965) tests the empirical validity of the random-walk model and finds evidence in support of this model. It is important to note that the random-walk model does not imply the erratic movement of stock prices. In fact it emphasises that because the market is so efficient, and because prices adjust instantaneously, it is not possible for buyers and sellers to react quick enough to benefit (Malkiel, 1990).

Although Samuelson (1965) is the first to formally publish empirical proof for a version of the efficient market hypothesis (EMH), it is Fama (1970) who provides both the theoretical and empirical account of the efficient market hypothesis. Based on the information structure, Fama (1970) divides the EMH into three forms: weak form, semi-strong form and strong form. Fama (1970: 388) explains, *"the categorization of the tests into weak, semi-strong and strong form will serve the useful purpose of allowing us to pinpoint the level of information at which the hypothesis breaks down"*. The *weak form* version of the EMH asserts that it is not possible to earn abnormal returns in the market using past information, since security price movements are random (random walk model). A security price change at time *t* and time *t+1* are uncorrelated. Thus technical analysis, the practice of using past prices to predict future prices (and based on the premise that "history repeats itself") is deemed fruitless. The *semi-strong form* version of the EMH posits that it is not possible to earn abnormal returns in the market based on public information, since all public information (published financial

statements, stock splits, company announcements, analyst's reports, to name a few) is already reflected in the share price. Thus fundamental analysis, the attempt to find "undervalued" shares by analysing company attribute values and earning's report, is rejected under this form of efficiency. Early tests of the semi-strong form efficiency includes tests conducted by Ball and Brown (1968), Fama, Fischer, Jensen and Roll (1969), Ibbotson (1975) and Jensen and Ruback (1983). On the other hand, the *strong form* version of the EMH suggests that, in addition to past and public information, it is not possible for an investor to earn abnormal risk-adjusted returns in the market by trading on private information, particularly inside information (insider trading).

The empirical tests of Fama (1970) are largely a motivation for the random walk model and whether security prices 'fully reflect' all available information. Fama (1970: 388) concludes, *"…the market may be efficient if "sufficient numbers" have ready access to available information. And disagreement among investors about the implications of given information does not in itself imply market inefficiency unless there are investors who can consistently make better evaluations of available information than are implicit in market prices"*.

Twenty one years later, in a follow-up article, Fama (1991) revised his original definitions of the three categories of market efficiency to *"tests of return predictability"*, *"event studies"* and *"tests for private information"*, respectively, to which he asserts, *"the cleanest evidence of market efficiency comes from event studies….event studies can give a clear picture of the speed of adjustment of prices to information"* (Fama, 1991: 1607).

2.3 Modern Portfolio Theory and Asset Pricing

The foundation of asset pricing is laid down by Harry Markowitz (1952) who introduces the first portfolio model that incorporates both portfolio risk *(portfolio variance or standard deviation of expected returns)* and return *(portfolio mean)*. Under this model, investors are assumed to be rational, wealth-maximising and risk-averse, where more wealth (returns) is preferred to less wealth (returns) and less risk is preferred to more risk. Thus if two investments are identical in all respects, yielding identical returns, but different risks, investors will choose the less risky investment. The only way an investor will be prepared to take on additional risk is if the investment offers a higher return. A curve representing the best risk-return combination portfolios, that is, the set of portfolios with the highest mean at each given level of variance, is called an efficient frontier.

The extent to which a risk-averse investor decides to take on more or less risk is dependent on that investor's attitude towards risk/degree of risk-aversion. An investor maximises his/her utility by choosing a point on the efficient frontier. The investor's attitude towards risk can be represented by examining his/her trade-off between risk and return, also known as the investor's utility curve. Markowitz (1952) derives the efficient frontier of risky assets by incorporating the effect of diversification into the portfolio concept. For a set of risky assets, input variables to the derivation include expected return estimation, variance/standard deviation of returns and pairwise covariance/correlation. Portfolio return is computed as a weighted average of the individual constituent assets' returns while variance is a function of the correlation of the constituent assets. Risk reduction takes place by holding a combination of assets which are not perfectly correlated. The inclusion of a risk-free asset allows the investor to split their investment between holding a combination of the optimal risky portfolio (market portfolio) identified by Markowitz and the risk free asset. A more risk-averse investor will tilt his investment towards holding more of the risk-free asset, while a less risk-averse investor will tilt his investment towards investing in more of the market portfolio. A combination of the market portfolio with the risk-free asset is termed the *capital market line*, which now represents the new efficient frontier.

Tobin's (1958) separation theorem divides the investment process into two separate steps

taken by the investor. The first step entails finding the efficient portfolio of risky assets. This is based on extrinsic factors (computing mean, variance and covariance of returns). The second step entails combining the efficient portfolio with a risk-free asset. This is based on intrinsic factors (investor's degree of risk-aversion).

The capital asset pricing model (CAPM) developed by Sharpe (1964) and Lintner (1965) extends Markowitz's portfolio theory in an attempt to simplify the portfolio decision faced by the investor. An investor is exposed to two types of risk, namely unsystematic (diversifiable) risk, and systematic (undiversifiable) risk. Unsystematic risk is firm-specific risk, which can be reduced/eliminated through the process of diversification. Systematic risk, on the other hand represents portfolio/market risk common to all assets, which cannot be eliminated by diversification. Since firm-specific risk can be diversified away in a large portfolio, according to Sharpe (1964) and Lintner (1965), systematic risk is the only risk that is relevant to this portfolio. Where Markowitz examines the relationship between expected return and total risk (variance) of a portfolio, the capital asset pricing model examines the relationship between expected return and beta of a security (or portfolio), where beta is a measure of systematic risk. The CAPM linear risk-return relationship is represented as follows.

$$E(R_i) = R_f + \beta_i\left[E(R_m) - R_f\right] \tag{2.1}$$

Where:

$E(R_i)$ = expected rate of return of asset *i*;

R_f = the risk-free rate, that is, the rate of return on a theoretical zero-risk asset or portfolio;

$\left[E(R_m) - R_f\right]$ = the market risk premium;

β_i = the measure of systematic risk which measures the tendency of return asset i to co-vary with the return of the market portfolio.

Thus, examining Equation 2.1, the CAPM implies that expected return is an increasing linear function of its tendency to co-vary with the market portfolio. Beta is represented as follows:

$$\beta_i = \frac{\sigma_{i,m}}{\sigma_m^2} \quad (2.2)$$

Where:

σ_m^2 = variance of the return on the market portfolio;

$\sigma_{i,m}$ = the covariance of asset i's return with the market return, and can be defined as:

$\rho_{im} \times \sigma_i \times \sigma_m$

Where:

ρ_{im} = the correlation coefficient, a measure of the correlation of the returns of i with the returns of m; and

σ = the standard deviation of returns.

Thus, based on the CAPM, all systematic risk factors are captured by the market movement. As a result, market risk is the only relevant risk that investors require compensation for. To date, the CAPM is regarded as the most generally accepted equilibrium pricing model for pricing assets in the capital market. Due to the fact that asset pricing models provide benchmarks for the intrinsic values of assets in the capital market, empirical tests on asset pricing models are inevitably tests of market efficiency. This is commonly known as the joint hypothesis. Empirical tests of the CAPM often raise questions regarding the stability of the beta coefficient, the linearity of the security market line (SML) and whether the market risk premium is the sole representation of risk in the capital market.

In the belief that the market is not the only factor in the return generating process, Ross (1976) offers an alternative to the CAPM by introducing a multifactor equilibrium pricing model known as the arbitrage pricing theory (APT). The APT gained significant support after Roll's critique on the testability of the CAPM due to the unobservability of the true market portfolio in 1977 and 1978, which led to the benchmark error in asset pricing.

The APT is a multifactor asset pricing model that is developed based on less restrictive assumptions than the CAPM: it does not require the observation of the true market portfolio; and it does not rely on the entire investment community to restore market equilibrium, but

assumes that few investors who identify the riskless arbitrage opportunity to take as large a position as possible to restore the market equilibrium. Equation 2.3 depicts the equilibrium pricing relationship under the APT:

$$E(R_j) = R_f + b_{j1}\lambda_1 + b_{j2}\lambda_2 + \cdots + b_{jn}\lambda_n \quad (2.3)$$

Where:

$E(R_j)$	=	expected rate of return of asset *j*;
R_f	=	the risk-free rate, that is, the rate of return on a theoretical zero-risk asset or portfolio;
b_{jk}	=	each b_{jk} coefficient represents the sensitivity of asset *j* to risk factor *k*; and
λ_k	=	represents the risk premium on factor *k*.

Although the APT proposes the use of multiple risk factors to determine asset prices, it does not specify what these factors are or how they should be determined. The empirical study conducted by Chen, Roll and Ross (1986) identify the unanticipated movements in industrial production, inflation, yield spread between the low grade bond and the government bond and the slope of the term structure of interest rates as possible APT factors in the U.S. equity market over the period from 1953 to 1983. These factors are proven to be robust in international research. Van Rensburg (1997) suggests that any variable that impacts on the magnitude of the explanatory variables of the Gordon-Shapiro constant growth dividend discount model, that is, current dividend payments (D_0), market capitalisation rate/discount rate (k), and the expected growth rate of dividends (g) is instrumental in explaining prices and as a result deserves consideration as 'candidate' APT factors.

2.4 Behavioural Finance

A body of research has challenged the efficient market hypothesis (EMH) both empirically and theoretically. Behavioural finance, "*the study of how investors systematically make*

errors in judgments, or mental mistakes" (Fuller, 2000: 2), offers an alternative paradigm to the EMH. Rather than being based on the assertion that investors act in a rational manner and take into consideration all available information when making investment decisions, behavioural finance attempts to focus on the extent to which human emotion and psychological biases influences the decision-making process and actions of investors. The major development of behavioural finance stems from prospect theory of Kahneman and Tversky (1979), which presents arguments against the basic tenets of the expectation theory. While the expectation theory utilises the function of diminishing marginal utility to emphasise the risk aversion of investors, prospect theory specifically indicates that investors are risk-averse for gains but exhibit diminishing marginal disutility for losses. The concept of loss aversion of prospect theory further indicates that the extent of disutility derived from losses is larger than the utility derived from an equal magnitude of gains. This proposition suggests that investors prefer to avoid making losses than to acquiring gains. Loss aversion results from seeking pride and avoiding regret and often leads investors to hold losers too long and sell winners too soon. This is commonly known as the disposition effect.

Behavioural biases lead investors to violate the assumptions underlying modern portfolio theory and not to make their decisions based on the mean, variance and covariance of asset returns. By examining the behaviour of investors, analysts and portfolio managers, researchers have identified and examined biases that impact on investment performance. For example, De Bondt and Thaler (1985) test the overreaction hypothesis where they assert that irrational investors persistently overweight recent information and underweight long-term fundamental information. As a result of this overreaction, the contrarian and momentum asset pricing anomalies arise. This over- or understated information is expected to correct to its long-term value, given that the fundamentals remain the same.

Other pricing anomalies relating to firm-specific attributes such as the size effect and the value effect are not explained by the CAPM of Sharpe (1964) and Lintner (1965), but possibly can be explained by a behavioural phenomenon. This will be explored further in Chapter 3.

2.5 Artificial Neural Networks

The capital asset pricing model (CAPM) of Sharpe (1964) and Linter (1965) and the Ross (1976) arbitrage pricing theory (APT) both assume a linear pricing relationship between returns and the respective explanatory variables. By relaxing the linearity assumption, an artificial neural network (ANN), a nonlinear modeling technique, has the potential to improve forecasting accuracy. With its roots in computational neuroscience and cognitive psychology, ANNs are designed to mimic the functioning of the human brain. Like the human brain which can be described as *"a non-linear parallel computer, highly complex and responsible for all the information-processing abilities that we possess"* (Attew, 2002: 157), Neural networks are composed of neurons (commonly referred to as 'nodes', 'processing elements' or 'perceptrons'). A biological neuron, in its most basic form, is a data-storing cell, responsible for the cognitive processes (for example, memory, problem-solving, ability to learn) of the brain. An interconnected set or group of these neurons is referred to as a neural network. An artificial neural network, which is an abstraction of this biological neural network, is thus comprised of highly interconnected artificial neurons operating in parallel. Being similar in architecture to the human brain, ANNs thus have the ability to model complex, nonlinear relationships in data.

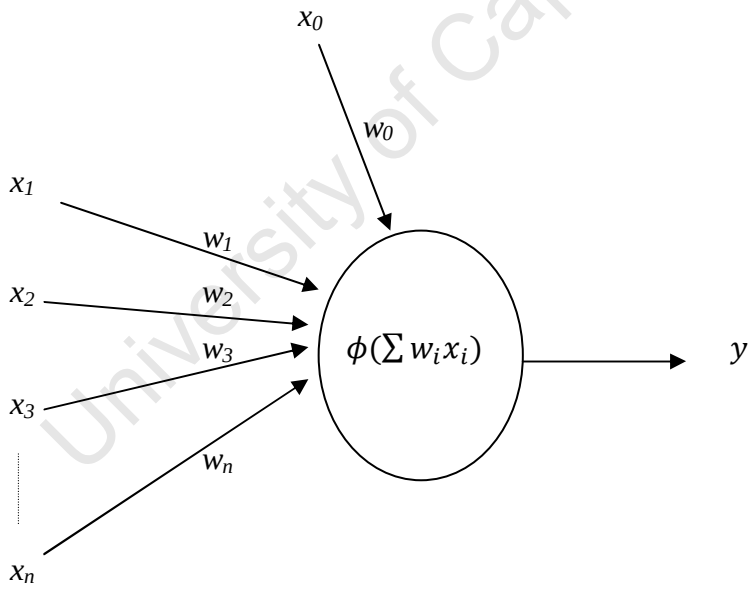
2.5.1 The Structure of a Neuron

A basic biological neuron is a decision unit consisting of synapses, the soma, the axon, and dendrites. The synapses are gaps between the neurons where electrical signals are able to pass from one neuron to the next. Electrical signals then pass to the soma, which in turn, sends its own electrical signal to the axon for transmission of the signal to the dendrites. The dendrites then transmit signals to various synapses. While the biological neuron describes and predicts biological processes, the artificial neuron's goal is one of computational efficiency. An artificial neuron is the basic component of an ANN model. Inputs ($x_0 \dots \dots x_n$) are fed into the neuron, weighted and summed and then passed through a non-linear activation (transfer) function to the output (y). This structure is depicted in Figure 2.1. Inputs, denoted (x_i) are equivalent to the independent variables in linear regression models, while output (y) is

equivalent to the dependent variable. The interconnection between the units are referred to as synaptic weights $(w_0 \dots \dots w_n)$, denoted (w_{ij}). x_0 and w_0 represent the unity bias input and weight respectively. These weights are equivalent to the coefficients of the regression with w_0 representing the intercept term. *"Nodes are basically nonlinear, ensuring that the nonlinearity is spread throughout the network- important for recognising patterns from an input signal whose underlying mechanism is itself inherently nonlinear"* (Attew, 2002: 158).

Figure 2.1 The Structure of a Neuron

The most basic component of an artificial neural network is the artificial neuron. Inputs $(x_0 \dots \dots x_n)$ are fed into the neuron, weighted and summed and then passed through a nonlinear activation function to the output (y). The connection (synaptic) weights $(w_0 \dots \dots w_n)$ represent the interconnection between the units. The weights are modified to represent the respective connection strengths. x_0 represents the bias inputs, while, w_0 represents the bias weight.

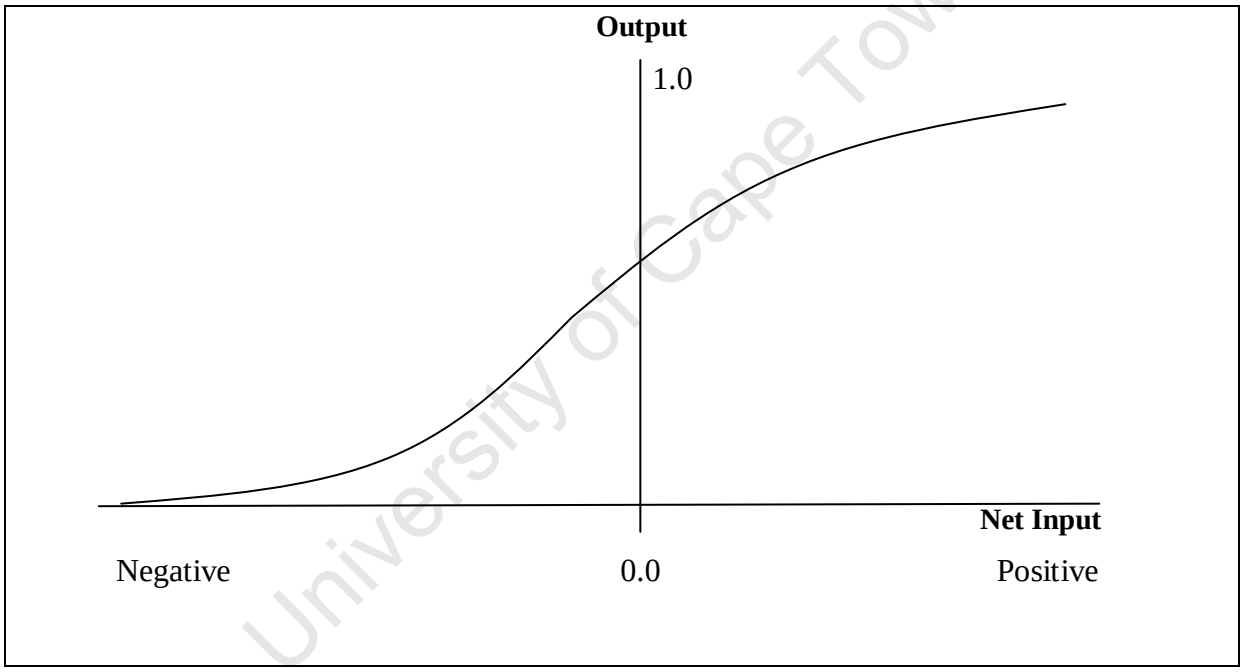


Nonlinearity is introduced into the network by way of an activation function. The sigmoid activation function (ϕ) is the most commonly used transfer function in neural networks. In linear regression the activation function is the identity function. Each neuron processes incoming values with an activation function. An activation value between 0 and 1 then

results. This activation function allows the network to recognise nonlinear patterns in the data. The sigmoid activation function is represented in Figure 2.2.

Figure 2.2 The Sigmoid Activation Function

Nonlinearity (the ability of the network to recognise nonlinear patterns) is introduced into the system via the sigmoid activation function which produces an activation value between 0.0 and 1.0. The sigmoid function is very close to 1.0 for large positive net input values, 0.5 when the net input value is 0.0, and very close to 0.0 for large negative input values, which results in smooth transition between the low and high outputs of the neuron.



2.5.2 Artificial Neural Network (ANN) Architecture

A group of artificial neurons, with weighted interconnections forms a neural network. Neural networks are structured in layers with parallel processing. There exists an input layer, one or more hidden layers and an output layer. Two common ANN architectures are the feed-forward neural networks and the recurrent neural networks. Examining feed-forward neural networks, although multiple layers, with multiple neurons occupying each layer may exist,

the "*single layer feed-forward or multiperceptron network with one hidden layer is the most basic and commonly used neural network in economic and financial applications"* (McNelis, 2005: 22). Figure 2.3 is an example of a (3-2-1) feed-forward neural network, that is, a network comprising 3 neurons in the input layer, 2 neurons in the hidden layer and 1 neuron in the output layer. This network has a fixed-architecture, where the number of neurons are determined in advance and do not change.

Figure 2.3 Artificial Neural Network Architecture

Figure 2.3 is an example of a (3-2-1) fixed-architecture feed-forward neural network, that is, a network comprising 3 neurons in the input layer, 2 neurons in the hidden layer and 1 neuron in the output layer. With fixed-architecture feed-forward networks, connections are unidirectional - they are fed forward. Connections do not pass back to the previous layer.

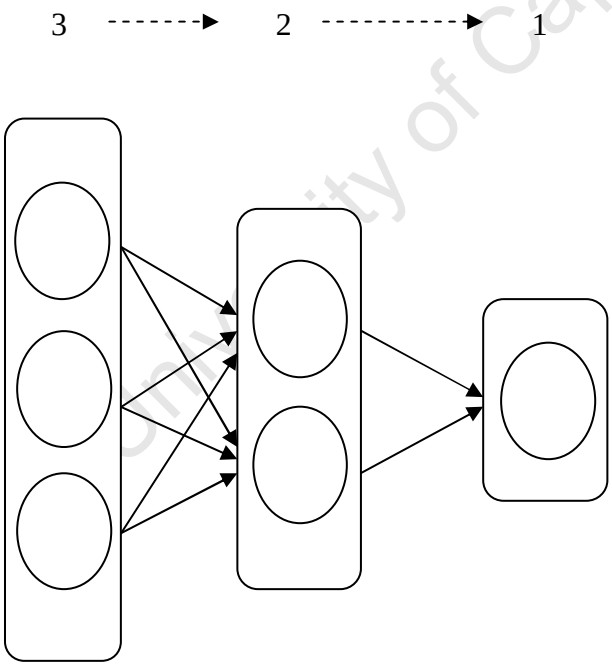


Figure 2.3 is an example of a fixed-architecture network where every neuron is specified in advance (as opposed to networks where the neurons evolve as the network learns). Feedforward neural networks are arranged in layers. Information passes, or is "fed forward", in one direction from left to right in the Figure 2.3. This is done via the interconnection

weights. Every neuron in the input layer is connected to every neuron in the hidden layer, which in turn is connected to the output layer. Examining each layer, an important observation is that the neurons occupying a specific layer have no interconnections between them. Nonlinearity is introduced into the system via the activation function. With feed-forward networks connections are unidirectional - they are fed forward. Connections do not pass back to the previous layer. Recurrent neural networks, on the other hand, *"has feedback from outputs or hidden layers back into earlier layers through a buffer"* (Taylor, 2002: 91). Linear regression models may be viewed as a special case of feed-forward neural network with no hidden layers and one output neuron with a linear activation function.

2.5.3 Feed-Forward ANN Modeling in the Cross-Section of Equity Returns

Equation 2.4, modified from Kanas and Yannopoulos (2001: 389), is the algebraic expression of a standard fixed-architecture ANN model with direct connections between the input neurons and the output neuron, that is used to estimate the return of share *i* in month *t*. This ANN model can be viewed as "*a standard linear regression model augmented with non-linear terms*" (Gonzales, 2000: 9).

$$E(R_{i,t}) = a_{i,0,t} + \sum_{j=1}^{J} g_{i,j,t} F_{j,t-1} + \sum_{k=1}^{K} b_{i,k,t} f\left(d_{i,0,t} + \sum_{j=1}^{J} c_{i,j,k,t} F_{j,t-1} \right) \tag{2.4}$$

Where:

$a_{i,0,t}$ = the bias term to the output layer for period *t*;

$g_{i,j,t}$ = the direct payoff to input neuron $F_{j,t-1}$ from the output neuron (the equity returns) for period *t*;

$F_{j,t-1}$ = the input neuron representing the transformed lagged value of attribute *j* in the input layer;

$b_{i,k,t}$ = the factor payoff to the hidden neuron *k* from the output neuron for period *t* (it measures the sensitivity of the equity returns to movements in hidden neuron *k*);

$d_{i,0,t}$ = the bias term to the hidden layer for period *t*;

$c_{i,j,k,t}$ = the factor payoff to input neuron $F_{j,t-1}$ from the hidden neuron *k* for period *t* (it measures the sensitivity of hidden neuron *k* to movements in input $F_{j,t-1}$);

$\sum_{j=1}^{J} g_{i,j,t} F_{j,t-1}$ = the linear function;

$f\left(d_{i,0,t} + \sum_{j=1}^{J} c_{i,j,k,t} F_{j,t-1}\right)$ = the nonlinear activation function (the sigmoid function); and

$d_{i,0,t} + \sum_{j=1}^{J} c_{i,j,k,t} F_{j,t-1}$ = the input signal within the nonlinear activation function.

Based on Equation 2.4, the expected return determination process receives signals from both the linear function and the nonlinear transformation function. The signal from the linear function is simply the sum of the products of the input neurons and their associated factor payoffs (that is, $\sum_{j=1}^{J} g_{i,j,t} F_{j,t-1}$). For the purpose of this research, each input neuron $F_{j,t-1}$ takes on the raw lagged value of the attribute *j*, or a transformed lagged value of the attribute *j* via one of the following transformation functions:

Natural logarithm function:	$log\ j$
Exponential function:	e^{j}
Exponential of Exponential:	$(e^{j})\ (e^{j})$
Square function:	j^{2}
Fourth Power function:	j^{4}
Square Root function:	$j^{0.5}$
Fourth Root function:	$j^{0.25}$
Inverse function:	$1 / j$
1/ (Square function):	$1 / j^{2}$
1/ (Fourth Power function):	$1 / j^{4}$
1/ (Square Root function):	$1 / j^{0.5}$

1/ (Fourth Root function): $1 / j^{0.25}$

Hyperbolic tangent function: $tanh\ (j)$

The input signal, $d_{i,0,t} + \sum_{j=1}^{J} c_{i,j,k,t} F_{j,t-1}$, has to be activated in order to reflect the nonlinearity in the expected return determination process. Other available activation functions include a *linear*, *Tanh*, *Gaussian* and *Sine* transfer function. The sigmoid logistic cumulative distribution function is employed for the research. The sigmoid logistic cumulative distribution function, $f(x)$, is defined as $1/(1+e^{-x})$. The nonlinear transformation function, $f\left(d_{i,0,t} + \sum_{j=1}^{J} c_{i,j,k,t} F_{j,t-1} \right)$ of the ANN models can thus be redefined as $\frac{1}{\left(1 + e^{-\left(d_{i,0,t} + \sum_{j=1}^{J} c_{i,j,k,t} F_{j,t-1} \right)} \right)}$.

2.5.4 The Multilayer Perceptron (MLP) Architecture of the ANN Models

The feed-forward network, together with the sigmoid activation function is referred to as the multilayer perceptron (MLP) network. According to McNelis (2005:25), the MLP network is *"the basic workhorse of the neural network forecasting approach, in the sense that researchers usually start with this network as the first representative network alternative to the linear forecasting model"*. Equation 2.5, provided by Gonzalez (2000: 9), displays an example of an ANN model with 2 input neurons and 2 hidden neurons in a single hidden layer. This equation maps how the input vectors (that is, the signals generated from the input neurons) are transferred to the output neurons. The input neurons are the lagged values of the style attributes of the sample shares, and the output neurons are the expected returns of the sample shares in time period *t*.

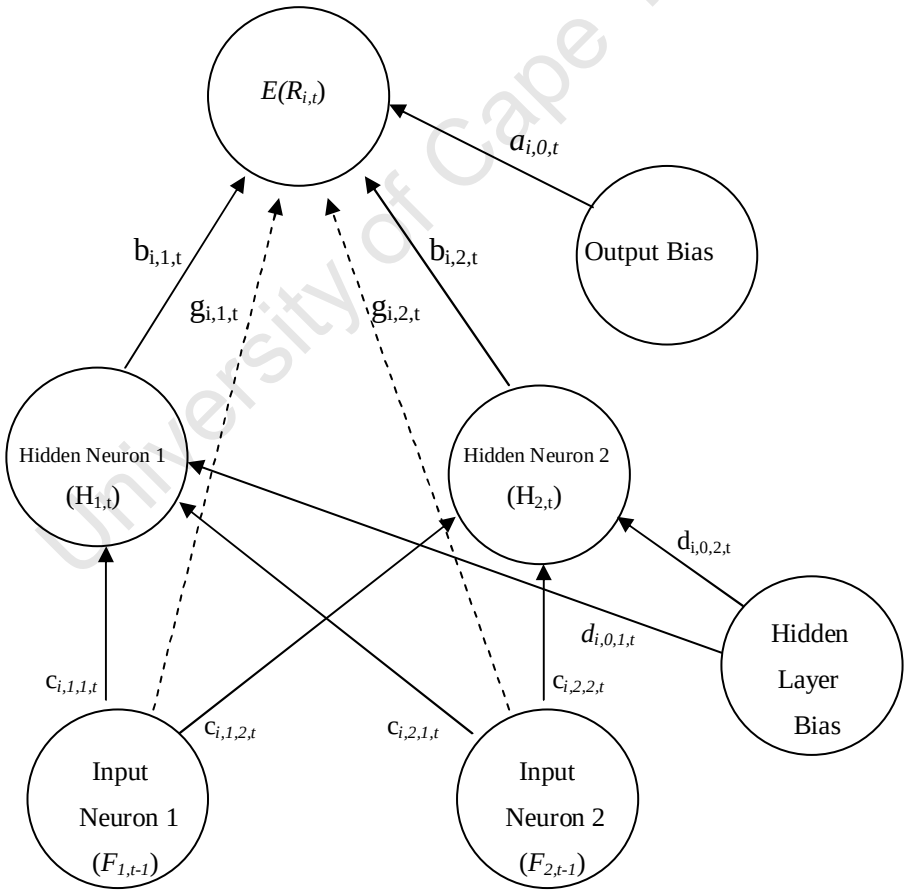
$$E(R_{i,t}) = a_{i,0,t} + (g_{i,1,t} \times F_{1,t-1}) + (g_{i,2,t} \times F_{2,t-1}) + \left(\frac{b_{i,1,t}}{1 + e^{-(d_{i,0,1,t} + c_{i,1,1,t} \times F_{1,t-1} + c_{i,2,1,t} \times F_{2,t-1})}} \right) + \left(\frac{b_{i,2,t}}{1 + e^{-(d_{i,0,2,t} + c_{i,1,2,t} \times F_{1,t-1} + c_{i,2,2,t} \times F_{2,t-1})}} \right) \quad (2.5)$$

The advantage of employing this model is that it separates the contribution of the non-linearity from that of the linearity in expected return estimation. As indicated by Gonzalez (2000: 9), the nonlinear factor payoffs, $b_{i,1,t}$ and $b_{i,2,t}$ in Equation 2.5 will be equal to zero if the share returns do not exhibit any nonlinear traits. Thus, the above model essentially provides an indication as to the effectiveness of mapping nonlinearity in estimating the sample share returns.

Figure 2.4 shows the architecture of Equation 2.5 employing the multilayer perceptron (MLP) network. The MLP is the graphical representation of the manner in which the processing elements (that is, the network neurons) are connected. While $(g_{i,1,t} \times F_{1,t-1}) + (g_{i,2,t} \times F_{2,t-1})$ represents the input signal from the linear transformation function to the output, the other input signals, $(d_{i,0,1,t} + c_{i,1,1,t} \times F_{1,t-1} + c_{i,2,1,t} \times F_{2,t-1})$ and $(d_{i,0,2,t} + c_{i,1,2,t} \times F_{1,t-1} + c_{i,2,2,t} \times F_{2,t-1})$, have to be transformed into the nonlinear format before they can be received by the output neuron. Once these two signals are transformed into the nonlinear format using the sigmoid logistic cumulative distribution function, the hidden neurons, $H_{1,t}$ *and* $H_{2,t}$ are established, with their values being depicted by $\left(\frac{1}{1 + e^{-(d_{i,0,1,t} + c_{i,1,1,t} \times F_{1,t-1} + c_{i,2,1,t} \times F_{2,t-1})}} \right)$ and $\left(\frac{1}{1 + e^{-(d_{i,0,2,t} + c_{i,1,2,t} \times F_{1,t-1} + c_{i,2,2,t} \times F_{2,t-1})}} \right)$ respectively. Following the activation of the hidden layer neurons, the final signal of the nonlinear transformation function connecting the hidden layer and the output neuron, $E(R_{i,t})$, is equal to the sum of the products of the hidden neuron values and their factor payoffs, as depicted by $\left(\frac{b_{i,1,t}}{1 + e^{-(d_{i,0,1,t} + c_{i,1,1,t} \times F_{1,t-1} + c_{i,2,1,t} \times F_{2,t-1})}} \right) + \left(\frac{b_{i,2,t}}{1 + e^{-(d_{i,0,2,t} + c_{i,1,2,t} \times F_{1,t-1} + c_{i,2,2,t} \times F_{2,t-1})}} \right)$.

Figure 2.4 The MLP Architecture of an Augmented ANN Model with 2 Inputs and 2 Neurons in a Single Hidden Layer

This representation is modified from Gonzales (2000: 9). The signal from the linear transformation function, $(g_{i,1,t} \times F_{1,t-1}) + (g_{i,2,t} \times F_{2,t-1})$, is passed from the inputs to the outputs directly without further transformation. The other input signals generated to capture nonlinearity in the expected return determination process, $(d_{i,0,1,t} + c_{i,1,1,t} \times F_{1,t-1} + c_{i,2,1,t} \times F_{2,t-1})$ and $(d_{i,0,2,t} + c_{i,1,2,t} \times F_{1,t-1} + c_{i,2,2,t} \times F_{2,t-1})$, are then transformed into the nonlinear format by the sigmoid logistic cumulative distribution function. The new signals, $\left(\frac{1}{1+e^{-(d_{i,0,1,t}+c_{i,1,1,t}\times F_{1,t-1}+c_{i,2,1,t}\times F_{2,t-1})}}\right)$ and $\left(\frac{1}{1+e^{-(d_{i,0,2,t}+c_{i,1,2,t}\times F_{1,t-1}+c_{i,2,2,t}\times F_{2,t-1})}}\right)$, represent the values of neurons $H_{1,t}$ *and* $H_{2,t}$ in the hidden layer respectively. The ultimate signal being generated by the above nonlinear transformation process to the output neuron, $E(R_{i,t})$ is depicted by $\left(\frac{b_{i,1,t}}{1+e^{-(d_{i,0,1,t}+c_{i,1,1,t}\times F_{1,t-1}+c_{i,2,1,t}\times F_{2,t-1})}}\right)+\left(\frac{b_{i,2,t}}{1+e^{-(d_{i,0,2,t}+c_{i,1,2,t}\times F_{1,t-1}+c_{i,2,2,t}\times F_{2,t-1})}}\right)$.



2.5.5 Neural Network Training

Learning/training the ANN means solving the problem that is being presented to the network. Neurons are connected to each other via the synaptic weights which determine the strength of the connection between the neurons. Training the network thus entails adjusting these connection weights until a certain cost function is minimised, that is, until the difference between the estimated and actual output is at a minimum. Based on the availability of the outcome class and on their learning behaviour, ANN models are categorised into supervised or unsupervised learning. Supervised learning occurs when the target (desired) value is known, whereas unsupervised learning occurs when the target value associated with each input is unknown.

A popular example of supervised learning is the backpropagation algorithm (Werbos (1974) and Rumelhart and McClelland (1986)). The backpropagation learning algorithm is a gradient descent mechanism of updating the weights of the network, until the network error is minimised. It consists of two phases, namely a forward pass and a backward pass. With the forward pass, the activations are propagated from the input to the output layer. The backward pass propagates the network error (the difference between the desired output and the network output) backwards into the network in order to modify the weights and the bias value. Another less popular learning rule in neural networks, which is introduced in this research as an appropriate learning rule for financial data, is the extended Kalman filter (EKF) learning algorithm. This technique is an extension of the popular Kalman filter (Kalman, 1960), extended to incorporate nonlinearity. It is based on state-space theory where the weights are assumed be the state to be estimated. It is through the training/learning process that ANN models are able to learn (by experience) to store and recognise patterns in a changing environment. An example of unsupervised learning is adaptive resonance theory (ART) and bi-directional associative memory.

2.5.6 Advantages and Disadvantages of Artificial Neural Networks

Artificial neural networks are powerful data-modeling tools, which have the advantage of being 'universal approximators', which means that that they can map any arbitrary nonlinear function well. ANNs can recognise patterns in data, even if there exists no functional relationship between the input and the output. Due to this flexibility, they are powerful for forecasting, pattern recognition and classification problems. The danger of this flexibility, however, is that because ANNs have the ability to map any function, it can also lead to overfitting of the data. An added advantage, according to Kaastra and Boyd (1996: 216) is that *"neural networks are less sensitive to error term assumptions and they can tolerate noise, chaotic components, and heavy tails better than most other methods"*. On the other hand, a disadvantage of ANNs, particularly fixed-architecture neural networks is the fact that the number of hidden neurons and number of hidden layers have to be determined in advance. Thus the choice of optimal network topology could severely hamper the accuracy of the model. Further to this is the potential problem of network overfitting where the neural network, due to its universal approximation property could potentially result in network overfitting since it is able to map both relevant and irrelevant patterns in data. This will be addressed in Chapter 7 and Chapter 8.

3

A Review of Prior Literature

3.1 Introduction

According to asset pricing models based on investor rationality, the cross-section of equity returns can be explained by betas or factor loadings on a set of common factors. Early evidence supports the Sharpe (1964) – Lintner (1965) version of the capital asset pricing model (CAPM) and the efficient market hypothesis (EMH). The first test examining the cross-section of equity returns was conducted by Fama and Macbeth (1973) who find the relationship between high beta stocks and average returns to be positive and approximately linear. The development of the arbitrage pricing theory (APT) by Ross (1976) led to a series of studies in which proxies for the APT factors replaced the market and betas of the CAPM. Around this time, researchers began testing whether non-risk characteristics, for example, the earnings-to-price ratio and size (as measured by market capitalisation), also affect security returns (see Basu (1977) and Banz (1981)). If this was the case, then the risk characteristics of securities (betas or factor loadings) are not the sole determinants of expected returns as originally postulated. This research into the effect of non-risk characteristics on security returns continued into the 1980s and 1990s, in an attempt to determine if attributes other than beta account for the variation in equity returns. Fama and French (1992), in an influential and controversial study, presents evidence that market betas are unable to explain the cross-sectional variation in equity returns and that size and the ratio of book-to-market equity (B/M) are the two major determinants explaining the cross-sectional variation in average returns.

Since the publication of Fama and French (1992)'s article, a large body of evidence, explaining the identification and nature of the CAPM-related anomalies, has emerged in developed as well as emerging economies. CAPM-related anomalies include the size effect (small-firm effect), the value effect, the momentum effect and long-term price reversal. The degree of commonality explaining the cross-sectional differences in average returns is

explored by Haugen and Baker (1996, 2009) and Hanna and Ready (2005) in developed economies and Serra (2002) in emerging economies.

Results obtained from numerous studies, for example Banz (1981), Fama and French (1992, 1993) and Van Rensburg and Robertson (2004), to name a few, suggests that there exists nonlinearity in explaining the cross-sectional differences in average returns. Eakins, Stansell and Buck (1998), Eakins and Stansell (2003), Cao, Leggio and Schniederjans (2005) and Cao, Parry and Leggio (2009) attempt to forecast equity returns from a set of firm-specific and macroeconomic variables using artificial neural network (ANN) model.

The remainder of the chapter is arranged as follows: Section 3.2 discusses the empirical research on the identities of asset pricing anomalies in the U.S. and developed economies; Section 3.3 discusses the empirical research on the identities of asset pricing anomalies in South Africa and emerging economies; Section 3.4 explores the possible reasons for these anomalies by exploring prior research on the nature of these anomalies. Section 3.5 and Section 3.6 discusses the prior research relating to linear and nonlinear multifactor modeling techniques using firm-specific and macroeconomic variables; Section 3.7 concludes.

3.2 Identification of Asset Pricing Anomalies in the U.S. and Developed Economies

Several deviations from the CAPM, known as "anomalies", were discovered more than three decades ago. Banz (1981) uses firm size (market value), in addition to beta to explain the cross-section of equity returns. Using a methodology similar to Fama and Macbeth (1973), Banz (1981) reports a negative relationship between average stock returns and firm size, after controlling for risk, of common stocks listed on the New York Stock Exchange (NYSE), over the period from 1927 to 1975. The author attributes these results to the misspecification of the CAPM. The results also indicate that the effect lacks a theoretical foundation and that it is not stable over time. Banz (1981: 16-17) concludes, *"we do not even know whether the factor is size itself or whether size is a proxy for one or more true but unknown factors correlated with size…the size effect exists, but it is not at all clear why it exists"*.

Criticism of Banz (1981)'s findings is reported by Roll (1981) who argues that the systematic risk estimates of returns are biased downwards, due to the less frequent trading of small stocks relative to large stocks. On the other hand, Keim (1983), Reinganum (1983) and Blume and Stambaugh (1983), find that the small-firm effect occurs in the month of January and is thus a small-firm-in-January effect. The size effect is further emphasised by Fama and French (1992), who sorts stocks according to both size and beta. Results conclude that high-beta stocks do not produce higher returns than low-beta stocks of the same size. Another explanation is that size is actually a proxy for some unknown factor.

During the same time that empirical tests on the size effect are conducted, tests on the "value effect" are also being explored. Basu (1977, 1983) finds that companies with high earnings-to-price ratio (or low price-to-earnings (P/E) ratio) earn positive abnormal returns relative to the CAPM. Basu (1977) forms portfolios from stocks listed on the NYSE over the period April 1957 to March 1971. The relationship between the performances of the portfolios and their P/E (or E/P) ratio is compared in terms of pre-specified measures. The results reveal that high E/P stocks earn significantly higher returns on a risk-adjusted basis than low E/P stocks.

The author attributes this finding to a violation of the joint hypothesis. According to Ball (1978) this evidence represents evidence against the CAPM but not evidence against the EMH. Banz (1981: 16) argues, *"Basu believed to have identified a market inefficiency but this P/E is just a proxy for the size effect. Given its longevity, it is not likely that it is due to market inefficiency but it is rather evidence of a pricing model misspecification"*. A follow-up study by Basu (1983) concludes that the size and earnings-to-price effect are separate from each other, and that, even controlling for E/P, small firms still tend to have higher returns.

Besides the E/P ratio, henceforth called the earnings yield, value proxies can also be formed from accounting ratios that examine the ratio of a company's market value relative to its fundamental value such as book value, dividends, cash flow and sales, to name a few. Follow-up research also finds evidence of a "value effect" where positive abnormal risk-adjusted returns accrue to portfolios of stocks possessing high ratios of dividend-to-price (D/P or dividend yield), book-to-market (B/M) and cash flow-to-price (C/P). For example, Litzenberger and Ramaswamy (1979) document a positive relationship between dividend yield and common stock returns over the period from 1936 to 1977; Bhandari (1988) documents a positive relationship between leverage and average returns; Statman (1980) and Rosenberg, Reid and Lanstein (1995) report a positive relation between B/M and average returns. By examining the ratio of a stock's market value relative to its fundamental value, stocks can be classified as either value stocks or growth stocks. Companies possessing high B/M, high C/P, high dividend yield and high earnings yield, are classified as value stocks, while companies possessing low B/M, low C/P, low earnings yield and low dividend yield are classified as growth stocks.

Fama and French (1992) combine the following attributes, namely, market beta, size, leverage, book-to-market equity and earnings yield in the cross-section of average returns on the NYSE, American Stock Exchange (AMEX) and the over-the-counter NASDAQ stocks over the period from 1963 to 1990. Stocks are sorted and ranked according to each respective attribute and portfolios formed. Portfolios range from extreme value-oriented to extreme growth-oriented. The results reveal that beta does not explain the cross-section of average returns, the central prediction of the Sharpe-Lintner model. Size and B/M have the strongest

relation to average returns, with a stronger role being afforded to B/M. When size and B/M are included in the regression, the explanatory power of the other attributes disappears. Fama and French (1993) propose a three-factor model with regressions that use, excess market returns ($R_m - R_f$) and mimicking returns for size and B/M factors, that is SMB ("small capitalisation stocks minus big capitalisation stocks") and HML ("high B/M stocks minus low B/M stocks), as explanatory variables. Fama and French (1993) find that the abnormal returns from the three-factor model are not reliably different from zero and thus conclude that a market factor and the proxies for risk factors related to size and B/M successfully explain the cross-section of returns. Fama and French (1992, 1993) thus contend that size (as measured by market capitalisation) and value (as measured by B/M) actually represent risk factors missing from the CAPM.

Lakonishok, Shleifer and Vishny (1994) examine the value effect using stocks listed on the NYSE and the AMEX over the period from 1963 to 1990. Stocks are classified into portfolios based on cash flow-to-price (C/P), earnings-to-price (E/P), book-to-market (B/M), as well as the average historical 5-year growth rate of sales. Portfolios with the highest C/P, highest E/P, highest B/M and lowest growth rates are classified as value portfolios, while those portfolios with the lowest C/P, lowest E/P, lowest B/M and highest growth rates are classified as growth portfolios. The results reveal evidence of a value premium where value stocks outperform growth stocks 5 years after formation. In order to explain the anomalies identified by Lakonishok, Shleifer and Vishny (1994), Fama and French (1996) argue that most of the CAPM anomalies are related and captured by their Fama and French (1993) three-factor model. Using the same anomalies suggested by Lakonishok, Shleifer and Vishny (1994), Fama and French (1996) find that when using portfolios based on these variables in their three-factor model, no estimates of abnormal performance that are reliably different from zero is evident.

On the other hand, Daniel and Titman (1997) argue that it is the actual size and B/M characteristics, and not their risk factors that explain the cross-sectional patterns of average returns. They thus argue in favour of a "characteristic model" which states that the return premia on small capitalisation and high B/M stocks are not related to the covariance structure

of returns. Rather it is their actual characteristics that explain the cross-sectional return variation in U.S stock returns over the period from 1973 to 1993. Daniel and Titman (1997) sort stocks based on size, B/M and risk loadings into portfolios. The authors reject the Fama and French (1993) three-factor model in favour of the characteristic model and conclude that size and B/M are not risk factors in an equilibrium pricing model.

In an attempt to test the robustness of Daniel and Titman (1997) results, Davis, Fama and French (2000) examine the pre-1973 period in order to determine if it is characteristics or factors that better explain U.S. average stock returns. B/M data for stocks listed on the NYSE, AMEX and NASDAQ are collected and analysed over the period from 1929 to 1997. The results reveal that the premium associated with value stocks is similar in the pre-1960 to the post 1960. The authors also find that the value effect subsumes the size effect in the earlier sample period. Davis, Fama and French (2000) find the results of Daniel and Titman (1997) to be not very robust in the out-of-sample period and conclude *"the evidence of Daniel and Titman (1997) in favour of the characteristic model is special to their rather short sample period"* Davis, Fama and French (2000: 405).

Tests on the value anomaly are also extended to international economies. Fama and French (1998) extend their tests to international economies, in order to determine if the value-growth effect is an international phenomenon. Data for the period from 1975 to 1995 is downloaded for stocks listed on NYSE, AMEX, NASDAQ and EAFE (Europe, Australia and Far East). Portfolios based on B/M, C/P, earnings yield and dividend yield are formed. The results reveal that value stocks earn higher risk-adjusted returns than growth stocks over the examination period. A value premium is reported with B/M being the most consistent value proxy. Apart from Italy, high B/M stocks outperform low B/M stocks in 12 out of 13 markets. Portfolios based on C/P, earnings yield and dividend yield produce similar value premia. The authors also report evidence of a value premium in emerging markets. According to the regression results, the standard CAPM fails to explain the international value premium over the examination period.

Bauman, Conover and Miller (1998) use the same value proxies of Fama and French (1998)

to test the size and value anomalies on EAFE and Canadian stocks for the period from 1986 to 1996. The results reveal that although the value anomaly is not evident every year, when it is evident, value stocks outperform growth stocks by a wide margin. There is also evidence of a size effect across the economies is most years over the examination period.

A further test examining the international value anomaly is conducted by Chan and Lakonishok (2004). The authors construct a composite value proxy comprising B/M, C/P, sales-to-price ratio and earnings yield. The sample includes large-caps in the MSCI (Morgan Stanley Capital International) EAFE index. The results reveal evidence of a value anomaly over the period 1989 to 2001. The highest ranked value portfolio earns a return of 12.3% per annum, while the EAFE composite index earns a return of 4.5% per annum over the period 1989 to 2001.

The final anomaly tested is when past stock returns are used as a firm-specific attributes in the cross-section of equity returns. Fama and French (1996), for example, examine the effect of past stock returns on the cross-section of equity returns. The first to examine the effect of long-term past stock returns on the cross-section of equity returns is De Bondt and Thaler (1985, 1987), while Jegadeesh and Titman (1993) examine the effect of short-term past returns on the cross-section of equity returns.

De Bondt and Thaler (1985) classify portfolios as “winners” and “losers” and analyse their subsequent performances. Losers are stocks possessing low returns in the prior 3 to 5 years, while winners are stocks with high returns in the prior 3 to 5 years. Winner and loser portfolios are formed from stocks listed on the NYSE over the examination period from 01 January 1933 to 31 December 1982. The average cumulative abnormal returns (ACARs) of the prior winners and losers are computed and subsequently compared. The results reveal an anomaly whereby past losers earn higher average returns than past winners (known as a “contrarian” effect). Over the examination period, the prior 36-month loser portfolios outperform their respective winner portfolios by 24.6%, 36 months after formation, on average. In addition to this, since formation, the loser portfolios accumulate positive abnormal returns, while the winner portfolios accumulate negative abnormal returns. In terms of the 36-month ACARs, the mean reversals of the loser portfolios are three times stronger

than that of the winner portfolios, on average. Since the majority of the positive abnormal returns of the loser portfolios are earned in January, there is an argument that the results are attributable to the tax-loss selling for the losers.

In a follow-up study, De Bondt and Thaler (1987) extend their research by incorporating factors such as firm size, seasonality and market risk in the study. The capital-gains tax lock-in effect is evident for prior winners, since January excess returns are negatively related to prior December excess returns. Tax-loss selling for losers is not evident in this study. Evidence reveals that the mean reversal of prior winners and losers is not explained by the size effect and market risk (CAPM-betas). On the other hand, Chan (1998) argues that the risks of winners and losers are not constant. The betas of the losers are found to increase (following a period of abnormal loss), while the betas of the winners are found to decrease (following a period of abnormal gain) over time. Chan (1988) concludes that examining the period from 1933 to 1985, when changes in risk are controlled for, abnormal returns between prior winners and losers are a minimum.

Chopra, Lakonishok and Ritter (1992) extend the study of De Bondt and Thaler (1985, 1987) in an attempt to determine the extent to which changes in portfolio beta or size bias the De Bondt and Thaler (1985, 1987) result. Regression analysis on the abnormal returns of prior winners and prior losers on the NYSE for the period from 1931 to 1986 is performed. Chopra, Lakonishok and Ritter (1992) report that even when time-varying betas are taken into account, there are still large differences in abnormal returns between prior winners and prior losers. Further to this, since size, prior returns and betas are all interrelated, all three variables have to be included in any study of cross-sectional returns performance in order to eliminate the omitted variables bias problem. Incorporating these three variables in a multiple regression model reveals that prior losers outperform prior winners by 4.8%, on average, 5 years after formation, after controlling for both size and beta. The authors, however, conclude that the degree of mean reversal is stronger for smaller firms.

While De Bondt and Thaler (1985, 1987) find evidence of a contrarian effect, Jegadeesh and

Titman (1993, 2001), on the other hand, find evidence of a momentum effect, where recent prior winner returns (that is, 1-year or less portfolio return) outperform recent prior loser returns. Jegadeesh (1990) finds that stocks that have performed well over the past few months tend to earn high returns over the next month, while stocks that have performed poorly over the past few months, tend to earn low returns over the next month. In a follow-up study, Jegadeesh and Titman (1993) examine the returns to buying past winner portfolios and selling past loser portfolios on the NYSE and the AMEX for the period from 1965 to 1989 based on 3- to 12-month prior return momentums. Significant abnormal returns are found for the relative strength strategies in the first year following formation. However, these abnormal returns dissipate in the next two years after formation. Evidence also indicates that the relative strength strategy profitability is not due to its systematic risk. Jegadeesh and Titman (2001), in an updated study, again confirm these findings.

Fama and French (1996), using their three-factor model, test the long-term reversal strategy of De Bondt and Thaler (1985). The results find no estimates of abnormal returns that are reliably different form zero. This finding, however, is not consistent when testing the momentum strategy in the Fama and French three-factor model. Since the intercepts are all reliably positive, Fama and French (1996) conclude that the short-term momentum strategy of Jegadeesh and Titman (1993) is left unexplained by the Fama and French (1993) three-factor model. Similar results is found by Brennan, Chordia and Subrahmanyam (1998) who conclude that, given the Fama and French (1993) three-factor model, size and B/M characteristics do explain average return differences. Further to this, the authors find consistency with Fama and French (1996) results, in that the Fama and French three-factor model was unable to explain the momentum effect. On the other hand, Carhart (1997) includes a momentum factor constructed by the monthly return difference between the returns on the high and low prior return portfolios in an attempt to capture the possible momentum anomaly. This can be viewed as an extension to the original Fama and French (1993) 3-factor model. The 4-factor model is found to capture momentum anomaly in the U.S. stock market.

Examining international economies, Schiereck, De Bondt and Weber (1999) form momentum portfolios based on 6-month and 12-month excess returns on the Frankfurt Stock Exchange

(FSE) for the period from 01 January 1961 to 31 December 1991. The results reveal that momentum portfolios outperform the market proxy on a risk-adjusted basis. When portfolios based on contrarian strategies using 36-, 48- and 60-month excess returns are formed, these portfolios outperform the DAX stock index over the examination period. Chan, Hameed and Tong (2000) explore the profitability of momentum investing in broad equity market indices over the period from 01 January 1980 to 30 June 1995. Relative strength strategies are constructed on 23 equity market indices. This includes 9 countries from the Asia-Pacific region, 11 countries from Europe and 2 countries from North America and South Africa. The strategy includes buying the winner country indices and simultaneously selling the loser country indices based on 1-, 2-, 4-, 12- and 26-week prior return momentums of the respective indices. The results suggest that profits earned by these momentum strategies are statistically significant for short holding periods of less than 4 weeks. On the other hand, Forner and Marhuenda (2003) conduct studies on the Spanish Stock Exchange for the period from 01 January 1963 to 31 December 1997. The results reveal that the 12-month momentum strategy and the 60-month contrarian strategy yield significant positive abnormal returns over the examination period.

3.3. Identification of Asset Pricing Anomalies in South Africa and Other Emerging Economies

Early South African tests on CAPM-anomalies include research conducted by De Villiers, Lowlings, Petit and Affleck-Graves (1986), Bradfield, Barr and Affleck-Graves (1988), Bradfield (1990), Page and Palmer (1991) and Page (1996). Tests on the overreaction hypothesis include research conducted by Page and Way (1992/1993) and Muller (1999). Page and Way (1992/1993) adopt the methodology of De Bondt and Thaler (1985) on the Johannesburg Stock Exchange (JSE) over the period from 1974 to 1989. Winner and loser portfolios are formed based on 36-month prior cumulative excess returns. The results reveal that the loser portfolios outperform the respective winner portfolios by 14.5%, on average, 36 months after formation. Consistent with the result of De Bondt and Thaler (1985), the asymmetrical reversals of winners and losers are observed. In a follow-up study, Muller (1999) forms portfolios from a sample of shares based on the largest 200 shares by market capitalisation on the JSE over the period from 1985 to 1998. The results reveal that positive abnormal returns initially accrue to both winner and loser portfolios. The loser portfolios lose their initial momentum after 340 days, while the winner portfolios lose their initial momentum after about 600 days.

Fraser and Page (2000) examine the value effect and the momentum effect on the JSE over the period from 1978 to 1997. By evaluating the interaction of these value and momentum strategies, the authors conclude that value and momentum investing independently explain the cross-sectional variation in returns on the JSE. Van Rensburg (2001) is one of the first publications to document tests on the identification of style-based effects in South Africa by examining a multitude of candidate style attributes on the JSE. Van Rensburg (2001) examines 23 candidate style attributes using shares listed on the industrial sector of the JSE over the period from 1983 to 1999. Style attributes are categorised into one of the following groups: (i) value, (ii) future earnings and growth and (iii) irrationality/neglect. By adopting a 'portfolio-based' approach, factor mimicking portfolios are formed with their risk premia assumed to be the spread between the risk premium of the fractile with high attribute and the

risk premium on the fractile with low attribute value. Two risk factors identified in this research include a value risk premium and a small-firm risk premium. Anomalies associated with the following nine attributes are identified: (i) earnings yield, (ii) past twelve month returns, (iii) market capitalisation, (iv) dividend yield, (v) six month's past returns, (vi) leverage, (vii) cash flow-to-debt, (viii) turnover and (ix) three month's past returns. The regression results of these nine attributes could not be explained by the 2-factor APT documented by Van Rensburg and Slaney (1997) in which market risk is segmented into the industrial risk factor and the resource risk factor. Cluster analysis on the returns of the style-based portfolios reveal that value (earnings yield), market capitalisation and momentum (past 12-month returns) represent sources of style-based risks on the JSE.

In a follow-up study, Van Rensburg and Robertson (2003a) adopt the 'characteristic-based' approach of Daniel and Titman (1997) on the JSE over the period from 1990 to 2000. The main focus of their research was on the identification of the style attributes that explain the cross-section of returns on the JSE. Using cross-sectional regression, share returns are regressed on the values of 24 style attributes. In a univariate test, the time-series factor payoff to each style attribute is estimated. Van Rensburg and Robertson (2003a: 1) explain *"the time-series of the slope coefficients estimated in this manner represent the 'rewards' accruing to each characteristic in each period while a share's exposure to the factor is directly observable by the value of the characteristic concerned"*. Style attributes with significant payoffs are then used in a stepwise permutation procedure. Attributes are added to the multifactor models until the explanatory power of the multifactor model ceases to improve. The univariate results provide evidence of a value effect and size effect where the following attributes, namely, price-to-NAV, dividend yield, price-to-earnings, cash flow-to-price, price-to-profit and market capitalisation are extracted. The multifactor results reveal a two-factor model with size (market capitalisation) and P/E as the resulting explanatory variables (two sources of style risk). None of the momentum factors are found to be significant in this study.

In an attempt to further test the two sources of style risk, namely, market capitalisation and price-to-earnings ratio, identified in Van Rensburg and Roberson (2003a), Van Rensburg and Robertson (2003b) adopt the methodology of Fama and French (1992) on the JSE. The

results reveal that low P/E shares, whilst earning higher returns, also have lower betas. Contrary to international evidence, which find small firms to actually be riskier investments, in this study small capitalisation shares are found to earn higher returns, while possessing lower betas. The results are also consistent with the evidence of Fraser and Page (2000) and Van Rensburg and Robertson (2003a) in that the size and value effects are found to operate independently of each other on the JSE.

Van Rensburg and Robertson (2004) adopt the methodology of Daniel and Titman (1997) and Daniel, Titman and Wei (2001) on the JSE and form simulated quintile portfolios over the period from 01 July 1990 to 30 June 2000. The objective of this test is to determine whether factor loadings on factor mimicking portfolios or the direct style attributes best explain the cross-section of JSE returns. The authors conclude, *"in both the case of size and of price-to-earnings, the attribute values are able to discriminate the spread in future returns more clearly than corresponding factor loadings"* (Van Rensburg and Robertson, 2004: 346). The results thus reveal that the cross-section of returns on the JSE is better explained using attributes rather than factor loadings.

With regard to emerging market studies, Bekaert and Harvey (1995) and Ferson and Harvey (1999) suggest that emerging market returns can be predicted with a set of both global and local variables, including variables such as dividend yield, credit spread, change in the short-term risk-free rate, P/E, GDP and the ratio of consumption to wealth, to name a few. Harvey (1995) analyses 20 equity markets in emerging economies and presents evidence suggesting that returns in emerging markets are generally more predictable than returns in developed markets. The author further suggests that emerging market returns are influenced by local and not global information variables. Harvey (1995: 812) concludes "*local information variables play a much more important role in predicting emerging market returns...over half of the predictable variance in the emerging market returns can be traced to local information*".

Claessens, Dasgupta and Glen (1998) examine the cross-section of stock returns for 19 emerging economies by examining the effect of several explanatory variables, in addition to beta, on asset returns over the period 1986 to 1993. The results reveal that, in addition to beta,

size and trading volume are found to have significant explanatory power in some economies. Earnings-to-price ratio and dividend yield are also found to have explanatory power in fewer economies. Exchange rate risk also has significant explanatory power. The authors conclude that while size, price-to-book ratio and dividend yield have explanatory power, the signs of the coefficients are not consistent with those of developed economies. This is most pronounced for size.

Rouwenhorst (1999) examines the sources of return variation in 20 emerging markets over the period 1982 to 1997. The authors find evidence that emerging market return factors are qualitatively similar to return factors in developed economies. Averaging across all emerging economies, results indicated evidence of a value effect, momentum effect and a size effect. This is consistent with the findings of Fama and French (1998) but inconsistent with the findings of Claessens, Dasgupta and Glen (1998). Regarding beta, evidence did not exist suggesting that high beta stocks outperform low beta stocks. With respect to liquidity, there is no evidence suggesting that average returns are related to liquidity. This is inconsistent with results from developed economies.

Achour, Harvey, Hopkins and Lang (1999) examine 28 firm-specific attributes in South Africa over the period from 1993 to 1998. The in-sample period spans the period from 1993 to 1995, while the out-of-sample period spans the period from 1996 to 1998. Employing a screening methodology, for each attribute, firms are ranked in order of expected returns and assigned to a particular fractile. Screening factors are divided into fundamental factors, expectation factors, technical factors and size factors. The results reveal that earnings-to-price and earnings forecast revisions yield the highest average annualised excess returns, while earnings forecast revisions and book-to-market yield the greatest average annualised underperformance. Momentum indicators are unsuccessful in this research.

Serra (2002) examines the role of a set of *a priori* specified factors in order to determine the commonality in the cross-section of returns across emerging economies (including South Africa). The results reveal that the important factors are common across emerging economies and similar to the factors identified in developed economies and that the driving factors in

emerging markets are consistent with Fama and French (1998). The six most important attributes in the cross-section of emerging market returns included technical factors (12-week lagged holding period returns), firm characteristics (earnings-price, book-to-market, dividend yield) and liquidity factors (size and price per share).

Results do not reveal evidence of a size effect. Serra (2002) concludes that contrary to evidence from developed markets, the average payoffs of liquidity factors are positive. Consequently, Serra (2002: 17) concludes, "*the size effect is thus not supported by the data*".

3.4 The Nature of Asset Pricing Anomalies

Some researchers explain the role of non-risk characteristics as being a consequence of the rational pricing of assets (Fama and French (1992, 1993, 1996, 1998)), investor irrationality (Chopra, Lakonishok and Ritter (1992), Lakonishok, Shleifer and Vishny (1994) and Haugen, (1995)) or as a result of methodological bias (Lo and MacKinlay (1990) and Brown and Goetzman (1995)). Each view, in turn, is discussed below.

3.4.1 Modern Finance View: Rational Asset Pricing

One interpretation of the anomalies is held by Fama and French (1992, 1993, 1995, 1996, 1998) who interpret their findings as being consistent with the efficient market hypothesis, but not with the single-factor CAPM. They argue that although the anomalies provide evidence against the CAPM, they do not provide evidence against a rational multifactor model in which there are multiple risk factors. Their argument is that size and book-to-market proxy some unobserved risk factors. According to Fama and French (1992), superior returns represent compensation for risk where portfolios formed based on respective firm-specific attributes are interpreted as mimicking portfolios. The returns of these mimicking portfolios are "*correlated with relevant state variables representing consumption or production opportunities*" (La Porta, Lakonishok, Shleifer and Vishny, 1997: 859). It is also found that some variables are redundant in explaining average returns. The relationship between average returns and other measures of value (earnings yield and leverage) are absorbed by the combination of size and B/M. Fama and French (1993) report that factor mimicking portfolios related to size and book-to-market add substantially to the variation in stock returns explained by the market portfolio. Regarding the relationship between beta and average returns, the authors conclude, "*if there is a role for beta in average returns, it is likely to be found in a multifactor model that transforms the flat simple relation between average return and beta into a positively sloped conditional relation*" (Fama and French, 1992: 449). Fama and French (1993) conclude that size and B/M proxy for sensitivity of risk factors in returns, thus consistent with rational pricing. Fama and French (1995, 1996) in support of this risk-

based view proposed by Fama and French (1993), argue that the value premium is compensation for risk missed by the Sharpe-Lintner (1964, 1965) CAPM. According to Fama and French (1993, 1996) poor performing small stocks having high B/M ratios are vulnerable to financial distress and thus command a 'distress premium'.

The anomaly that is not explained by the above interpretation is the momentum effect of Jegadeesh and Titman (1993). In explaining the momentum effect, Fama and French (1996), instead of providing a risk-based explanation, argue that the momentum effect arises as a result of data-snooping or survivorship bias. Fama and French (2008: 1653) conclude, *"like the patterns in average return associated with net stock issues, accruals, profitability and asset growth, return momentum is left unexplained by the 3-factor model of Fama and French (1993) as well as by the CAPM"*.

3.4.2 Behavioural View: Irrational Mispricing

Another interpretation is that the anomalies arise as a result of investor irrationality due to psychological biases which causes investors to make irrational forecasts. De Bondt and Thaler (1985, 1987) attribute their results to the overreaction hypothesis. De Bondt and Thaler (1985) assert that irrational investors persistently overweight recent information and underweight long-term fundamental information. This over- or understated information is expected to correct to its long-term value, given that the fundamentals remain the same. It is this overreaction hypothesis that result in the profitability of the momentum and contrarian strategies. Momentum strategies can thus be devised to profit from the temporary overshooting of asset prices before the market corrections take place, while contrarian strategies can be used to take advantage of the reversal of asset prices when the market is ready to correct. Jegadeesh and Titman (1993) also provide evidence of overreaction and subsequent correction/ mean reversal. Barberis, Shleifer and Vishny (1998), Daniel, Hirshleifer and Subrahmanyam (1998) and Hong and Stein (1999), present models based on investor behavior, which assume that momentum profits are due to the biases in the manner in which investors interpret and act on information.

According to Chopra, Lakonishok and Ritter (1992), Lakonishok, Shleifer and Vishny (1994) and Haugen (1995), differentials in predicted returns come as a surprise to investors. Haugen and Baker (1996: 402) argue that the differentials could be as a result of over- or underreaction to various events and that *"distortions in the pattern of expected realised returns, caused by bias in the pricing of stocks, can mask the true nature of the relation between expected return and risk, whatever its nature"*. Lakonishok, Shleifer and Vishny (1994) argue that due to irrationality, naïve investors incorrectly overestimate the difference in future growth rates between glamour and value stocks. Investors extrapolate past earnings growth into the future and thus become overly optimistic about stocks that have performed well in the past (glamour stocks) and overly pessimistic about stocks that have performed poorly in the past (value stocks). Glamour stocks are companies that are characterised by low B/M (or low E/P) ratios. These companies subsequently underperform once their growth in earnings disappoints investors. Value stocks, on the other hand, characterised by high B/M (or high E/P) ratios, subsequently outperform once their earnings growth positively surprise investors. Evidence in support of this rationale is provided by La Porta (1996).

An alternative view for the value premium is offered by Daniel and Titman (1997). This view posits that the value premium can be attributed to a behavioural explanation that is not based on investor overreaction, but rather one that traces the attribute characteristics. The value premium argument is not consistent with Fama and French (1993)'s risk argument. For example, investors may have a preference for growth stocks over value stocks because they perceive growth companies to be stronger companies. The resultant value premium has nothing to do with risk. Davis, Fama and French (2000: 390) conclude, *"the behavioural overreaction story can also be viewed as a variant of the characteristic model…the characteristics model covers anything that produces a premium for the value characteristic relative to the growth characteristic and is not the result of risk"*.

3.4.3 Biases in Asset Pricing

With regard to biases in asset pricing, several sources of bias include 'survivorship' bias, 'data snooping' bias, 'look ahead' bias and spurious regressions. Kothari, Shanken and Sloan (1993) suggest that the observed predictive ability is an artifact of the research design and database used to conduct the study. They emphasise that the predictive power of certain variables will reduce if a different dataset and methodology is employed. Lo and Mackinlay (1990) on the other hand, argue that the research results are spurious and a result of data snooping, which has accidentally found patterns in historical data. Chapter 4 addresses issues of methodological concern.

3.5 Linear Multifactor Quantitative Modeling

Haugen and Baker (1996) test the validity of a number of factors that includes approximately 50 measures of accounting information and past return behaviour to determine which factors explain security returns within a country over the period from 1979 to 1993 over five economies, namely United States of America, United Kingdom, Japan, Germany and France. These factors were categorised into risk factors, liquidity factors, price level factors, growth potential factors and price history factors. Monthly cross-sectional regression on the Russell 3000 stocks is conducted. Expected return factor models are constructed to forecast monthly returns. For the first half of the examination period, factor payoffs associated with the respective factors are averaged and the factors subsequently ranked based on the absolute values of their *t*-statistics associated with the respective means. These results are then compared with the second half of the examination period. The results reveal stability in the signs and coefficient values. The results also reveal that three of the four 12 most significant factors are technical. This is followed by the liquidity factors. Prices level factors are next, that is, earnings yield, book-to-price ratio and cash flow-to-price ratio are all positive and significant. An important finding is that none of the CAPM or APT factors are important determinants. Thus, according to Haugen and Baker (1996: 411), *"comparing the two periods shows a high degree of commonality in the signs and sizes of the coefficients. All the important factors maintain their signs, and the sizes of the coefficients are remarkably similar"*.

In conducting out-of-sample tests, Haugen and Baker (1996) find that it is the collective power of many of the factors that accounts for the returns prediction accuracy. Haugen and Baker (1996: 435) suggest, "*…the most plausible explanation for the predictive power of the multifactor model seems to be its exploitation of important forms of bias in pricing in the five markets"*. Overall, the authors conclude that expected return factor models are quite accurate in predicting future returns in the five major world economies. There exists a degree of commonality in the important factors.

In a follow-up study, Haugen and Baker (2009) extend the application of their original study

to longer period spanning forty four years from 1963 to 2007. The results are consistent with the results of the original study. "*We find power and stability in the factors that are most influential in determining the structure of stock returns. In addition, to its explanatory power, the model has amazing and consistent power in predicting which stocks will have relatively high and relatively low future returns*" (Haugen and Baker 2009: 2). The results prove to be inconsistent with the efficient market hypothesis.

Frankish (2004) adopts the methodology of Haugen and Baker (1996) and Van Rensburg and Robertson (2003) in an attempt to investigate the cross-section of returns on the All Share Index components of the JSE in South Africa and the shares comprising the S&P 500 in the U.S. market. The author concludes that in building multifactor models, it is the combination of factors and not the factors individually that have an important role to play in explaining the cross-section of returns.

Hanna and Ready (2005) extend the Haugen and Baker (1996) study by adding the examination of transaction costs and after-transaction returns to the study. The authors adopt the same sample, time period, as well as examine the period after the original study. The strategies of the Haugen and Baker (1996) study are examined and compared to the B/M strategy of Fama and French (1992) and the momentum strategy of Jegadeesh and Titman (1993). Hanna and Ready (2005) conclude that taking transactions costs into account, the Haugen and Baker strategies do not provide attractive returns if the investor had formed portfolios based on B/M and momentum.

3.6 Nonlinear Multifactor Quantitative Modeling

Since the introduction of the Sharpe-Lintner (1964, 1965) capital asset pricing model, numerous studies testing the predictability of stock returns in a linear single factor and multifactor framework have emerged. Abhyanker, Copeland and Wong (1997) suggest that there exists growing evidence that stock return predictability can perhaps be explained by a nonlinear model. For example, Banz (1981: 16) concludes, *"The size effect is not linear in the market proportion (or the log of the market proportion) but is most pronounced for the smallest firms in the sample"*. Motivation for a model that accounts for nonlinear behavior is also dated back to Fama and French (1992). Stehle (1997) argues that Fama and French assume a nonlinear relationship between expected returns and the untransformed values of the independent variables by using the logarithm of size and the logarithm of book-to-market ratio as independent variables in their cross-sectional regression models. Further to this, Stehle (1997) argues that Fama and French use this form of nonlinear relationship without providing any reason.

Research conducted by Eakins, Stansell and Buck (1998) is one of the first studies to examine nonlinearities in financial data by examining a set of firm-specific attributes for firms invested in by institutional investors. The sample includes companies listed on the NYSE, AMEX or those trades through the national association of security dealers. The data used in the study details the institutional ownership of about 3000 firms for the period from 1988 to1991. An ANN model is constructed which includes the lagged values of ten firm-specific variables as inputs, namely, net profit margin, operating profit margin, current ratio, total asset turnover, debt-to-asset ratio, return-on-assets, price-earnings ratio, market value, trading volume and percentage ownership as the model output. The model is trained via the gradient descent backpropagation learning algorithm. Motivation for an ANN model is due to results of the first phase of their study in which tests for linearity revealed that beta, firm size and trading liquidity are nonlinear in their relationship with institutional ownership. This ANN model is then compared to a tobit regression model. In terms of predictive power, the ANN model outperforms the tobit regression model over the examination period.

In a similar study Eakins and Stansell (2003), in an attempt to determine whether value stocks provide superior investment returns, construct an ANN model using the descriptors of value, that is, price-to-cash flow, price-to-book, dividend yield, earnings yield, sales and market capitalisation as inputs in the neural network, while percentage total return is specified as the output. All stocks listed on COMPUSTAT over the period from 1975 to 1996, with a market capitalisation value greater than a current value of US$150 million are included in the study. The model is trained using the backpropagation training algorithm. The results provided evidence that the risk-adjusted returns of portfolios selected by the ANN model are greater than the results achieved by other forecasting models. Consistent with Fama and French (1992, 1993) value stocks provide higher returns with lower risk than can be obtained from the random walk process. Eakins and Stansell (2003: 87) concludes, "*the results of this paper show that when presented with a portfolio of securities, neural networks select portfolios of securities that yield superior total returns by those firms commonly defined as value securities*".

Cao, Leggio, Schniederjans (2005) examine firm-specific attributes in emerging economies over the period from 1999 to 2002. Daily closing prices, quarterly book value and common shares outstanding for 367 public corporations traded on the Shanghai Stock Exchange are downloaded. ANN models are constructed to predict stock price movement in the Shanghai Stock Exchange and compared to linear regression models. Four models are constructed, two linear and two ANN models. The linear model constructed is a univariate CAPM where stock returns are regressed on beta. A multivariate model is a replication of the Fama and French three-factor model. The ANN models are the nonlinear counterparts of the univariate and multivariate models. The backpropagation algorithm is used to train the ANN models. The results indicate that, in terms of the linear models, The CAPM outperform the three-factor model in stock return prediction. It is also found that the ANN models outperform the linear counterparts in terms of their predictive power.

Cao, Parry and Leggio (2009) extend the methodology of Cao, Leggio, Schniederjans (2005) by comparing three linear models with three ANN models on the Shanghai Stock Exchange over two sub periods, namely, 1999 to 2002 and 2003 to 2008. The models include a linear

and an ANN univariate time series model; a multivariate linear and nonlinear version of the CAPM (where stock returns are regressed on market returns); and the linear and nonlinear version of Fama and French (1993) three-factor model. The results once again confirm the forecasting superiority of the ANN models relative to the linear models. There is also no significant difference between the CAPM and the three-factor model, both in the linear and the nonlinear case.

Other tests that support the use of ANN models in financial research include studies conducted by Kanas and Yannopoulos (2001) and Hung, Liang and Liu (1996). Kanas and Yannopoulos (2001) construct both a linear and an ANN model in order to examine and compare (out-of-sample) monthly returns for the Dow Jones (DJ) and Financial Times (FT) indices, using lagged return series of dividends and trading volume as the explanatory variables. With regard to both indices, the results reveal the superiority of the out-of-sample ANN forecasts. In addition to this, the ANN forecasts explain the forecasting errors of the linear models for both indices. This is not the case with the linear model. The authors' findings provide support for the underlying nonlinear relation between stock returns and fundamental attributes.

Hung, Liang and Liu (1996) conduct research on the Taiwan Stock Exchange in order to determine the feasibility of integrating the portfolio management process (and asset pricing models) and artificial neural networks. This approach includes blending the arbitrage pricing theory (APT), artificial neural networks (ANN) and a portfolio constructor. The task of factor analysis and the APT is to identify the factors and price the risk assets available for portfolio building. Then, an ANN model is used to predict the future trend of each risk factor. A fixed-architecture network is trained using the backpropagation algorithm. Finally, based on investor preference, the portfolio constructor is used to select the optimal portfolio from the candidates. The results provide evidence that the integrated approach provides a more optimal solution than when the APT or ANN is used alone, by eliminating the shortcomings of each technique. Integrating APT and ANN also outperformed the benchmark (TSEWPI, the weighted stock price index published by the Taiwan Stock Exchange), as well as other integrating techniques such as blending ARIMA with APT.

3.7 Conclusion

Early evidence supports the Sharpe (1964) and Lintner (1965) capital asset pricing model (CAPM) as well as the efficient market hypothesis (EMH). The CAPM postulates that different stocks have different expected rates of return due to their differing betas (non-diversifiable risk). However, evidence of "anomalies" cast doubt on the validity of the joint (EMH-CAPM) hypothesis. For example, Banz (1981) finds evidence of a negative relation between the size of a firm and average stocks returns. Fama and French (1992) find that size and B/M ratio capture the cross-sectional variation of U.S stock returns over the period from 1963 to 1990. It is found that size and B/M ratio are able to capture the cross-sectional variation better than other combinations of variables. Beta is found to possess almost no explanatory power over the period. Fama-French (1993, 1995) studies support the evidence of the 1992 study. Fama and French (1993) propose a three-factor model with regressions that use excess market returns and mimicking returns for size and B/M factors as explanatory variables. Fama and French (1992, 1993) thus conclude that size and value represent the missing risk factors of the CAPM.

There are various different interpretations of the Fama and French (1992, 1993) results. For example, Fama and French (1992, 1993, 1995, and 1996) provide a risk-based explanation for these anomalies arguing that although the anomalies are inconsistent with the CAPM, they do not provide evidence against a rational multifactor model with multiple risk factors. The anomaly, however, not explained by Fama and French three-factor model is the momentum anomaly of Jegadeesh and Titman (1993). A second interpretation of the anomalies is that they are a result of investor irrationality which causes investors to overreact/underreact. Advocates of this interpretation include De Bondt and Thaler (1986, 1987), Lakonishok, Shleifer and Vishny (1994) and Haugen (1995).

On the other hand, Daniel and Titman (1997) reject the Fama and French (1993) three-factor model and argue in favour of a behavioural explanation not based on investor reactions, but rather the actual characteristics themselves. This characteristic model posits that the return

risk premium on small capitalisation and high B/M stocks are not related to the covariance structure of returns, rather it is the actual characteristics themselves that explain the cross-section of returns. These characteristics can be used as model inputs in building expected return factor models. Studies in support of this includes tests conducted by Haugen and Baker (1996, 1999), Serra (2002), Frankish (2004) and Hanna and Ready (2005).

Tests conducted by Abhyanker, Copeland and Wong (1997), Banz (1981), Fama and French (1992, 1993), Van Rensburg and Robertson (2004), Eakins, Stansell and Buck (1998, 2003), Cao, Leggio, Schniederjans (2005) and Cao, Parry and Leggio (2009) suggest that there exists growing evidence that stock return predictability can perhaps be explained by a nonlinear model. This provides a motivation to incorporate nonlinearity in an expected return factor model on the JSE Securities Exchange.

4

Data and Methodology

4.1 Introduction

Empirical tests of market efficiency reveal anomalies that cannot be explained by the capital asset pricing model (CAPM) of Sharpe (1964) and Lintner (1965). The CAPM-related anomalies are firm-specific and can be applied to form potential alpha-generating investment styles that capture the characteristics of the anomalies. The factor loading approach of Fama and French (1993) suggests that risk-adjusted abnormal returns are fully explained by the value and size factor mimicking portfolios. This proposition is rejected by Daniel and Titman (1997) who argue that style attributes (characteristics) themselves explain the cross-section of equity returns. While Van Rensburg (2001) identifies three major style clusters, namely size, value and momentum on the JSE Securities Exchange (JSE), the results of Van Rensburg and Robertson (2003a) suggest that the cross-section of JSE equity returns are explained by size and price-earnings attributes, rather than the return covariance of the factor mimicking portfolios. Further investigations of Van Rensburg and Robertson (2004) suggest that style attributes rather than factor loadings better explain empirical anomalies on the JSE. Empirical implications from the series of investigations in this field motivate this research to develop expected return factor models on the JSE from style attributes themselves rather than using factor mimicking portfolios.

Results obtained from numerous studies, for example Banz (1981), Fama and French (1992, 1993) and Van Rensburg and Robertson (2004), to name a few, suggest that there exist nonlinearity in explaining the cross-sectional equity returns. This motivates the research to develop nonlinear models, in addition to linear models, to forecast JSE equity returns. Section 4.2 describes the problem statement and the research objectives. Section 4.3 describes the research database and sample selection. Section 4.4 discusses the potential biases of the research and their possible remedies. Section 4.5 presents an overview of the research methodology.

4.2 Problem Statement and Research Objectives

This research attempts to explore market anomalies by developing a series of linear and non-linear models to forecast the JSE equity returns over the period from 1 January 1997 to 31 December 2007. The expected return factor models are evaluated based on the pre-specified performance measurement criteria. The question to be answered includes whether robust expected return factor models can be developed from firm-specific attributes. If so, determine whether nonlinearity in the asset pricing relationship aids the power of the factor models in forecasting forward equity returns, and the successful identification of future outperforming and underperforming shares on the JSE. The desired outcome of the research is a blended stock selection model from the best linear and the best nonlinear model that achieves the highest attainable forecasting accuracy with diversified forecasting errors of the constituent models.

The objectives of this research are as follows:

1. Identify candidate firm-specific style attributes that potentially have the ability to explain the cross-sectional equity returns on the JSE over the examination period from 1 January 1997 to 31 December 2007.
2. Estimate the cross-sectional factor payoffs to the firm-specific attributes over the examination period, with the aim of determining their relative strengths in explaining cross-sectional sample share returns.
3. Develop style-based linear and nonlinear expected return factor models over the examination period based on pre-specified performance measurement criteria.
4. Determine the relative power of the expected return factor models in forecasting sample share returns over the examination period.
5. Construct a blended stock selection model from the best linear and the best nonlinear model identified in the multivariate analysis, and subsequently evaluate the performance of the blended model relative to that of the constituent models over the examination period.
6. Conduct fractile analysis on the factor models to determine the relative power of the factor models in classifying the best and the worst performers over the examination period.

4.3 Data and Sample Selection

The examination period for the research is from 01 January 1997 to 31 December 2007 (a total of 132 months). As of 31 December 2007, 159 shares comprising the FTSE/JSE All Share Index are selected as the sample shares for this research. The monthly data of the closing sample share prices, indices and style attributes are downloaded from DataStream International in the finance laboratory at the University of Cape Town.

Candidate style attributes are divided into six categories, namely, fundamental values relative to share price, solvency and liquidity, fundamental growth, operating performance, size and return momentum and consensus analyst forecast. Style variables in the fundamental values relative to share price category distinguish value stocks from growth (glamour) stocks. Value stocks have small fundamental values per share relative to their share prices compared to growth stocks. Solvency and liquidity ratios serve as indications for the companies' financial positions. Style variables in the fundamental growth category are historical growth rates in cash, earnings, profit margins, dividends and sales for the companies. Profit margins are indications of operational profitability while return on equity and return on assets are indications of sustainable profitability for the companies. The operating efficiency of the sample shares is indicated by the total asset turnover. Return momentum measures the growth rate in the total return index of the sample shares. The natural logarithm of share price and market capitalisation are indications of the current market values of sample shares. Style attributes in the size and return momentum category are price-sensitive attributes. Finally, attributes in the consensus analyst forecast category incorporates analyst forecasts regarding future earnings and dividends. The descriptors of the candidate style attributes are displayed in Table 4.1. The computations of the candidate style attributes are demonstrated in Appendix A.

Table 4.1 List of Firm-Specific Style Attributes

The 38 candidate style attributes are grouped into 6 categories, namely, fundamental values relative to share price, solvency and liquidity, fundamental growth, operating performance, size and return momentum and consensus analyst forecast. The computations of these attributes are displayed in Appendix A.
`

No.	Descriptor	Style Attribute	No.	Descriptor	Style Attribute
(I)	**FUNDAMENTAL VALUES RELATIVE TO SHARE PRICE**				
1.	BVTP	Book value-to-price	4.	EY	Trailing earnings yield
2.	CFTP	Cash flow-to-price	5.	SALESTP	Sales-to-price
3.	DY	Trailing dividend yield			
(II)	**SOLVENCY AND LIQUIDITY**				
6.	CFTCURRLIABS	Cash flow-to-current liabilities	9.	DEBTTMVE	Debt-to-market value of equity
7.	CFTDEBT	Cash flow-to-debt	10.	DEBTTBVE	Debt-to-book value of equity
8.	CURRENTRATIO	Current ratio	11.	ICBT	Interest coverage before tax
(III)	**FUNDAMENTAL GROWTH**				
12.	G12MCPS	12-month cash holdings growth	16.	G12MNPMARGIN	12-month net profit margin growth
13.	G12MDPS	12-month dividend growth			
14.	G12MEPS	12-month earnings growth	17.	G12MSALES	12-month sales growth
15.	G12MGPMARGIN	12-month gross profit margin growth	18.	G24MEPS	24-month earnings growth
			19.	Growth	Dividend growth rate
(IV)	**OPERATING PERFORMANCE**				
20.	GPMARGIN	Gross profit margin	23.	ROA	Return of assets
21.	NPMARGIN	Net profit margin	24.	ROE	Return on equity
22.	PAYOUT	Dividend payout ratio	25.	TATURNOVR	Total asset turnover
(V)	**SIZE AND RETURN MOMENTUM**				
26.	LAGLPRICE	Lagged log of market price	30.	MOM12	12-month return
27.	LPRICE	Log of market price	31.	MOM12-1	Lagged 11-month return
28.	LSIZE	Log of market capitalisation	32.	MOM24	24-month return
29.	MOM1	1-month return	33.	MOM3	3-month return
			34.	MOM6	6-month return
(VI)	**CONSENSUS ANALYST FORECAST**				
35.	EARNREV	Earnings forecast revision	37.	FOREY1	1-year forward earnings yield
36.	EG1	1-year forward earnings growth	38.	FOREY2	2-year forward earnings yield

4.4 Possible Biases in Research and Their Remedies

In order to ensure that the sample shares have sufficient liquidity, a turnover ratio is computed by dividing the average number of shares traded daily for a particular month by the total number of outstanding shares on the first day of the month. Shares with equal to or less than a turnover ratio of 0.01% are excluded for that month. This correction method is suggested by Van Rensburg and Robertson (2003a) in order to ensure that each sample share is traded at least once in any particular month. In addition to the thinly-traded shares, companies that exhibit major corporate restructuring, mergers and acquisitions, and share splits are excluded as of the date of the above mentioned corporate events. After filtering out non-representative shares based on the above procedures, the final sample of shares is nevertheless subject to survivorship bias since only shares listed on the JSE as at 31 December 2007 are considered in the initial sample. This bias is partially addressed through the retention of liquid shares in the research sample, which are generally more established firms that are less likely to be non-survivors.

Prior to the implementation of the univariate test, candidate style attributes are winsorised and standardised based on methods described in van Rensburg and Robertson (2003a). The maximum and minimum monthly values of each style variable are set at 99.5^{th} and 0.5^{th} percentiles to remove the extreme outliers over the examination period. Once the monthly style attributes are free of outliers, the monthly cross-sectional mean of each style attribute is subtracted from the cross-sectional distribution of each style attribute. The cross-sectional distribution of each style attribute is subsequently divided by its respective cross-sectional standard deviation. Repeating this procedure monthly enables the monthly distribution of each style variable to be normally distributed over the examination period. According to Van Rensburg and Robertson (2003a), standardising style attributes allows comparison to be made amongst style factor payoffs estimated in the cross-sectional regressions. Due to the fact that all style attributes are computed directly from the published financial statement information downloaded from DataStream International with the attributes lagging the corresponding share returns in the regression analysis, the studies conducted in this research are not subject to look-ahead bias.

4.5 Overview of Research Methodology

This paper first examines the sensitivities of the cross-sectional sample share returns to each of the style attributes over the examination period from 01 January 1997 to 31 December 2007. Based on the univariate linear regression model of Fama and Macbeth (1973), the monthly sample share returns are regressed on the lagged values (one-month) of each of the style attributes over the examination period. In this manner, the monthly cross-sectional payoffs (that is, the slope coefficient of the regression) for each of the style attributes are estimated over the examination period. The monthly average of each style payoff is subsequently subject to the Student's *t*-test in order to determine their relative significance in explaining the cross-sectional sample share returns.

Multivariate analysis entails developing multifactor linear and nonlinear models that best explain the cross-section of sample share returns. Following the methodology of Haugen and Baker (1996), in conjunction with the stepwise permutation procedure of Van Rensburg and Robertson (2003a), four linear expected return factor models are constructed from the style attributes that collectively maximise the in-sample scores of the four performance measurement criteria, namely the average information coefficient (Average IC), the slope *t*-statistic, the Grinold information ratio (Grinold IR) and the Qian and Hua information ratio (QH IR). On the other hand, the nonlinear expected return factor models are constructed using the cascade correlation artificial neural network (ANN) architecture. The ANN models are trained using either the embedded backpropagation learning rule or the embedded extended Kalman filter (EKF) learning rule, within the cascade correlation architecture.

The relative stock picking abilities of the factor models are established by conducting a fractile analysis proposed by Achour, Harvey, Hopkins and Lang (1999). Assuming that nonlinearity forms an integral component of the return determination process, the final stage of the thesis attempts to construct a blended stock selection model from the best linear and the best nonlinear models. The hypothesis being tested in this stage is whether such a model possesses superior stock-picking ability relative to the linear and nonlinear models, if they were to be used in isolation.

5

Style Attributes and the Cross-Section of Equity Returns: Univariate Linear Approach

5.1 Introduction

The purpose of the univariate tests is to identify potential style attributes that explain the cross-sectional equity returns statistically significantly. The first univariate test to be conducted involves estimating the time-series mean of the monthly factor payoff to each style attribute in isolation. Subsequently, the relative power of the candidate style attributes in explaining the cross-sectional equity returns are evaluated based on the magnitudes (irrespective of the signs) of the style attributes' mean factor payoffs. This procedure is often employed to filter out less important attributes as the preliminary step in building expected return factor models by setting a benchmark to judge whether the candidate attributes receive consistent rewards in the form of factor payoffs over the examination periods. For example, Van Rensburg and Robertson (2003a) eliminate attributes whose mean factor payoffs are insignificant at a 5% significance level using the Student's *t*-test, while Frankish (2004) and Moore (2008) discard attributes with *t*-statistics of less than one for the absolute values of their mean factor payoffs.

Besides the relative magnitudes of the payoffs to the style attributes, Haugen and Baker (1996) emphasise that important factors should also have consistent signs regarding their payoffs across independent examination periods. Moore (2008) also suggests that in order to ensure that the style attributes are fit for forecasting equity returns, it is necessary to conduct tests on the signs of the payoffs to the style attributes. Thus, non-parametric sign tests are conducted on all candidate style attributes to achieve this objective.

Another objective of conducting sign tests is to observe whether the style attributes explain the cross-sectional equity returns in the same direction as theoretically expected (Haugen and Baker (1996) and Serra (2003)). In addition, the cumulative geometric factor payoffs of these

attributes are illustrated graphically to enhance the comparison of their respective periodic results. The shape of the cumulative geometric factor payoffs to a particular style attribute not only illustrates the consistency of the size and signs of its payoffs over time, but also indicates the degree to which the payoffs are time-varying.

An important issue requiring further attention, as indicated by Van Rensburg and Robertson (2003a), is the possibility of multicollinearity in the construction of expected return factor models. An overfitted expected return factor model, as a consequence of multicollinearity amongst model inputs, often results in good in-sample, but poor out-of-sample scores. Achour, Harvey, Hopkins and Lang (1999) emphasise the need to examine the correlation structure of the style attributes in order to eliminate potentially redundant inputs when building expected return factor models. To address the issue of multicollinearity, the correlation matrix of style attributes in the examination period is constructed to identify the simultaneous inclusion of the pairs of style attributes that are potentially subject to multicollinearity in a factor model.

The univariate test is conducted on the candidate style attributes over two sub-periods: 01 January 1997 to 31 December 2001 and 01 January 2002 to 31 December 2007. Comparison of the sub-period results enhances the determination of the consistency of the candidate style attributes in explaining the cross-section of equity returns over time.

5.2 Descriptive Statistics and Methodology

The univariate factor model of Fama and Macbeth (1973) is employed to determine the relative strength (in terms of the univariate score) of each candidate style attribute in explaining the cross-section of equity returns as depicted in Equation 5.1.

$$R_{i,t+1} = a_{t+1} + b_{t+1} \times F_{i,t} + e_{i,t+1} \tag{5.1}$$

Where:

$R_{i,t+1}$ = the realised return on share *i* for month *t+1*;

b_{t+1} = the estimated cross-sectional factor payoff for the style attribute; and

$F_{i,t}$ = the lagged standardised value of the style attribute.

The monthly cross-sectional factor payoff (that is, b_{t+1}) to each style attribute in month $t + 1$ can hence be estimated by regressing the realised returns of all sample shares in month $t + 1$ on their respective lagged style attributes in month t using the ordinary least squares (OLS) approach. Note that the style attributes (model inputs) lag the sample share returns (model outputs) by one month to allow the market participants to act on the latest released information. As discussed in Chapter 4, this approach serves to mitigate the effect of look-ahead bias in the research.

Once the cross-sectional factor payoffs to the style attributes for all the months in the examination period (that is from 01 January 1997 to 31 December 2007) are estimated, the relative significance of the style attributes (that is, their univariate scores) are determined by conducting the Student's *t*-test on the time-series means of their respective factor payoffs for the sub-periods and the entire examination period. In assessing the style consistency of the candidate style attributes, a binomial sign test, Wilcoxon signed ranks test and Van der Waerden (normal scores) test are employed to perform this task. The binomial sign test assumes that the sample proportion above and below the median of a binomial distribution should be exactly 50 percent. Consequently, the null hypothesis of the binomial sign test is that the median of the factor payoff to a given style attribute is equal to zero. With regard to the Wilcoxon signed ranks test, the absolute value of the distance of each observation from the mean is calculated and subsequently ranked. Once the observations are ranked, the sum of

the ranks of the positive observations is compared to that of the negative observations to determine the strength of the direction of the reward to a given style attribute. On the other hand, the Van der Waerden test, derived from Wilcoxon signed ranks test, is based on smoothed ranks by converting ranks into quantiles of the normal distribution (see Conover, 1999 and Sheskin, 1997). Since these tests are tests that are based on the median of the distribution, the median of each style factor payoff being tested is set to zero. Thus, if the null hypothesis is rejected for a given style attribute, it is statistically more likely for the style attribute to receive either a consistent positive or negative payoff.

Finally, in an attempt to identify possible multicollinearity that may arise in building the expected return factor models, the monthly cross-correlations of the style attributes are computed and subsequently averaged over the examination period. The computation of the correlation coefficient between any two styles attributes (for example, X and Y) for any given month based on Pearson's correlation coefficient is as follows:

$$\rho_{X,Y} = \frac{\sigma_{X,Y}}{\sigma_X \times \sigma_Y} \tag{5.2}$$

Where:

$\sigma_{X,Y}$ = monthly covariance of style attribute *X* and style attribute *Y*;

$$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X})(Y_i - \overline{Y});$$

σ_X = monthly cross-sectional standard deviation of style attribute *X*;

$$= \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X});$$

σ_Y = monthly cross-sectional standard deviation of style attribute *Y*;

$$= \frac{1}{n-1}\sum_{i=1}^{n}(Y_i - \overline{Y});$$

n = number of sample shares comprising information pertaining to both style attributes (*X* and *Y*) in a particular month;

X_i and Y_i = the values of style attribute *X* and style attribute *Y* of share *i*; and

$\overline{X}$ and $\overline{Y}$ = the arithmetic averages of monthly cross-sectional style attribute *X* and style attribute *Y* respectively.

It is important to note that a significant correlation among independent variables is not a necessary condition for multicollinearity. DeFusco, McLeavey, Pinto and Runkle (2001) suggest that when the absolute value of the correlation coefficient between two variables is reasonably low, multicollinearity should not be a matter of concern. Following the decision rule of Van Rensburg and Janari (2008), potential multicollinearity problems are detected in the situation where the correlation coefficients between two or more factors are greater than the absolute value of 0.70. In such instance, the relatively more important attribute, as indicated by the order of the inclusion of the attributes in the stepwise regression (when building linear expected return factor models), will be retained in the expected return factor model.

As discussed earlier, eliminating candidate model inputs based on their univariate scores speeds up the construction of the expected return factor models. However, this method fails to take into account the joint explanatory power of the style attributes. Certain style attributes may not be able to distinguish cross-sectional equity returns in isolation, yet receive significant payoffs when employed with their complimentary counterparts. In other words, incorrect rejections may occur when a style attribute work better with other attributes. Since the pool of the style attributes are not unreasonably large, the style attributes will not be eliminated based on their mean factor payoffs at this stage. As a result, all 38 style attributes are employed as candidate inputs in the construction of the expected return factor models. The univariate scores of the style attributes that are included in the expected return factor models are re-examined to determine whether the univariate test results are effective indications of the relative significance of the style attributes in building expected return factor models.

5.3 Results

The characteristics of the factor payoffs to the candidate style attributes are demonstrated in Table 5.1. PANEL (A) through PANEL (F) of Table 5.1 displays the univariate test results of the style attributes in the fundamental values relative to share price, solvency and liquidity, fundamental growth, operating performance, size and return momentum and consensus analyst forecast categories respectively. Each panel consists of the time-series means of the factor payoffs to the style attributes within the category, their respective *t*-statistics, and the results of the various forms of the sign tests on the time-series median of the payoffs to each attribute.

Both the Student's *t*-test on the time-series means of the factor payoffs and the sign tests on the time-series median of the factor payoffs are set at 5 percent significance level. Significant *t*-statistics are highlighted in bold. Although the time-series mean of the factor payoffs to a particular style attribute may be statistically significant, the factor payoffs to the style attributes are not necessarily consistent (or robust) across different time-periods. Thus, the results of the sign tests are necessary in order to confirm the consistency of the payoff to a particular style attribute in cross-sectional equity pricing. In addition to the traditional binomial sign test, the results of Wilcoxon signed ranks test and Van Der Waerden sign test are also displayed in the Table 5.1. The null hypothesis is that the payoff to a particular style attribute does not exhibit a consistent sign. Thus, expected return factor models incorporating those style attributes that reject the null hypothesis of the sign tests are more likely to predict equity returns successfully.

Examining PANEL (A) through PANEL (F) of Table 5.1, there are 10 (out of 38) style attributes in total that exhibit significant *t*-statistics for the time-series mean of their factor payoffs over the first sub-period. Out of these 10 attributes, 3 attributes are from the fundamental values relative to share price category, 6 attributes are from the size and return momentum category, and the remaining attribute is from the consensus analyst forecast category. The stand-alone explanatory power of candidate style attributes improves drastically in the second sub-period. During this period, 20 attributes significantly explain the cross-sectional equity returns. Out of these 20 attributes, 4 attributes are from the

fundamental values relative to share price category, 1 attribute is from the solvency and liquidity category, 5 attributes are from the fundamental growth category, 7 attributes are from the size and return momentum category, and 3 attributes are from the consensus analyst forecast category. Although the majority (5 out of 8) of the style attributes in the fundamental growth category have significant mean factor payoffs over the second sub-period, none of the attributes from this category appears to be significant in the first sub-period.

Over the entire examination period from 01 January 1997 to 31 December 2007, all attributes from the fundamental values relative to share price category, 3 attributes from the fundamental growth category, 1 attribute from the solvency and liquidity category, 7 attributes from the size and return momentum category and all attributes from the consensus analyst forecast category demonstrate their abilities to explain the cross-sectional equity returns in their individual capacities. An important observation is that all attributes from the liquidity and solvency category and the operating performance category, with the exception of DEBTTBVE, are not able to explain the cross-sectional equity returns over all sub-periods as well as the overall examination period in their individual capacities. As discussed earlier in Section 5.1, this observation does not entail that the attributes in these two categories are unimportant in explaining cross-sectional equity returns since the attributes are examined in isolation. In the next chapter, the joint explanatory power of the attributes is examined under the stepwise regression procedure.

Examining the results of the median tests on the factor payoffs of the style attributes reveals that most of the significant attributes are associated with significant signs. In the situation where the style attributes are associated with signs that correspond to their theoretical expectations, the payoffs to these attributes are deemed to be fair. On the other hand, since the style attributes are lagged by one month of their realised returns, when the style attributes possess signs that are unexpected, their payoffs are either overvalued or undervalued, and short-term reversals (that is, market corrections) have taken place.

Another way of assessing the style consistency of the candidate style attributes is to determine whether the magnitudes and the directions of the style payoffs are consistent over time. This objective is achieved by observing the geometric cumulative factor payoffs of the

style attributes from month to month. The geometric cumulative factor payoffs over the entire examination period (01 January 1997 to 31 December 2007) for the 38 candidate style attributes are computed and depicted in Figure 5.1. The cumulative factor payoffs to style attributes in different categories are depicted in CHART (A) to CHART (F) respectively to facilitate the comparison between the rewards to the style attributes within the same category. As discussed in Chapter 4, the style attributes are standardised and thus their factor payoffs are directly comparable to each other.

5.3.1 Fundamental Values Relative to Share Price

The *t*-statistics of the mean factor payoffs to the style attributes in PANEL (A) of Table 5.1 demonstrate that all style attributes in this category (BVTP, CFTP, DY, EY and SALESTP) have significant mean payoffs over the examination period. In addition, all style attributes in this category possess significant positive signs from the fundamental values relative to share price category. This means that firms with relatively high fundamental values to their share prices (that is, the value firms) are temporarily undervalued, and the market will correct shortly and reward these firms with higher payoffs relative to that of the glamour firms.

The positive abnormal returns to the value firms are well documented in both local and international studies (see Page and Palmer (1993), Achour *et al* (1999) Van Rensburg (2001), Van Rensburg and Robertson (2003) and Rousseau and Van Rensburg (2004) for South African research; Basu (1977), Fama and French (1992), Lakonishok *et al* (1994), Haugen and Baker (1996), La Porta *et al* (1997), and Hanna and Ready (2005) for the U.S. studies; Van Rensburg and Janari (2008) for the Australian research; Serra (2002) for the emerging market studies; and Fama and French (1998) for the international research in the U.S. and EAFE countries).

Examining the cumulative geometric factor payoffs to the attributes in CHART (A) of Figure 5.1, the reward to BVTP, amongst all other fundamental indicators, has accelerated since the millennium and has taken a separate path from the rewards to other attributes. On the other hand, CFTP accumulates the second-highest payoff over the examination period. However,

the payoff to CFTP declines in the first few months of 2002 and accumulate at a steady rate towards the end of the examination period. Due to the fact that 01 January 2002 distinguishes the two sub-periods, the time-varying payoffs to BVTP and CFTP are also observed in the *t*-statistic of their respective mean factor payoffs in PANEL (A) of Table 5.1. The *t*-statistic of the mean factor payoff to BVTP increases from 3.669 in the first sub-period to 4.007 in the second sub-period. On the other hand, the *t*-statistic of the mean factor payoff to CFTP reduces from 3.935 in the first sub-period to 3.015 in the second sub-period.

Overall, the style attributes in this category have significant, consistent, positive factor payoffs over both sub-periods as well as the entire examination period, and hence are good candidates for model inputs for building expected return factor models. Although the payoffs to BVTP and CFTP changes in magnitudes from the first sub-period to the second sub-period, the cumulative geometric factor payoffs to these two attributes exhibit steady trends over individual sub-periods. Due to the fact that BVTP and CFTP accumulate the highest factor payoffs in the category, and their payoffs are not volatile, these two attributes are potential candidates for forecasting equity returns.

Table 5.1 Characteristics of Factor Payoffs to Style Attributes

PANEL (A) Fundamental Values Relative to Share Price

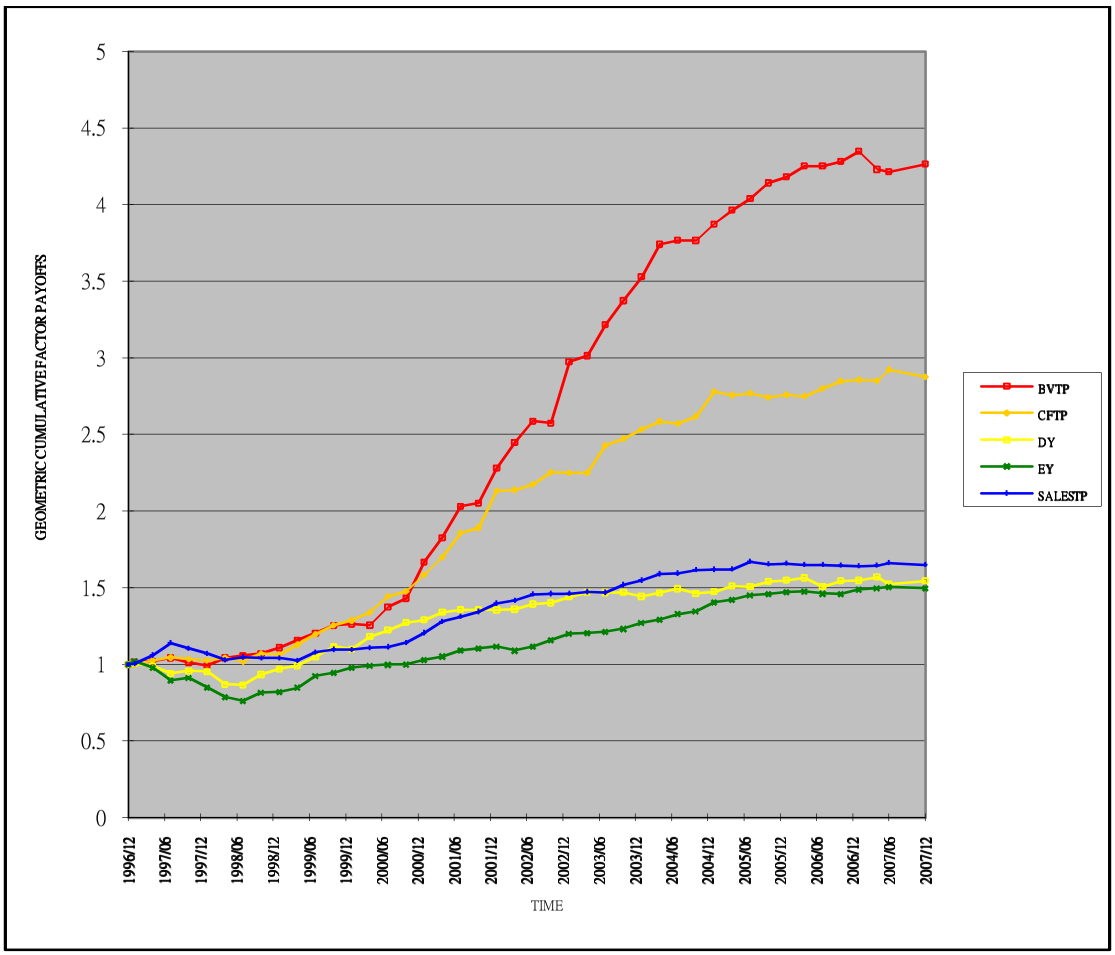
Table 5.1 displays the time-series means of the factor payoffs to the style attributes, their respective *t*-statistics, and the test results of various forms of sign tests on the time-series median of the payoffs to each attribute in the fundamental values relative to share price category over the two sub-periods and the entire examination period. The style attributes under examination are BVTP, CFTP, DY, EY and SALESTP. Values that are statistically significant at the 5 percent significance level are highlighted in bold. In order to determine the relative consistency of the factor payoff to each style attribute, both the sizes and the signs of the factor payoffs to the candidate style attributes must be estimated. The size of the factor payoff to a particular attribute is determined by the *t*-statistic of the attribute's mean payoff. On the other hand, a binomial sign test, Wilcoxon signed ranks test and Van der Waerden sign test are employed to determine the consistency of the signs associated with the factor payoffs. Sign tests are tests on the medians of the factor payoffs, and the null hypothesis is that a particular factor payoff does not exhibit a consistent sign. Therefore, the rejection of the null hypothesis for a particular sign test (highlighted in bold) means that there is a consistent positive/negative reward to the style attribute in question.

	BVTP	CFTP	DY	EY	SALESTP
SUB-PERIOD 1					
Factor Payoffs (97m01 to 01m12):	0.014	0.013	0.005	0.002	0.006
Mean Tests:					
t-statistic:	***3.669***	***3.935***	1.847	0.771	***2.298***
Rank:	2	1	13	24	8
Median Tests:					
Sign (exact binomial):	***39.000***	***42.000***	38.000	32.000	***42.000***
Sign (normal approximation):	***2.195***	***2.969***	1.936	0.387	***2.969***
Wilcoxon signed rank:	***3.213***	***3.677***	2.058	1.034	***2.551***
Van der Waerden (normal scores):	***3.334***	***3.690***	1.956	0.902	***2.392***
No. of Observations > 0:	39.000	42.000	38.000	32.000	42.000
No. of Observations < 0:	21.000	18.000	22.000	28.000	18.000
SUB-PERIOD 2					
Factor Payoffs (02m01 to 07m12):	0.010	0.005	0.002	0.005	0.003
Mean Tests:					
t-statistic:	***4.007***	***3.015***	1.314	***3.408***	***3.051***
Rank:	6	13	26	9	11
Median Tests:					
Sign (exact binomial):	***51.000***	***46.000***	44.000	***47.000***	44.000
Sign (normal approximation):	***3.570***	***2.339***	1.846	***2.585***	1.846
Wilcoxon signed rank:	***4.197***	***2.881***	1.731	***3.501***	***2.689***
Van der Waerden (normal scores):	***4.206***	***2.935***	1.361	***3.457***	***2.843***
No. of Observations > 0:	51	46	44	47	44
No. of Observations < 0:	21	26	28	25	28
WHOLE PERIOD					
Factor Payoffs (97m01 to 07m12):	0.012	0.009	0.004	0.003	0.004
Mean Tests:					
t-statistic:	***5.284***	***4.803***	***2.248***	***2.267***	***3.286***
Rank:	2	4	19	18	10
Median Tests:					
Sign (exact binomial):	***90.000***	***85.000***	***79.000***	76.000	***83.000***
Sign (normal approximation):	***4.187***	***3.831***	***2.762***	2.227	***3.474***
Wilcoxon signed rank:	***5.059***	***4.672***	***2.693***	***3.036***	***3.801***
Van der Waerden (normal scores):	***5.110***	***4.705***	***2.463***	***2.735***	***3.613***
No. of Observations > 0:	90	88	82	79	86
No. of Observations < 0:	42	44	50	53	46

Figure 5.1 Geometric Cumulative Payoffs to Style Attributes

CHART (A) Fundamental Values Relative to Share Price

The geometric cumulative payoffs to the style attributes in the fundamental values relative to share price category over the entire examination period from 01 January 1997 to 31 December 2007 are computed and depicted. Style attributes under examination are BVTP, CFTP, DY, EY and SALESTP. The factor payoffs to the style attributes are estimated by the Fama-MacBeth (1973) single factor model.



5.3.2 Solvency and Liquidity

Examining the characteristics of the payoffs to style attributes in the solvency and liquidity category (see PANEL (B) of Table 5.1), the majority of the attributes have failed to explain the cross-sectional returns over the examination period in their individual merits. DEBTTBVE, amongst other 5 attributes, is the only attribute that exhibits significant payoffs in the examination period. The significant negative payoffs to DEBTTBVE over the second sub-period and the overall examination period reveals that high leveraged firms with substantial financial risk do not impress investors, and are penalised with lower payoffs, compared to firms with lower financial risk.

The cumulative geometric factor payoffs to style attributes demonstrated in CHART (B) of Figure 5.1 indicate that the payoff to DEBTTBVE only exhibits consistent negative payoff from January 2002. The factor payoffs to other style attributes in this category (CFTCURRLIABS, CFTDEBT, CURRENTRATIO, DEBTTMVE and ICBT) not only exhibit inconsistent signs, but also extremely volatile over the examination period. Unless these are compliments to attributes in other categories in explaining cross-sectional equity returns, they do not add values in forecasting equity returns.

Table 5.1 Characteristics of Factor Payoffs to Style Attributes

PANEL (B) Solvency and Liquidity

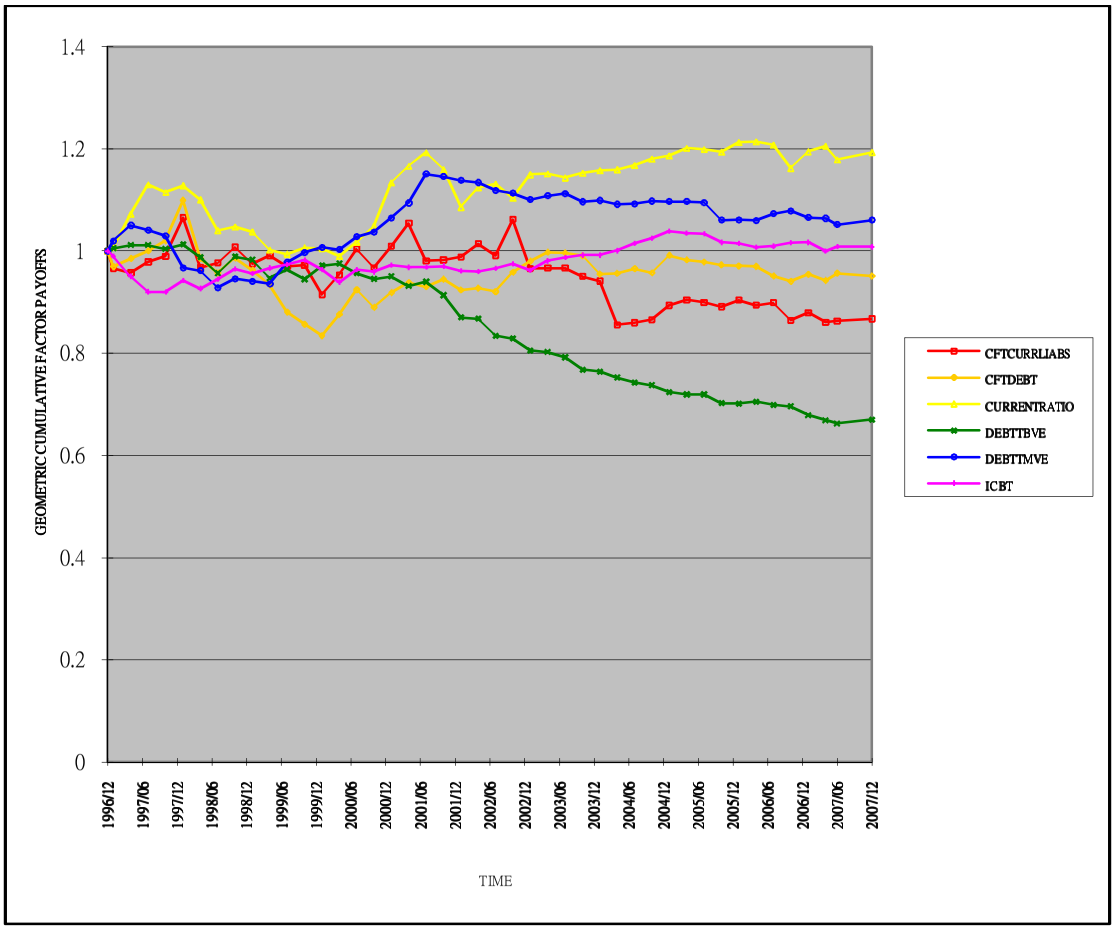
This table displays the time-series means of the factor payoffs to the style attributes, their respective *t*-statistics, and the test results of various forms of sign tests on the time-series median of the payoffs to each attribute in the solvency and liquidity category over the two sub-periods and the entire examination periods. The style attributes under examination are CFTCURRLIABS, CFTDEBT, CURRENTRATIO, DEBTTBVE, DEBTTMVE and ICBT. Values that are statistically significant at the 5 percent significance level are highlighted in bold. In order to determine the relative consistency of the factor payoff to each style attribute, both the sizes and the signs of the factor payoffs to the candidate style attributes must be estimated. The size of the factor payoff to a particular attribute is determined by the *t*-statistic of the attribute's mean payoff. On the other hand, a binomial sign test, Wilcoxon signed ranks test and Van der Waerden sign test are employed to determine the consistency of the signs associated with the factor payoffs. Sign tests are tests on the medians of the factor payoffs, and the null hypothesis is that a particular factor payoff does not exhibit a consistent sign. Therefore, the rejection of the null hypothesis for a particular sign test (highlighted in bold) means that there is a consistent positive/negative reward to the style attribute in question.

	CFTCURRLIABS	CFTDEBT	CURRENTRATIO	DEBTTBVE	DEBTTMVE	ICBT
SUB-PERIOD 1						
Factor Payoffs (97m01 to 01m12):	0.000	-0.001	0.002	-0.002	0.003	-0.001
Mean Tests:						
t-statistic:	0.029	-0.342	0.572	-0.857	1.127	-0.330
Rank:	38	33	28	22	19	34
Median Tests:						
Sign (exact binomial):	35.000	31.000	31.000	31.000	33.000	31.000
Sign (normal approximation):	1.162	0.129	0.129	0.129	0.645	0.129
Wilcoxon signed rank:	1.093	0.526	0.188	0.828	1.042	0.217
van der Waerden (normal scores):	0.674	0.191	0.060	-1.027	1.232	-0.204
No. of Observations > 0:	35.000	31.000	29.000	31.000	33.000	31.000
No. of Observations < 0:	25.000	29.000	31.000	29.000	27.000	29.000
SUB-PERIOD 2						
Factor Payoffs (02m01 to 07m12):	-0.002	0.001	0.001	-0.004	-0.001	0.001
Mean Tests:						
t-statistic:	-0.983	0.588	0.803	***-3.777***	-1.126	0.937
Rank:	29	34	32	7	28	30
Median Tests:						
Sign (exact binomial):	38.000	37.000	40.000	***52.000***	40.000	38.000
Sign (normal approximation):	0.369	0.123	0.862	***3.816***	0.862	0.369
Wilcoxon signed rank:	0.524	0.741	0.843	***3.711***	1.533	0.888
van der Waerden (normal scores):	-0.715	0.780	0.802	***-3.517***	-1.426	0.935
No. of Observations > 0:	34	37	40	20	32	38
No. of Observations < 0:	38	35	32	52	40	34
WHOLE PERIOD						
Factor Payoffs (97m01 to 07m12):	-0.001	0.000	0.002	-0.003	0.001	0.000
Mean Tests:						
t-statistic:	-0.476	-0.112	0.830	***-2.681***	0.393	0.130
Rank:	30	36	26	14	31	35
Median Tests:						
Sign (exact binomial):	69.000	68.000	69.000	***81.000***	67.000	69.000
Sign (normal approximation):	0.445	0.267	0.445	***2.584***	0.089	0.445
Wilcoxon signed rank:	0.657	1.023	0.270	***3.063***	0.071	0.438
van der Waerden (normal scores):	0.249	0.736	0.400	***-3.027***	0.297	0.402
No. of Observations > 0:	69	68	69	51	65	69
No. of Observations < 0:	63	64	63	81	67	63

Figure 5.1 Geometric Cumulative Payoffs to Style Attributes

CHART (B) Solvency and Liquidity

The geometric cumulative payoffs to the style attributes in the solvency and liquidity category over the entire examination period from 01 January 1997 to 31 December 2007 are computed and depicted. Style attributes under examination are CFTCURRLIABS, CFTDEBT, CURRENTRATIO, DEBTTBVE, DEBTTMVE and ICBT. The factor payoffs to the candidate style attributes are estimated by the Fama-MacBeth (1973) single factor model.



5.3.3 Fundamental Growth

PANEL (C) of Table 5.1 shows that none of the individual style attributes in the fundamental growth category are significant in explaining the cross-sectional equity returns in isolation over the in-sample period. However, 5 out of the 8 attributes in this category (G12MDPS, G12MEPS, G12MNPMARGIN, G24MEPS and GROWTH) exhibit significant payoffs over the second sub-period. Although only G12MDPS and G12MEPS pass all four sign tests in the overall examination period, the factor payoffs to the style attributes (with the exception of GROWTH) in this category are positive over the second sub-period and the entire examination period. This provides some degree of evidence that firms that have steady growth in fundamental values are rewarded by the market investors.

Examining the cumulative geometric factor payoffs to the style attributes in CHART (C) of Figure 5.1, it is apparent that the factor payoffs to the style attributes in this category are extremely volatile and change signs quite frequently over the examination period. This phenomenon is also observed in the second sub-period within which the majority of the attributes exhibit significant payoffs. This implies that the style attributes in this category are not fit for forecasting equity returns in their individual capacities.

Table 5.1 Characteristics of Factor Payoffs to Style Attributes

PANEL (C) Fundamental Growth

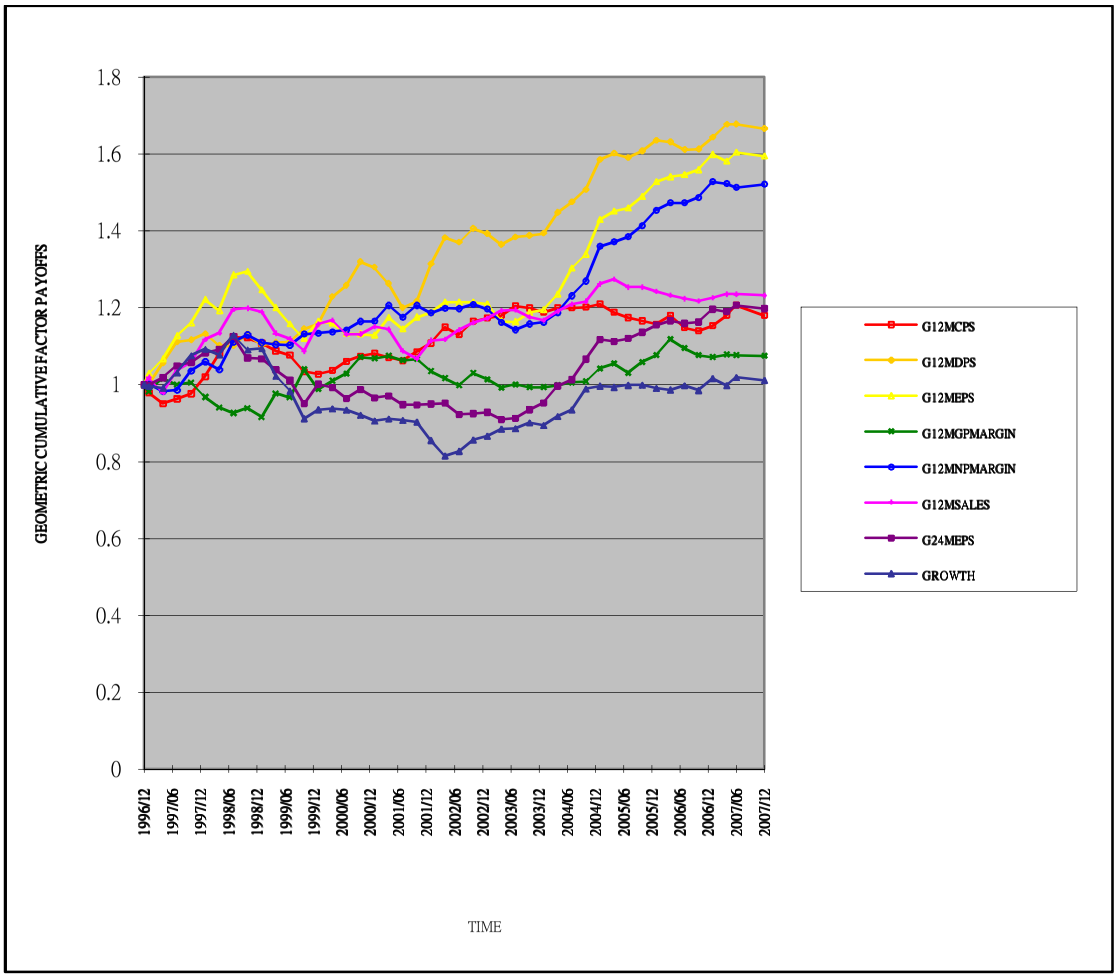
This table displays the time-series means of the factor payoffs to the style attributes, their respective *t*-statistics, and the test results of various forms of sign tests on the time-series median of the payoffs to each attribute in the fundamental growth category over the two sub-periods and the entire examination period. The style attributes under examination are G12MCPS, G12MDPS, G12MEPS, G12MGPMARGIN, G12MNPMARGIN, G12MSALES, G24MEPS and GROWTH. Values that are statistically significant at the 5 percent significance level are highlighted in bold. In order to determine the relative consistency of the factor payoff to each style attribute, both the sizes and the signs of the factor payoffs to the candidate style attributes must be estimated. The size of the factor payoffs to a particular attribute is determined by the *t*-statistic of the attribute's mean payoff. On the other hand, a binomial sign test, Wilcoxon signed ranks test and Van der Waerden sign test are employed to determine the consistency of the signs associated with the factor payoffs. Sign tests are tests on the medians of the factor payoffs, and the null hypothesis is that a particular factor payoff does not exhibit a consistent sign. Therefore, the rejection of the null hypothesis for a particular sign test (highlighted in bold) means that there is a consistent positive/negative reward to the style attribute in question.

	G12MCPS	G12MDPS	G12MEPS	G12MGPMARGIN	G12MNPMARGIN	G12MSALES	G24MEPS	GROWTH
SUB-PERIOD 1								
Factor Payoffs (97m01 to 01m12):	0.002	0.004	0.003	0.001	0.003	0.002	-0.001	-0.002
Mean Tests:								
t-statistic:	0.767	1.709	1.136	0.399	1.269	0.676	-0.194	-0.802
Rank:	25	14	18	31	15	26	36	23
Median Tests:								
Sign (exact binomial):	32.000	39.000	34.000	31.000	33.000	30.000	33.000	31.000
Sign (normal approximation):	0.387	2.195	0.904	0.129	0.645	-0.129	0.645	0.129
Wilcoxon signed rank:	0.637	1.734	1.115	0.151	0.836	0.320	0.306	0.718
van der Waerden (normal scores):	0.711	1.576	0.973	0.245	1.049	0.582	-0.047	-0.803
No. of Observations > 0:	32.000	39.000	34.000	31.000	33.000	30.000	33.000	29.000
No. of Observations < 0:	28.000	21.000	26.000	29.000	27.000	30.000	27.000	31.000
SUB-PERIOD 2								
Factor Payoffs (02m01 to 07m12):	0.002	0.004	0.005	0.000	0.004	0.002	0.004	0.002
Mean Tests:								
t-statistic:	1.256	***3.033***	***3.146***	0.314	***2.688***	1.547	***2.811***	***2.025***
Rank:	27	12	10	35	17	23	16	20
Median Tests:								
Sign (exact binomial):	39.000	42.000	***46.000***	39.000	41.000	38.000	***46.000***	42.000
Sign (normal approximation):	0.615	1.354	***2.339***	0.615	1.108	0.369	***2.339***	1.354
Wilcoxon signed rank:	0.952	***2.472***	***3.162***	0.134	2.229	1.163	***2.555***	1.999
van der Waerden (normal scores):	1.133	***2.682***	***3.115***	0.183	***2.458***	1.408	***2.629***	1.972
No. of Observations > 0:	39	42	46	33	41	38	46	42
No. of Observations < 0:	33	30	26	39	31	34	26	30
WHOLE PERIOD								
Factor Payoffs (97m01 to 07m12):	0.002	0.004	0.004	0.001	0.003	0.002	0.002	0.000
Mean Tests:								
t-statistic:	1.325	***3.005***	***2.645***	0.508	***2.543***	1.332	1.133	0.195
Rank:	23	12	15	29	16	22	25	33
Median Tests:								
Sign (exact binomial):	71.000	***81.000***	***77.000***	65.000	74.000	68.000	79.000	71.000
Sign (normal approximation):	0.802	***2.584***	***2.405***	0.267	1.336	0.267	2.227	0.802
Wilcoxon signed rank:	1.083	***3.099***	***2.866***	0.253	2.089	0.903	1.984	0.747
van der Waerden (normal scores):	1.206	***2.941***	***2.630***	0.342	***2.334***	1.172	1.521	0.491
No. of Observations > 0:	71	81	80	64	74	68	79	71
No. of Observations < 0:	61	51	52	68	58	64	53	61

Figure 5.1 Geometric Cumulative Payoffs to Style Attributes

CHART (C) Fundamental Growth

The geometric cumulative payoffs to the style attributes in the fundamental growth category over the entire examination period from 01 January 1997 to 31 December 2007 are computed and depicted. Style attributes under examination are G12MCPS, G12MDPS, G12MEPS, G12MGPMARGIN, G12MNPMARGIN, G12MSALES, G24MEPS and GROWTH. The factor payoffs to the candidate style attributes are estimated by the Fama-MacBeth (1973) single factor model.



5.3.4 Operating Performance

As discussed earlier, the significant attributes from the fundamental growth category are rewarded for the fast growth of the respective fundamental values (that is, DPS, EPS and NPMARGIN). However, firms are not rewarded significantly for the attributes that symbolise their most recent operating performance (GPMARGIN, NPMARGIN, PAYOUT, ROA, ROE and TATURNOVR). The characteristics of the factor payoffs in PANEL (D) of Table 5.1 show that TATURNOVR is the only attribute in this category that exhibits mild statistical significance (without passing any of the four sign tests) in the second sub-period. None of the attributes in this category receive significant payoffs, or pass any of the sign tests over the entire examination period.

As expected, the cumulative geometric factor payoffs to the style attributes in CHART (D) of Figure 5.1 reveals that the factor payoffs to the style attributes in this category are extremely volatile and inconsistent in their signs over the examination period. This implies that the style attributes in this category can only be used as one of the model inputs amongst attributes from other categories in the expected return forecasting models.

Table 5.1 Characteristics of Factor Payoffs to Style Attributes

PANEL (D) Operating Performance

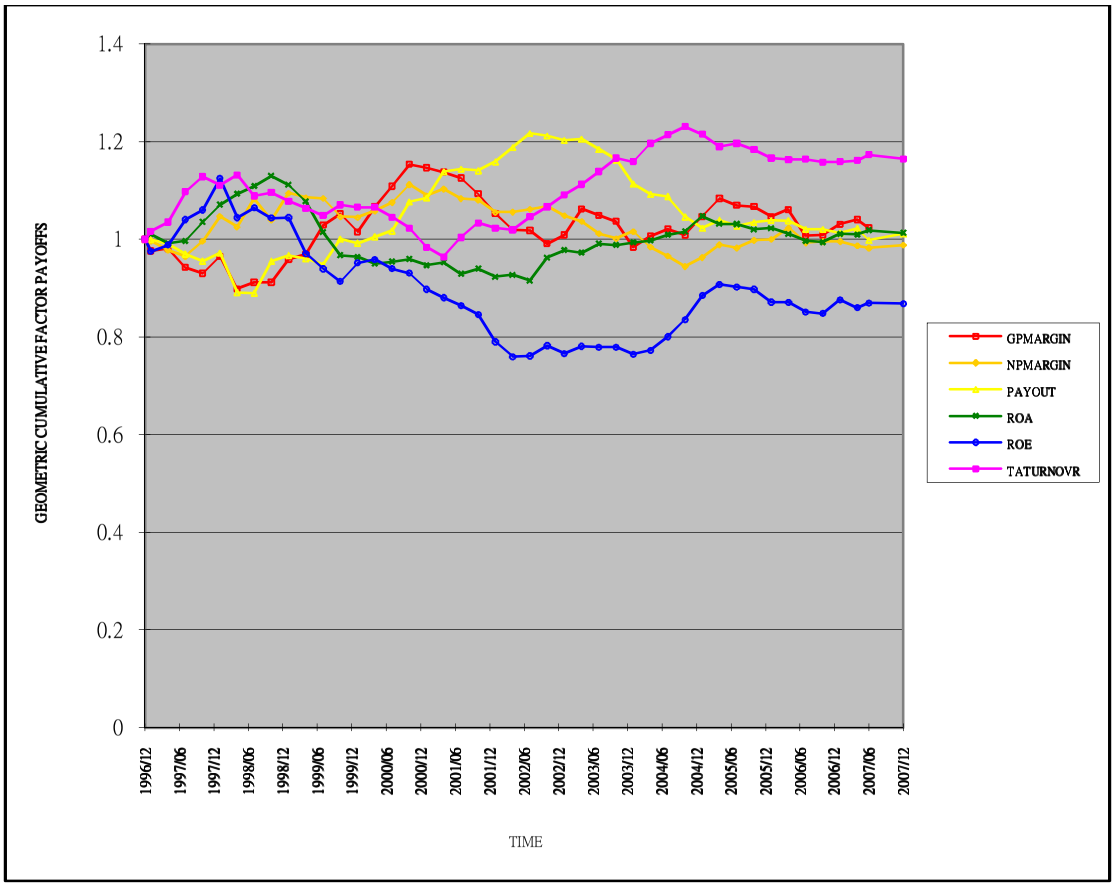
This table displays the time-series means of the factor payoffs to the style attributes, their respective *t*-statistics, and the test results of various forms of sign tests on the time-series median of the payoffs to each attribute in the operating performance category over the two sub-periods and the entire examination period. The style attributes under examination are GPMARGIN, NPMARGIN, PAYOUT, ROA, ROE and TATURNOVR. Values that are statistically significant at the 5 percent significance level are highlighted in bold. In order to determine the relative consistency of the factor payoffs to each style attribute, both the sizes and the signs of the factor payoffs to the candidate style attributes must be estimated. The size of the factor payoff to a particular attribute is determined by the *t*-statistic of the attribute's mean payoff. On the other hand, a binomial sign test, Wilcoxon signed ranks test and Van der Waerden sign test are employed to determine the consistency of the signs associated with the factor payoffs. Sign tests are tests on the medians of the factor payoffs, and the null hypothesis is that a particular factor payoff does not exhibit a consistent sign. Therefore, the rejection of the null hypothesis for a particular sign test (highlighted in bold) means that there is a consistent positive/negative reward to the style attribute in question.

	GPMARGIN	NPMARGIN	PAYOUT	ROA	ROE	TATURNOVR
SUB-PERIOD 1						
Factor Payoffs (97m01 to 01m12):	0.001	0.001	0.002	-0.001	-0.003	0.001
Mean Tests:						
t-statistic:	0.405	0.389	0.925	-0.565	-1.104	0.260
Rank:	30	32	21	29	20	35
Median Tests:						
Sign (exact binomial):	31.000	33.000	31.000	31.000	32.000	31.000
Sign (normal approximation):	0.129	0.645	0.129	0.129	0.387	0.129
Wilcoxon signed rank:	0.313	0.379	1.130	0.431	0.777	0.247
van der Waerden (normal scores):	0.375	-0.170	1.135	-0.581	-0.958	0.234
No. of Observations > 0:	31.000	27.000	31.000	29.000	28.000	29.000
No. of Observations < 0:	29.000	33.000	29.000	31.000	32.000	31.000
SUB-PERIOD 2						
Factor Payoffs (02m01 to 07m12):	0.000	-0.001	-0.002	0.002	0.001	0.002
Mean Tests:						
t-statistic:	-0.122	-0.873	-1.524	1.514	0.707	2.003
Rank:	38	31	24	25	33	21
Median Tests:						
Sign (exact binomial):	36.000	40.000	37.000	41.000	37.000	45.000
Sign (normal approximation):	-0.123	0.862	0.123	1.108	0.123	2.093
Wilcoxon signed rank:	0.351	0.862	0.856	1.258	0.747	2.229
van der Waerden (normal scores):	0.145	-0.864	-1.134	1.409	0.686	2.117
No. of Observations > 0:	36	31	37	41	37	45
No. of Observations < 0:	36	40	35	31	35	27
WHOLE PERIOD						
Factor Payoffs (97m01 to 07m12):	0.000	0.000	0.000	0.000	-0.001	0.001
Mean Tests:						
t-statistic:	0.217	-0.001	0.076	0.192	-0.575	1.165
Rank:	32	38	37	34	28	24
Median Tests:						
Sign (exact binomial):	67.000	73.000	68.000	70.000	67.000	74.000
Sign (normal approximation):	0.089	1.158	0.267	0.624	0.089	1.336
Wilcoxon signed rank:	0.409	0.894	0.085	0.523	0.102	1.337
van der Waerden (normal scores):	0.290	-0.697	0.030	0.291	-0.352	1.190
No. of Observations > 0:	67	59	68	70	65	74
No. of Observations < 0:	65	73	64	62	67	58

Figure 5.1 Geometric Cumulative Payoffs to Style Attributes

CHART (D) Operating Performance

The geometric cumulative payoffs to the style attributes in the profitability category over the entire examination period from 01 January 1997 to 31 December 2007 are computed and depicted. Style attributes under examination are GPMARGIN, NPMARGIN, PAYOUT, ROA, ROE and TATURNOVER. The factor payoffs to the candidate style attributes are estimated by the Fama-MacBeth (1973) single factor model.



5.3.5 Size and Return Momentum

PANEL (E) of Table 5.1 shows that with the exception of MOM3, the majority of the style attributes in this category exhibit significant payoffs over the two sub-periods. The payoffs to the three size attributes (LAGLPRICE, LPRICE and LSIZE) are both significant and consistently negative over the examination periods. This observation is consistent with the well-documented small-firm effects in both local and international research (see Van Rensburg (2001), Van Rensburg and Robertson (2003a) and Van Rensburg and Robertson (2004) for studies on the JSE; and Banz (1981), Reinganum (1981) and Fama and French (1992), and Haugen and Baker (1996) for research conducted on the U.S. markets).

On the other hand, the mean reversal of MOM1 observed in Van Rensburg and Robertson (2003a) is only found in the first sub-period, where the underperformance of the firms with relatively higher most recent return is documented. As suggested by Hodnett and Hsieh (2009), the factor payoffs to other indicators of short-term return momentum (MOM12, MOM12-1, MOM24 and MOM6) are associated with positive signs, indicating that there is still rooms for profits before the burst of speculative bubbles. The positive rewards to these momentum attributes might be even larger with the removal of shares with negative momentums as suggested by Fraser and Page (1999), who found that shares with negative momentum earn returns just below the momentum shares in the highest quintile. Subtracting the most recent 1-month momentum from MOM12, MOM12-1 appears to be the most rewarding short-term momentum attribute.

With regard to the style consistency of the attributes in this category (see CHART (E) of Figure 5.1), MOM3, amongst other attributes, is the only attribute that does not accumulate payoffs over the examination period. Overall, the payoffs to short-term momentum attributes are associated with clear trends, but volatile over time. This implies that after addressing the time-varying factor payoffs, the momentum attributes are potentially good candidates to model equity returns in their individual capacity. On the other hand, the size attributes (LAGLPRICE, LPRICE and LSIZE) are highly correlated with clear, steady negative trends.

Table 5.1 Characteristics of Factor Payoffs to Style Attributes

PANEL (E) Size and Return Momentum

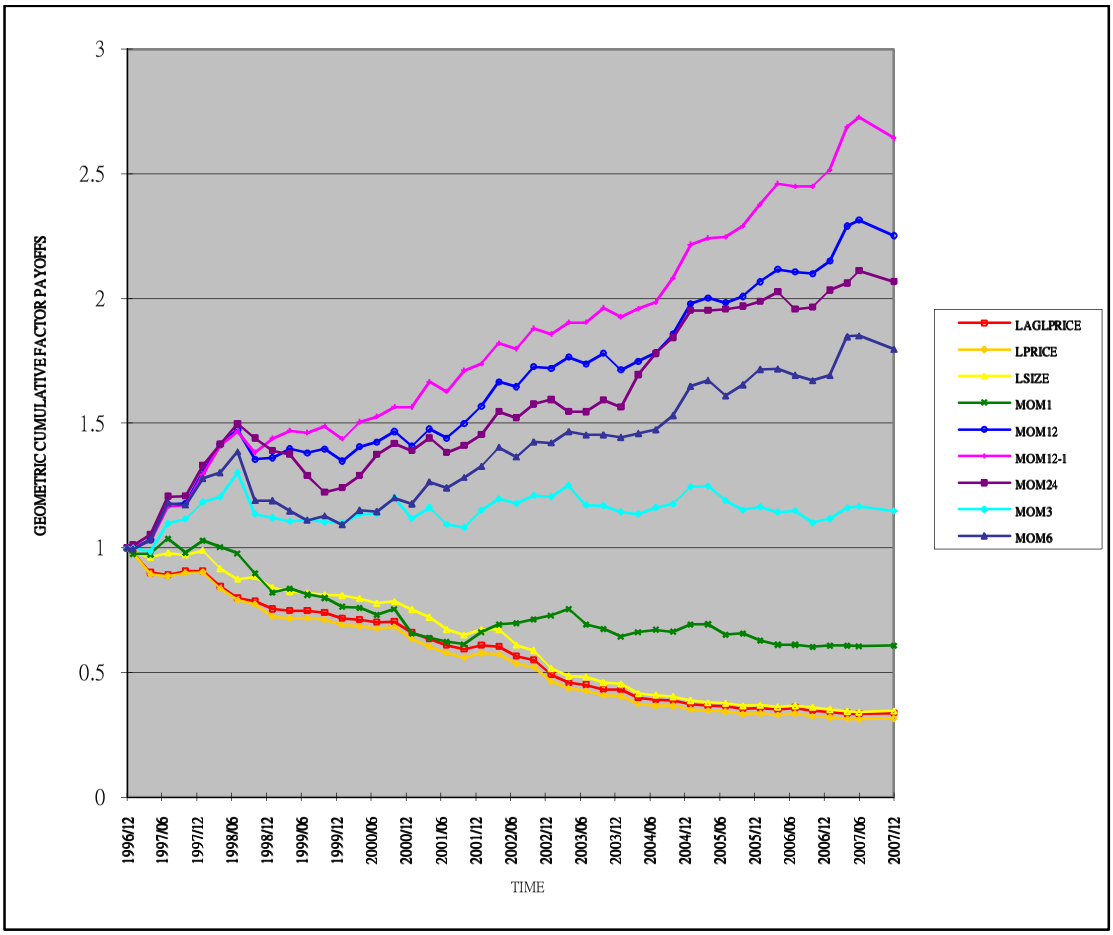
This table displays the time-series means of the factor payoffs to the style attributes, their respective *t*-statistics, and the test results of various forms of sign tests on the time-series median of the payoffs to each attribute in the size and return momentum category over the two sub-periods and the entire examination period. The style attributes under examination are LAGLPRICE, LPRICE, LSIZE, MOM1, MOM12, MOM12-1, MOM24, MOM3 and MOM6. Values that are statistically significant at the 5 percent significance level are highlighted in bold. In order to determine the relative consistency of the factor payoffs to each style attribute, both the sizes and the signs of the factor payoffs to the candidate style attributes must be estimated. The size of the factor payoff to a particular attribute is determined by the *t*-statistic of the attribute's mean payoff. On the other hand, a binomial sign test, Wilcoxon signed ranks test and Van der Waerden sign test are employed to determine the consistency of the signs associated with the factor payoffs. Sign tests are tests on the medians of the factor payoffs, and the null hypothesis is that a particular factor payoff does not exhibit a consistent sign. Therefore, the rejection of the null hypothesis for a particular sign test (highlighted in bold) means that there is a consistent positive/negative reward to the style attribute in question.

	LAGLPRICE	LPRICE	LSIZE	MOM1	MOM12	MOM12-1	MOM24	MOM3	MOM6
SUB-PERIOD 1									
Factor Payoffs (97m01 to 01m12):	-0.008	-0.009	-0.006	-0.007	0.007	0.009	0.006	0.002	0.005
Mean Tests:									
t-statistic:	***-2.800***	***-3.009***	***-2.019***	***-2.061***	***2.454***	***3.242***	1.866	0.661	1.236
Rank:	6	5	11	10	7	3	12	27	16
Median Tests:									
Sign (exact binomial):	36.000	37.000	39.000	35.000	36.000	***41.000***	37.000	32.000	36.000
Sign (normal approximation):	1.420	1.678	2.195	1.162	1.420	***2.711***	1.678	0.387	1.420
Wilcoxon signed rank:	***2.558***	***2.646***	1.991	1.616	***2.345***	***3.132***	2.072	1.115	1.366
van der Waerden (normal scores):	***-2.689***	***-2.746***	-1.971	-1.901	***2.373***	***3.093***	1.903	1.102	1.279
No. of Observations > 0:	24.000	23.000	21.000	25.000	36.000	41.000	37.000	32.000	36.000
No. of Observations < 0:	36.000	37.000	39.000	35.000	24.000	19.000	23.000	28.000	24.000
SUB-PERIOD 2									
Factor Payoffs (02m01 to 07m12):	-0.009	-0.009	-0.010	-0.001	0.006	0.007	0.006	0.001	0.006
Mean Tests:									
t-statistic:	***-5.098***	***-5.075***	***-5.131***	-0.295	***2.824***	***3.417***	***2.891***	0.302	***2.569***
Rank:	2	3	1	37	15	8	14	36	18
Median Tests:									
Sign (exact binomial):	***51.000***	***53.000***	***51.000***	43.000	***50.000***	***50.000***	***47.000***	37.000	38.000
Sign (normal approximation):	***3.570***	***4.062***	***3.570***	1.600	***3.323***	***3.323***	***2.585***	0.123	0.369
Wilcoxon signed rank:	***4.395***	***4.478***	***4.491***	0.805	***3.060***	***3.469***	***2.753***	0.600	2.153
van der Waerden (normal scores):	***-4.315***	***-4.363***	***-4.419***	-0.373	***2.876***	***3.298***	***2.775***	0.544	***2.375***
No. of Observations > 0:	21	19	21	29	50	50	47	37	38
No. of Observations < 0:	51	53	51	43	22	22	25	35	34
WHOLE PERIOD									
Factor Payoffs (97m01 to 07m12):	-0.008	-0.009	-0.008	-0.004	0.007	0.008	0.006	0.001	0.005
Mean Tests:									
t-statistic:	***-5.227***	***-5.365***	***-4.626***	-1.809	***3.702***	***4.681***	***3.206***	0.720	***2.442***
Rank:	3	1	6	21	7	5	11	27	17
Median Tests:									
Sign (exact binomial):	***87.000***	***90.000***	***90.000***	78.000	***86.000***	***91.000***	***84.000***	69.000	74.000
Sign (normal approximation):	***3.653***	***4.187***	***4.187***	2.049	***3.474***	***4.365***	***3.118***	0.445	1.336
Wilcoxon signed rank:	***4.787***	***4.894***	***4.414***	1.607	***3.747***	***4.607***	***3.389***	1.220	***2.496***
van der Waerden (normal scores):	***-4.760***	***-4.821***	***-4.267***	-1.645	***3.630***	***4.468***	***3.232***	1.160	***2.484***
No. of Observations > 0:	45	42	42	54	86	91	84	69	74
No. of Observations < 0:	87	90	90	78	46	41	48	63	58

Figure 5.1 Geometric Cumulative Payoffs to Style Attributes

CHART (E) Size and Return Momentum

The geometric cumulative payoffs to the style attributes in the size and return momentum category over the entire examination period from 01 January 1997 to 31 December 2007 are computed and depicted. Style attributes under examination are LAGLPRICE, LPRICE, LSIZE, MOM1, MOM12, MOM12-1, MOM24, MOM3 and MOM6. The factor payoffs to the candidate style attributes are estimated by the Fama-MacBeth (1973) single factor model.



5.3.6 Consensus Analyst Forecast

The data for the attributes in this category is only available from 1999. The characteristics of the payoffs to the style attributes displayed in PANEL (F) of Table 5.1 show that all four analyst forecasts (EARNREV, EG1, FOREY1 and FOREY2) are rewarded with significant payoffs. This implies that either the analysts have superior abilities in forecasting earnings, or investors, to some degree, follow the recommendations of the analysts.

Examining the style consistency of the attributes in this category reveals that only FOREY1 and FOREY2 passed all 4 sign tests in the overall examination period. In addition, the cumulative geometric factor payoffs to EG1 and FOREY1 seem to have steady trends in the second sub-period (see CHART (F) of Figure 5.1).

Table 5.1 Characteristics of Factor Payoffs to Style Attributes

PANEL (F) Consensus Analyst Forecast

This table displays the time-series means of the factor payoffs to the style attributes, their respective *t*-statistics, and the test results of various forms of sign tests on the time-series median of the payoffs to each attribute in the consensus analyst forecast category over the two sub-periods and the entire examination period. The style attributes under examination are EARNREV, EG1, FOREY1 and FOREY2. Values that are statistically significant at the 5 percent significance level are highlighted in bold. In order to determine the relative consistency of the factor payoffs to each style attribute, both the sizes and the signs of the factor payoffs to the candidate style attributes must be estimated. The size of the factor payoff to a particular attribute is determined by the *t*-statistic of the attribute's mean payoff. On the other hand, a binomial sign test, Wilcoxon signed ranks test and Van der Waerden sign test are employed to determine the consistency of the signs associated with the factor payoffs. Sign tests are tests on the medians of the factor payoffs, and the null hypothesis is that a particular factor payoff does not exhibit a consistent sign. Therefore, the rejection of the null hypothesis for a particular sign test (highlighted in bold) means that there is a consistent positive/negative reward to the style attribute in question.

	EARNREV	EG1	FOREY1	FOREY2
SUB-PERIOD 1				
Factor Payoffs (97m01 to 01m12):	0.008	0.001	0.005	0.008
Mean Tests:				
t-statistic:	***2.239***	0.148	1.217	***3.113***
Rank:	9	37	17	4
Median Tests:				
Sign (exact binomial):	21.000	22.000	20.000	26.000
Sign (normal approximation):	0.658	0.811	0.162	2.109
Wilcoxon signed rank:	1.720	1.001	0.812	***2.784***
van der Waerden (normal scores):	1.979	-0.685	1.106	***2.903***
No. of Observations > 0:	21.000	16.000	18.000	26.000
No. of Observations < 0:	16.000	22.000	20.000	12.000
SUB-PERIOD 2				
Factor Payoffs (02m01 to 07m12):	0.003	0.005	0.006	0.002
Mean Tests:				
t-statistic:	1.958	***4.191***	***4.913***	***2.211***
Rank:	22	5	4	19
Median Tests:				
Sign (exact binomial):	36.000	41.000	***46.000***	38.000
Sign (normal approximation):	0.615	1.846	***3.077***	1.108
Wilcoxon signed rank:	1.335	***3.482***	***4.331***	2.198
van der Waerden (normal scores):	1.676	***3.693***	***4.357***	2.214
No. of Observations > 0:	36	41	46	38
No. of Observations < 0:	30	25	20	28
WHOLE PERIOD				
Factor Payoffs (97m01 to 07m12):	0.005	0.003	0.006	0.004
Mean Tests:				
t-statistic:	***2.968***	***1.993***	***3.487***	***3.690***
Rank:	13	20	9	8
Median Tests:				
Sign (exact binomial):	57.000	57.000	***64.000***	***64.000***
Sign (normal approximation):	0.985	0.883	***2.255***	***2.255***
Wilcoxon signed rank:	2.045	1.769	***3.420***	***3.325***
van der Waerden (normal scores):	***2.523***	1.870	***3.491***	***3.481***
No. of Observations > 0:	57	57	64	64
No. of Observations < 0:	46	47	40	40

Figure 5.1 Geometric Cumulative Payoffs to Style Attributes

CHART (F) Consensus Analyst Forecast

The geometric cumulative payoffs to the style attributes in the consensus analyst forecast category over the entire examination period from 01 January 1997 to 31 December 2007 are computed and depicted. Style attributes under examination are EARNREV, EG1, FOREY1 and FOREY2. The factor payoffs to the candidate style attributes are estimated by the Fama-MacBeth (1973) single factor model.

After examining the characteristics of the factor payoffs to the candidate style attributes in isolation. The possibility of multicollinearity that may arise as a result of two or more factors being highly correlated with each other must be determined before examining the joint explanatory power of the candidate style attributes. A consequence of multicollinearity is that it *"it becomes practically impossible to distinguish the individual impacts of the independent variables on the dependent variable"* (DeFusco, McLeavey, Pinto and Runkle, 2004: 473). Table 5.2 demonstrates the mean of the monthly correlation coefficients between each pair of the candidate style attributes over the examination period. Values highlighted in grey represent statistically significant correlation between the two attributes. Correlation coefficients of greater than 0.70 or less than -0.70 are in white font and highlighted in black as an indication for possible multicollinearity in the construction of multifactor expected return models.

Examining the monthly mean correlation coefficients between candidate style attributes in Table 5.2, it is apparent that style attributes within the same designated category exhibit statistically significant correlation. DEBTTMVE and DEBTTBVE, FOREY1 and FOREY2, LAGLPRICE, LPRICE and LSIZE, MOM6 and MOM12, MOM6 and MOM3, and ROE and Growth are the only pairs of style attributes with greater than 0.70 for the absolute values of their mean monthly correlation coefficients. To address the problem of multicollinearity, the attribute being included in the stepwise regression in the latter iteration will be removed from the linear expected return factor model.

Table 5.2 Correlation Matrix of Style Attributes

The mean of the monthly correlation coefficients between style attributes are computed and averaged over the examination period from 01 January 1997 to 31 December 2007. Values shaded in grey represents statistically significant correlation between the two attributes. Correlation coefficients of greater than 0.70 or less than -0.70 are shaded in black as an indication for possible multicollinearity in the construction of multifactor expected return models.

	CFTCURRLIABS	CFTDEBT	CFTP	CURRENTRATIO	DEBTTBVE	DEBTTMVE	DY	EARNREV	EG1	EY	FOREY1	FOREY2	G12MCPS	G12MDPS	G12MEPS	G12MGPMARGIN	G12MNPMARGIN	G12MSALES	G24MEPS	GPMARGIN	GROWTH	ICBT	LAGLPRICE	LPRICE	LSIZE	MOM1	MOM12	MOM12-1	MOM24	MOM3	MOM6	NPMARGIN	PAYOUT	ROA	ROE	SALESTP	TATURNOVR
BVTP	-0.151	-0.114	0.375	0.149	-0.121	0.187	0.306	-0.063	-0.068	0.351	0.019	0.187	-0.033	-0.173	-0.082	-0.004	-0.007	-0.060	-0.116	0.087	-0.239	-0.030	-0.368	-0.375	-0.398	-0.065	-0.161	-0.154	-0.197	-0.106	-0.136	0.113	0.129	-0.208	-0.218	0.184	-0.213
CFTCURRLIABS		0.575	0.102	0.096	-0.134	-0.164	0.124	-0.011	-0.041	-0.001	-0.059	0.172	0.089	0.074	0.051	-0.007	0.030	0.040	0.013	0.274	0.059	0.200	0.291	0.292	0.235	0.007	0.021	0.023	0.084	-0.006	-0.010	0.353	0.128	0.369	0.271	-0.210	-0.247
CFTDEBT			0.107	0.011	-0.185	-0.234	0.188	0.020	-0.038	-0.008	-0.020	0.072	0.153	0.081	0.068	0.009	0.079	0.106	0.039	0.246	0.131	0.067	0.012	0.012	-0.010	0.013	0.062	0.068	0.134	0.002	0.017	0.264	0.138	0.483	0.224	-0.150	0.004
CFTP				-0.120	0.073	0.256	-0.013	0.013	-0.009	0.165	0.077	0.049	-0.011	0.016	-0.032	0.009	0.004	-0.017	-0.033	-0.085	-0.019	0.019	-0.092	-0.093	-0.070	-0.009	-0.022	-0.023	-0.063	-0.005	-0.007	-0.140	-0.133	-0.074	0.003	0.200	0.043
CURRENTRATIO					-0.191	-0.145	0.063	-0.018	-0.030	0.010	0.013	0.207	-0.073	0.035	-0.028	0.010	0.002	-0.090	-0.020	0.155	-0.029	0.066	0.029	0.030	-0.092	0.010	0.022	0.024	0.009	0.010	0.020	0.237	0.016	0.106	-0.016	-0.005	-0.116
DEBTTBVE						**0.762**	-0.084	-0.043	-0.051	-0.020	-0.068	-0.086	-0.089	0.011	-0.031	0.062	-0.033	-0.051	-0.021	-0.161	-0.010	-0.074	0.129	0.127	0.265	-0.032	-0.086	-0.081	-0.089	-0.054	-0.066	-0.134	-0.080	-0.287	-0.025	0.080	-0.113
DEBTTMVE							-0.028	-0.041	-0.096	0.132	-0.027	0.001	-0.091	-0.064	-0.073	0.011	-0.043	-0.059	-0.070	-0.169	-0.071	-0.055	0.012	0.008	0.095	-0.050	-0.138	-0.135	-0.162	-0.083	-0.110	-0.156	-0.090	-0.325	-0.133	0.229	-0.143
DY								-0.084	-0.047	0.481	0.121	0.084	-0.060	0.117	-0.097	-0.024	0.007	0.066	-0.057	0.577	-0.362	0.022	-0.306	-0.311	-0.349	-0.039	-0.167	-0.156	-0.143	-0.122	-0.157	0.423	0.627	0.143	0.020	-0.050	-0.281
EARNREV									0.054	-0.066	0.032	-0.035	-0.027	-0.032	0.032	-0.013	0.025	0.067	0.036	-0.004	-0.010	0.023	-0.013	-0.015	-0.002	0.031	0.063	0.066	0.044	0.062	0.070	-0.050	-0.049	-0.012	0.012	-0.027	0.048
EG1										-0.183	**0.780**	0.082	0.176	0.016	-0.053	0.024	-0.047	0.003	-0.060	0.080	-0.044	-0.004	-0.030	-0.028	0.015	0.035	0.145	0.140	0.166	0.047	0.082	-0.015	0.168	-0.075	0.042	0.214	-0.053
EY											0.138	0.187	-0.019	-0.010	0.191	-0.004	0.076	-0.008	0.155	0.132	0.083	0.017	-0.259	-0.267	-0.346	-0.076	-0.161	-0.147	-0.123	-0.136	-0.162	0.153	-0.139	0.153	0.039	0.151	-0.051
FOREY1												0.189	0.056	0.049	0.018	0.010	0.014	0.009	0.022	0.094	0.039	0.027	-0.144	-0.146	-0.105	-0.005	0.038	0.040	0.082	0.001	0.017	0.009	-0.019	-0.002	0.087	0.330	-0.022
FOREY2													-0.009	0.044	0.034	0.057	0.034	-0.069	0.020	-0.087	-0.055	-0.003	-0.052	-0.051	-0.072	0.002	0.022	0.025	-0.019	0.011	0.034	0.159	-0.049	0.083	-0.116	0.069	-0.053
G12MCPS														0.044	0.205	0.075	0.073	0.167	0.132	0.001	0.136	0.000	-0.003	-0.003	0.021	0.019	0.122	0.123	0.218	0.031	0.062	0.001	-0.073	0.083	0.108	-0.012	0.023
G12MDPS															0.329	0.129	0.160	0.099	0.354	0.045	0.013	0.050	0.081	0.085	0.093	0.066	0.258	0.253	0.327	0.113	0.162	0.017	0.077	0.108	0.146	-0.017	0.095
G12MEPS																0.208	0.589	0.147	0.555	-0.010	0.247	0.040	0.037	0.042	0.064	0.063	0.309	0.310	0.374	0.116	0.184	0.031	-0.247	0.154	0.195	-0.018	0.053
G12MGPMARGIN																	0.100	0.040	0.149	0.061	-0.009	-0.005	0.004	0.004	-0.018	0.003	0.086	0.091	0.116	0.013	0.050	-0.075	-0.067	-0.032	0.002	0.034	0.080
G12MNPMARGIN																		-0.117	0.269	-0.058	0.092	0.040	0.013	0.016	0.029	0.051	0.239	0.240	0.251	0.094	0.150	0.175	-0.059	0.145	0.092	-0.035	-0.006
G12MSALES																			0.194	0.101	0.063	-0.035	-0.111	-0.109	-0.035	0.022	0.120	0.122	0.247	0.039	0.066	0.051	0.015	0.118	0.087	-0.064	-0.041
G24MEPS																				-0.012	0.295	0.049	0.068	0.070	0.061	0.033	0.259	0.265	0.467	0.082	0.155	0.060	-0.217	0.232	0.252	-0.042	0.049
GPMARGIN																					-0.142	0.047	-0.150	-0.150	-0.174	0.010	-0.004	-0.002	0.030	-0.019	-0.018	0.433	0.425	0.100	0.107	-0.229	-0.438
GROWTH																						0.049	0.085	0.087	0.096	0.015	0.150	0.154	0.283	0.063	0.091	0.028	-0.534	0.285	**0.759**	-0.003	0.203
ICBT																							0.063	0.063	0.024	0.002	0.011	0.013	0.019	0.001	-0.001	0.042	0.023	0.140	0.097	-0.048	-0.018
LAGLPRICE																								**0.997**	**0.726**	-0.074	0.018	0.035	0.046	-0.031	-0.013	-0.077	-0.064	0.035	0.161	-0.184	-0.064
LPRICE																									**0.728**	0.000	0.039	0.039	0.060	0.011	0.017	-0.078	-0.065	0.035	0.162	-0.184	-0.062
LSIZE																										-0.020	-0.009	-0.007	-0.001	-0.021	-0.024	-0.145	-0.090	-0.051	0.138	-0.195	-0.152
MOM1																											0.286	0.062	0.181	0.564	0.401	0.004	0.001	0.017	0.016	-0.008	0.021
MOM12																												0.964	0.692	0.514	**0.733**	0.029	-0.102	0.123	0.122	-0.021	0.075
MOM12-1																													0.674	0.395	0.660	0.034	-0.102	0.127	0.128	-0.020	0.071
MOM24																														0.328	0.474	0.076	-0.131	0.246	0.234	-0.035	0.062
MOM3																															**0.704**	-0.005	-0.044	0.028	0.029	-0.005	0.057
MOM6																																0.001	-0.068	0.053	0.046	-0.006	0.071
NPMARGIN																																	0.304	0.386	0.057	-0.192	-0.322
PAYOUT																																		-0.027	-0.089	-0.128	-0.260
ROA																																			0.285	-0.212	0.108
ROE																																				-0.052	0.144
SALESTP																																					0.330

5.4 Conclusion

The univariate tests consist of tests on the mean, the median, and the correlations of the time-series factor payoffs to the style attributes in each category over the two sub-periods and the overall examination period. The results of the univariate tests give indication as to whether the monthly payoffs to the candidate attributes are significant and associated with consistent signs. The cumulative factor payoffs to each style attribute are geometrically linked and graphed over the examination period to enhance the analysis of the style consistency.

The style attributes that pass the criteria specified in the univariate tests are able to explain the cross-sectional equity returns in a consistent manner at their individual capacities over the examination period, and are potentially good candidates for modeling future equity returns. On the contrary, insignificant univariate scores indicate that the style attributes in question cannot explain the cross-sectional equity returns in isolation. In addition, these attributes only have the power to model equity returns if they are compliments to other attributes and must be used in conjunction with their compliments in modeling equity returns. The results of the univariate tests show that attributes in the fundamental values relative to share price category, the fundamental growth, the size and momentum category and the consensus analyst forecast category are able to explain the cross-sectional equity returns in their individual capacities. It should be noted that the number of significant attributes increases drastically between the first sub-period (10 attributes) and the second sub-period period (21 attributes). The significant attributes, in general, are associated with consistent signs for their payoffs, and their cumulative factor payoffs are not volatile over the examination period.

The consistent signs of the payoffs to the style attributes gives indications as to whether the payoffs are fairly priced as theoretically expected, or mispriced. When the signs of the payoffs to the attributes are not in line with their theoretical expectation, the apparent market anomalies are documented and the market corrections/reversals have taken place. Important market anomalies found in the univariate tests include the value effect, the size effect, and the short-term momentum effect. In addition, there

seems to be a mild reversal of one-month momentum in the first sub-period. Firms that have relatively high fundamental values relative to their share prices are deemed undervalued and are rewarded positively. Amongst all attributes in the fundamental values relatively to share price category, BVTP and CFTP are the attributes that accumulate the highest and the second highest payoffs over the examination periods respectively. An important observation is that BVTP seems to receive higher payoffs, and CFTP receives lower payoffs in the second sub-period compared to their payoffs in the first sub-period. The time-varying characteristics of the factor payoffs are also observed for the attributes in the fundamental growth, the return momentum and the consensus analyst forecast categories. Although there are clear upward trends for the factor payoffs to short-term momentum attributes, the payoffs to these attributes appear to be volatile from month to month. These findings suggest that the time-varying payoffs to the attributes must be taken into account in order to model equity returns successfully.

Besides the market anomalies documented above, the results of the univariate tests show that investors do not pay much attention to the immediate operating performance such as profit margins, dividend payouts and earnings. It is, however, the growth in these fundamental values such as G12DPS, G12MEPS G12MNPMARGIN that matters to the investors. Firms that show consistent growth in their fundamental operating performance seem to be rewarded with higher returns over the examination period. It is also noted that the consensus analyst forecasts regarding EPS seems to be more rewarding in the second sub-period than in the first sub-period.

Finally, although the majority of the attributes in the solvency and liquidity category and the operating performance category do not appear to receive significant payoffs over the examination period on their individual merits, they may well be good complements to other attributes in modeling equity returns. As a result, all 38 attributes will be employed in the multivariate tests in building expected return factor models. The univariate scores of the attributes being included in the expected return factor models will be revisited to determine whether it is fair for researchers to eliminate the attributes that failed the univariate tests at the initial stage of the construction of expected return factor models.

6

Style Attributes and the Cross-Section of Equity Returns: Multivariate Linear Approach

6.1 Introduction

The multivariate linear approach in forecasting equity returns entails the construction of a series of multifactor linear models and the subsequent implementations of these models in forecasting equity returns. The main objective of this chapter is to determine the relative strengths of the various combinations of style attributes in explaining and forecasting equity returns when they are employed as independent variables in the pre-specified linear multifactor models. One of the advantages of the multivariate approach over the univariate approach is that it examines the collective explanatory power of the attributes, and hence the complementary properties of the attributes can be utilised in forecasting equity returns.

Four linear multifactor models are constructed in this chapter, each of them having an objective of maximising one of the four pre-specified performance measurement criteria, namely, the average information coefficient (average IC), the Grinold (1989) information ratio (Grinold IR), the Qian and Hua information ratio (QH IR) and the slope *t*-statistic respectively. While the average IC, Grinold IR and the slope *t*-statistic are adapted from Grinold and Khan (2000), QH IR is adapted from Qian and Hua (2003).

There are two decisions to be made before developing the pre-specified linear factor models. Firstly, the method in which the attributes are selected as model inputs must be determined. A model's full potential in forecasting returns cannot be utilised if it omits important attributes as its inputs. On the other hand, an overfitted model that overloads attributes as its inputs is subject to multicollinearity. As a result, it may achieve great in-sample score with extremely poor out-of-sample forecasting performance. Secondly, the frequency and the manner in which the attributes and their factor payoffs are updated must be determined. A model that updates its inputs too frequently may omit past yet still relevant information for forecasting

returns. On the other hand, a model that updates its inputs less frequently than it should fails to take into account the dynamics of the economic and investment activities in forecasting equity returns.

For the purpose of this research, the stepwise regression procedure is used to select model inputs, meaning that the style attributes are added to the factor models one at a time until the performance measurement score cannot be improved any further. As discussed in Chapter 5, restricting the pool of candidate inputs based on their univariate scores may filter out some important attributes. Thus, all 38 attributes are employed as candidate model inputs in the stepwise regression procedure.

On the other hand, the dynamic moving-window procedure is employed for updating the attributes and their payoffs employed by the respective models. Although Haugen and Baker (1996) and Frankish (2004) update the coefficients of the factor payoffs in the expected return factor models to accommodate the time-varying factor payoffs discussed in Chapter 5, the actual model inputs are selected once-off in the in-sample period without being updated periodically in their studies. This research undertakes to examine the possibilities of the time-varying of the actual model inputs and the possible improvement of the forecasting power through updating model inputs.

6.2 Descriptive Statistics

The detailed description of the performance measurement criteria are discussed in Section 6.2.1. On the other hand, the specifications of the linear expected return multifactor models are demonstrated in Section 6.2.2.

6.2.1 The Performance Measurement Criteria

The first linear factor model is developed to maximise the average IC score, which is simply the time-series mean of the monthly correlation coefficients between the realised returns and the model projected returns. As demonstrated in Equation 6.1, the monthly correlation coefficients between the realised returns and the model's projected returns are computed over the period in which the model performance is evaluated. The monthly correlation coefficients are then averaged to obtain the IC score for the model over the evaluation period.

$$Average\ Information\ Coefficient = \frac{\sum_{t=1}^{T} \rho_t(R_{i,t}; E(R_{i,t}))}{T} \tag{6.1}$$

Where:

$R_{i,t}$ and $E(R_{i,t})$ = the realised return for share i in month t and the model forecasted return for share i in month t respectively;

$\rho_t(R_{i,t}; E(R_{i,t}))$ = the cross-sectional correlation coefficient between $R_{i,t}$ and $E(R_{i,t})$ in month t; and

T = the number of months in the evaluation period.

The average IC score is regarded as the skill of an investment strategy (in this case, the expected return multifactor models), and it increases the probabilities of successful forecasts (Grinold and Khan, 2000). Although the average IC score of a model provides indication as to the extent to which the model forecasted returns match the realised returns, it fails to take into account the model's breadth and the volatility of the model's forecasting accuracy.

According to Grinold and Khan (2000), the breadth of an investment strategy (in this case, the expected return multifactor models) is the number of independent projected returns over the evaluation period. The value added by an expected return factor model will increase with an increase in the number of forecasts (that is, the breadth) without compromising its average IC score. Thus, Grinold (1989) derives an approximation of the information ratio by making adjustments for the breadth of the monthly forecasts in the performance measurement as shown in Equation 6.2.

$$Grinold\ Information\ Ratio = \frac{\sum_{t=1}^{T} \rho_t(R_{i,t}; E(R_{i,t})) \times \sqrt{N_t}}{T} \tag{6.2}$$

Where:

$\sqrt{N_t}$ = the adjustment for the breadth of the forecast in month *t*. N_t is the number of sample share returns being projected in month *t*.

On the other hand, a model with a high average IC score (that is, high average correlations between the forecasted and the realised returns) adds little value if it made extremely poor forecasts from time to time. Such a model is considered unreliable for active bets in the equity markets due to the high volatility of its forecasting accuracy. Qian and Hua (2003) provide a remedy to this problem by dividing the average IC score by the standard deviation of the time-series of ICs as depicted in Equation 6.3. Based on this equation, the model's performance is indirectly proportional to the volatility of its forecasting accuracy.

$$Qian\ and\ Hua\ Information\ Ratio = \frac{\overline{IC}}{\sigma_{IC}} \tag{6.3}$$

Where:

$\overline{IC}$ = the average IC score of the factor model; and

σ_{IC} = the standard deviation (that is, the volatility) of the time-series ICs.

In addition to QH IR, the volatility of the forecasting accuracy can be incorporated in the performance evaluation measure by firstly regressing the monthly realised returns on the monthly model forecasted returns to obtain the monthly slope of the regression for the evaluation period. The Student's *t*-statistic for the time-series mean of the slope, known as the slope *t*-statistic, can therefore be computed using Equation 6.4.

$$Slope\ t\text{-}Statistic = \frac{\bar{b}}{\sigma_{\bar{b}} / \sqrt{T}} \tag{6.4}$$

Where:

$\bar{b}$ = the time-series mean of the slope coefficients of the model forecasted return over the evaluation period;

$\sigma_{\bar{b}} / \sqrt{T}$ = the standard error of $\bar{b}$; and

$\sigma_{\bar{b}}$ = the standard deviation of the time-series slope coefficients of the model forecasted returns.

The slope *t*-statistic measures how well the model forecasted returns explain the realised returns, and the mean slope coefficient (the numerator) is seen as an alternative measure of the average IC. The higher the volatility of the model's forecasting accuracy, the lower is the model's performance in terms of its slope *t*-statistic score through the increase in the standard error of the denominator.

6.2.2 The Linear Expected Return Multifactor Models

The monthly cross-sectional payoffs to the various style attributes are computed using the OLS technique in the multiple linear regression analysis shown in Equation 6.5. The difference between Equation 6.5 and the Haugen and Baker (1996) multifactor model lies on the estimation of the constant term in Equation 6.5. Introducing the common threshold of the cross-sectional equity returns into the factor model avoids forcing the regression line passing

through the origin, and should be able to reduce the estimation error of the original Haugen and Baker (1996) model.

$$R_{i,t} = \hat{a}_t + \sum_{j=1}^{J} \hat{b}_{j,t} \times F_{i,j,t-1} + \hat{u}_{i,t} \tag{6.5}$$

Where:

$R_{i,t}$ = share *i*'s realised return in month *t*;

$\hat{a}_t$ = the estimated common constant for the cross-sectional sample share returns in month *t*;

$\hat{b}_{j,t}$ = the factor payoff to style attribute *j* in month *t*;

$F_{i,j,t-1}$ = the style attribute *j* for share *i* at the end of month *t-1*; and

$\hat{u}_{i,t}$ = unexplained component of return for share *i* in month *t*.

Once the monthly factor payoffs and the common constants are estimated, the expected return factor models can be established to forecast the JSE sample share returns over the out-of-sample period from 01 January 2002 to 31 December 2007 as shown in Equation 6.6. The time-series mean of the monthly payoffs over the trailing 12 months is used as the proxy of the expected payoffs to the respective style attributes in Equation 6.6. This is considered as a necessary step to ensure that a history of the payoffs to the various factors is taken into account due to the fact that the payoffs to the style factors may be time-varying. In other words, *"we can use the information embodied in the payoff histories to make out-of-sample projections on the sizes of the future payoffs in later periods"* (Haugen and Baker, 1996: 409). Similarly, the expected constant for a particular month is computed by averaging the trailing 12-month monthly constant estimated for the monthly cross-sectional sample share returns.

$$E(R_{i,t}) = E\left(\hat{a}_t\right) + \sum_{j=1}^{J} E\left(\hat{b}_{j,t}\right) \times F_{i,j,t-1} \tag{6.6}$$

Where:

$E\left(\hat{a}_t\right)$ = the estimated constant term of the factor model representing the expected common threshold of the cross-sectional sample share return in month *t*;

$E(R_{i,t})$ = the expected return on share *i* in month *t*;

$E\left(\hat{b}_{j,t}\right)$ = the expected payoff to style attribute *j* in month *t* (the arithmetic mean of the estimated payoff over the trailing 12 months); and

$F_{i,j,t-1}$ = style attributes of share *i* at the end of month *t-1*.

For example, to estimate the expected return of sample shares for January 2002 (the first month in the out-of-sample period), the prior 12-month (that is, from 01 January 2001 to 31 December 2001) cross-sectional constants and factor payoffs to various style attributes are estimated by Equation 6.5. The cross-sectional constants and factor payoffs are then averaged to represent the projections of the common constant and factor payoffs for January 2002. Given the lagged values of the style attributes and the projected payoffs for these attributes, the expected return for each sample share can be computed for January 2002 using Equation 6.6.

6.3 Methodology

Based on the model specifications described in section 6.2, the 38 style attributes are employed as the candidate inputs in the stepwise regression procedure to construct expected return multifactor models for the 4 pre-specified performance measurement criteria in each the 6 overlapping in-sample periods. The details of the stepwise regression procedure are discussed in section 6.3.1, and the details of the moving window procedure are discussed in section 6.3.2.

6.3.1 The Stepwise Regression Procedure

The stepwise regression procedure is initialised by obtaining the four performance scores for each of the 38 style attributes when they are applied as inputs in the single-factor model. Each style attribute is then ranked under each of the performance measurement criterion. The style attribute that obtains the highest score for a particular performance measurement criterion will be the starting model input to build the expected return multifactor model for that criterion. Next, other attributes are used in combination with the starting attribute one at a time to obtain their respective 2-factor performance scores. The attribute that improves and contributes the most to the particular performance measurement criterion for the existing single-factor is retained as the second factor in the model. This stepwise procedure is continued, and factors added, until the performance score cannot be improved any further for each of the performance measurement criteria. The final products of this procedure are 4 expected return multifactor models that optimise the scores of their respective performance measurement criteria in each of the 6 overlapping in-sample periods.

6.3.2 The Dynamic Moving-Window Procedure

Under the dynamic moving-window procedure, the inputs of the expected return factor models are updated over time as new data/information arrives and older data/information is discarded. The success of the dynamic moving-window procedure depends greatly on the frequency with which the model is updated. This study employs the moving-window procedure whereby the inputs of the multifactor models are updated every 12 months.

For the purpose of this study, the multifactor models are trained and their inputs (that is, attributes to be included in the models) determined via the stepwise regression procedure within the period of 60 months to forecast the immediate 12-month returns following the training period.[1] The first training dataset under this design spans the period from 01 January 1997 to 31 December 2001 to forecast the 12-month returns in 2002. The second training window covers the dataset spanning the period from 01 January 1998 to 31 December 2002. This training window discards the first 12-month data (that is, the data of 1997) and incorporates new 12-month data of 2002 to forecast the 12-month returns in 2003. This overlapping moving window procedure is then repeated for a further four times until all monthly returns in the out-of-sample period (01 January 2002 to 31 December 2007) are forecasted. The details of the dynamic moving-windows regarding the overlapping in-sample periods and their corresponding out-of-sample periods are summarised in Table 6.1.

[1] However, the expected model constants and the expected factor payoffs in Equation 6.6 are updated within each of the 12 months based on the average of their trailing 12-month estimates using Equation 6.5 (see Section 6.2.2). This is done to accommodate the possible time-varying nature of the expected constants and the expected factor payoffs discussed in Haugen and Baker (1996).

Table 6.1 The Dynamic Moving-Window Periods

Under the dynamic moving-window procedure, the inputs of the expected return multifactor models are updated over time as new data/information arrives and older data/information is discarded. For the purpose of this study, the multifactor models are trained with the identities of the model inputs determined via the stepwise regression procedure over a period of 60 months starting from 01 January 1997 to forecast the 12-month returns subsequent to the training period. The training period (and thus the model inputs) is updated every 12 months for 6 times to form 6 overlapping in-sample periods, and their respective 6 non-overlapping out-of-sample periods. The combined 6 out-of-sample periods in which sample share returns are forecasted and subsequently evaluated starts from 01 January 2002 to 31 December 2007 (a total of 72 months). The length of each in-sample period is 60 months, and the out-of-sample periods are 12 months in length. This procedure is known as the 60-for-12 moving- window procedure.

No.	Overlapping In-Sample (Training) Periods	Corresponding Out-of-Sample (Forecasting) Periods
1	01 January 1997 to 31 December 2001 (60 months)	01 January 2002 to 31 December 2002 (12 months)
2	01 January 1998 to 31 December 2002 (60 months)	01 January 2003 to 31 December 2003 (12 months)
3	01 January 1999 to 31 December 2003 (60 months)	01 January 2004 to 31 December 2004 (12 months)
4	01 January 2000 to 31 December 2004 (60 months)	01 January 2005 to 31 December 2005 (12 months)
5	01 January 2001 to 31 December 2005 (60 months)	01 January 2006 to 31 December 2006 (12 months)
6	01 January 2002 to 31 December 2006 (60 months)	01 January 2007 to 31 December 2007 (12 months)

6.4 Results

The style attributes selected via the stepwise regression procedure as the inputs for the pre-specified linear expected return multifactor models in each moving-window training period are demonstrated in Table 6.2. The order in which each of the style attributes being included in the average IC model, the Grinold IR model, the QH IR model and the slope *t*-statistic model are displayed in PANEL (A) through PANEL (D) of Table 6.2 respectively. The style attributes with significant time-series means of their factor payoffs during the training periods are in italics and highlighted in bold. The style attributes marked with a single asterisk in Table 6.2 are the attributes in the last iterations of the stepwise regression procedure. These attributes are eliminated because no further improvements in the performance scores of the designated criteria can be achieved through their inclusions. On the other hand, the style attributes marked with double asterisks are further eliminated from the designated models due to the possibilities of multicollinearity with their inclusions in the model.[2]

As the room for further improvements in the performance scores become minimal towards the last few iterations of the stepwise regression procedure, the problem of overfitting through multicollinearity is more likely to occur. This phenomenon is observed in Table 6.2 as all the style attributes marked with double asterisks are directly above the attributes in the last iterations of the respective moving-window training periods.

[2] The pairs of the style attributes that are potentially subject to the consequences of multicollinearity for their simultaneous inclusions if the absolute values of their mean monthly correlations are not less than 0.70 (see Section 5.2). The highly correlated pairs of the style attributes are extracted in Section 5.3 by examining the mean monthly correlation table displayed in Table 5.2. The pairs of style attributes identified are DEBTTMVE and DEBTTBVE, FOREY1 and FOREY2, LAGLPRICE, LPRICE and LSIZE, MOM6 and MOM12, MOM6 and MOM3, and ROE and GROWTH.

Table 6.2 The Stepwise Selection Procedure of the Style Attributes

The order in which each of the style attributes being included in the average IC model, the Grinold IR model, the QH IR model and the slope *t*-statistic model are displayed in PANEL (A) through PANEL (D) respectively. The style attributes with significant mean factor payoffs during the respective training window periods are in italics and highlighted in bold. The style attributes assigned with a single asterisk are those being eliminated in the model due to the fact that they do not contribute to the improvements of the respective performance scores. On the other hand, the style attributes assigned with double asterisks are those attributes that are highly correlated (correlation coefficient of greater than 0.70 or less than -0.70) to one or more of the attributes in the prior iterations. These attributes are presumed to introduce overfitting problems through the possibility of multicollinearity with their inclusions in the respective models, and are subsequently eliminated from the models.

PANEL (A) Average IC Model:

Factor No.	1 Jan 1997 to 31 Dec 2001		1 Jan 1998 to 31 Dec 2002		1 Jan 1999 to 31 Dec 2003		1 Jan 2000 to 31 Dec 2004		1 Jan 2001 to 31 Dec 2005		1 Jan 2002 to 31 Dec 2006	
1	***BVTP***	0.0891	***BVTP***	0.1202	***BVTP***	0.1502	***BVTP***	0.1549	***BVTP***	0.1390	***LSIZE***	0.1145
2	***FOREY2***	0.1512	MOM12	0.1417	***ROE***	0.1621	ROE	0.1830	ROE	0.1712	***G12MNPMARGIN***	0.1288
3	***CFTP***	0.1657	DEBTTBVE	0.1464	***CFTP***	0.1748	***CFTP***	0.1943	***CFTP***	0.1859	***CFTP***	0.1523
4	CURRENTRATIO	0.1936	*MOM1	0.1457	DEBTTMVE	0.1810	G12MGPMARGIN	0.2093	G12MGPMARGIN	0.2159	GROWTJ	0.1580
5	DEBTTMVE	0.2088			***DEBTTBVE***	0.1820	DEBTTMVE	0.2101	*ICBT	0.2151	MOM12-1	0.1668
6	EY	0.2249			***LPRICE***	0.1774	MOM1	0.2123			***MOM3***	0.1777
7	***LSIZE***	0.2291					*CFTCURRLIABS	0.2116			***BVTP***	0.1839
8	***MOM1***	0.2299									TATURNOVR	0.1852
9	***SALESTP***	0.2282									***LPRICE***	0.1861
10											***MOM12***	0.1858

PANEL (B) Grinold IR Model:

Factor No.	1 Jan 1997 to 31 Dec 2001		1 Jan 1998 to 31 Dec 2002		1 Jan 1999 to 31 Dec 2003		1 Jan 2000 to 31 Dec 2004		1 Jan 2001 to 31 Dec 2005		1 Jan 2002 to 31 Dec 2006	
1	***BVTP***	0.9459	***BVTP***	1.2934	***BVTP***	1.6194	***BVTP***	1.6848	***BVTP***	1.5344	***LSIZE***	1.3190
2	***FOREY2***	1.3190	MOM12	1.4549	***DEBTTBVE***	1.7190	ROE	1.9478	ROE	1.8477	MOM12-1	1.3737
3	***CFTP***	1.4356	DEBTTBVE	1.4761	***CFTP***	1.8486	***CFTP***	2.0697	***CFTP***	2.0066	***EY***	1.5570
4	CURRENTRATIO	1.4712	***CFTP***	1.4438	ROA	1.8655	DEBTTMVE	2.1031	***LPRICE***	2.0753	***MOM12***	1.7153
5	DEBTTMVE	1.5789			*DEBTTMVE	1.8467	***LPRICE***	2.1144	***MOM12***	2.0929	***BVTP***	1.7687
6	EY	1.6714					*MOM12	2.1054	ROA	2.1710	ROE	1.8460
7	***LSIZE***	1.7039							MOM3	2.1993	***CFTP***	2.0081
8	***MOM1***	1.6933							***LSIZE***	2.1829	***MOM3***	2.0352
9											***LPRICE***	2.0443
10											***LAGLPRICE***	2.0780
11											*TATURNOVR	2.0595

PANEL (C) Qian and Hua IR Model:

Factor No.	1 Jan 1997 to 31 Dec 2001		1 Jan 1998 to 31 Dec 2002		1 Jan 1999 to 31 Dec 2003		1 Jan 2000 to 31 Dec 2004		1 Jan 2001 to 31 Dec 2005		1 Jan 2002 to 31 Dec 2006	
1	***FOREY2***	0.6072	***BVTP***	0.5755	***BVTP***	0.8016	***BVTP***	0.8757	***BVTP***	0.8623	***LPRICE***	0.7131
2	***CFTP***	0.9036	***MOM12-1***	0.8555	MOM12	1.0095	***MOM12-1***	1.1890	***MOM12-1***	1.1445	***G12MNPMARGIN***	0.7562
3	CURRENTRATIO	1.0412	DEBTTMV	0.8853	***DEBTTBVE***	1.0445	***DEBTTBVE***	1.2857	GROWTH	1.3161	***CFTP***	0.9288
4	***BVTP***	1.1982	EY	0.8877	***CFTP***	1.0781	**DEBTTMVE	1.3005	***CFTP***	1.4066	***BVTP***	1.0170
5	EY	1.3863	***CFTP***	0.9295	***EY***	1.1239	***LAGLPRICE***	1.2841	MOM1	1.4366	***DEBTTBVE***	1.0933
6	TATURNOVR	1.4730	***FOREY2***	0.9267	CFTDEBT	1.2480			ICBT	1.4623	MOM1	1.1630
7	***SALESTP***	1.5324			*TATURNOVR	1.2075			G12MCPS	1.4826	***MOM24***	1.1923
8	***LAGLPRICE***	1.5325							*MOM3	1.4806	***MOM12***	1.2251
9	***MOM1***	1.5299									***MOM3***	1.2665
10											*TATURNOVR	1.2449

PANEL (D) Slope *t*-Statistic Model:

Factor No.	1 Jan 1997 to 31 Dec 2001		1 Jan 1998 to 31 Dec 2002		1 Jan 1999 to 31 Dec 2003		1 Jan 2000 to 31 Dec 2004		1 Jan 2001 to 31 Dec 2005		1 Jan 2002 to 31 Dec 2006	
1	EY	3.5741	**FOREY2**	3.6970	***BVTP***	4.9454	***BVTP***	5.5341	***BVTP***	6.1666	***BVTP***	4.5921
2	***MOM12-1***	5.2957	CURRENTRATIO	3.7277	***MOM12-1***	6.1820	***G12MDPS***	8.0023	***MOM12***	7.6327	***G12MEPS***	6.9009
3	***CFTP***	6.5400	***CFTP***	5.7766	***SALESTP***	6.6801	***SALESTP***	8.0513	GROWTH	8.2291	G12MCPS	7.9254
4	***MOM1***	6.6670	***BVTP***	6.0671	***CFTP***	7.1391	*DEBTTMVE	7.9714	***CFTP***	8.8478	***CFTP***	8.9107
5	***LAGLPRICE***	6.5610	*TATURNOVR	6.0121	***EY***	7.9872			MOM1	9.4003	***DEBTTBVE***	8.9473
6					TATURNOVR	8.2272			***LPRICE***	9.5930	*MOM1	8.8900
7					***DEBTTBVE***	7.9761			MOM3	9.8246		
8									*ICBT	9.8191		

6.4.1 Stepwise Selection Procedure of the Style Attributes

Examining the style attributes being included in the pre-specified linear expected return multifactor models in Table 6.2 reveals that most of the model inputs consist of attributes with significant mean payoffs (in italics and highlighted in bold) during the training window periods. This study result supports the use of the univariate score based on the mean factor payoffs to filter out attributes that are less likely to be important in the constructions of the expected return multifactor models of Van Rensburg and Robertson (2003a), Frankish (2004) and Moore (2008). Nevertheless, it is apparent that the in-sample performance of the models can be improved by including attributes that are not significant in terms of their mean factor payoffs in the model constructions. On the other hand, certain attributes with significant payoffs might not necessarily be included in the model due to the fact that their contributions to the explanation of the cross-sectional equity returns overlap with other attributes selected in the prior iterations in the stepwise regression procedure. For example, attributes related to earnings, such as EY, EARNREV, G12MEPS, FOREY1 and FOREY2, are not selected for the inclusions in all four multifactor models during the fourth and the fifth moving-window training periods. However, these attributes do have significant mean factor payoffs during these training periods.

6.4.2 The Periodic Model Performance

The periodic performances of the average IC model, the Grinold IR model, the QH IR model and the slope *t*-statistic model are evaluated by their respective average IC score, Grinold IR score, QH IR score, and the slope *t*-statistic during the evaluation periods. The performance scores of the pre-specified models for the 6 moving-window in-sample (training) periods and their corresponding out-of-sample (forecasting) periods are displayed in PANEL (A) through PANEL (F) of Table 6.3 respectively. In addition, the performance scores of the models that do not take out any of the highly correlated attributes are shown in brackets. As mentioned earlier, the aim of removing one or more highly correlated attribute in the construction of the linear expected return multifactor model is to reduce the problem of overfitting and hence improve the robustness and the forecasting power of the model. The study results

demonstrated in Table 6.3 support this view in that the out-of-sample scores of the models that attempt to address the multicollinearity issue are higher than the scores of the models that fail to eliminate highly correlated attributes in 14 out of 16 cases. The two exceptions are the QH IR out-of-sample score and the slope *t*-statistic out-of-sample score for the average IC model in PANEL (C) of Table 6.3, whereas the average IC model that attempts to address the overfitting problem ends up with lower out-of-sample performance scores in 2004. Overall, Table 6.3 shows that the out-of-sample performance scores of the pre-specified models constructed under the dynamic moving-window procedure are consistent with their respective in-sample performance scores.[3]

Figure 6.1 extracts the periodic in-sample performance information from Table 6.3. CHART (A) through CHART (D) of Figure 6.1 illustrate the comparisons of the in-sample performances of the pre-specified models under each of the performance measurement criteria. Thus, the direct comparisons of the relative performance of the pre-specified linear expected return multifactor models under each of the performance measurement criteria can be made. Frankish (2004) suggests that the stepwise regression procedure has an unfair bias for the inclusion of the first variable. The possible consequence might be that models start with alternative attributes in the model constructions may achieve better performance scores. Observing the histograms of the periodic in-sample model performance of the pre-specified linear expected return multifactor models displayed in Figure 6.1 shows that the models constructed under their designated performance measurement criteria have great potential in achieving the best in-sample scores for the designated criteria. More specifically, in competing with the alternative models, the average IC model is more likely to achieve better in-sample average IC score; the Grinold IR model is more likely to achieve better in-sample average Grinold IR score; the QH IR model is more likely to achieve better in-sample QH IR score; and the slope *t*-statistic model is more likely to achieve better in-sample slope *t*-statistic score. This finding indicates that the implications of the weakness of the stepwise regression procedure suggested by Frankish (2004) are not as severe as it seems.

[3] That is, there are no large discrepancies between the in-sample performance scores and the out-of-sample performance scores.

Table 6.3 The Periodic Model Performance Results

The periodic performances of the average IC model, the Grinold IR model, the QH IR model and the slope *t*-statistic model are evaluated by their respective average IC score, Grinold IR score, QH IR score, and the slope *t*-statistic during the evaluation periods. The performance scores of the pre-specified models for the 6 moving-window in-sample (training) periods and their corresponding out-of-sample (forecasting) periods are displayed in PANEL (A) through PANEL (F) respectively. The figures in brackets represent the performance scores obtained by the models that do not attempt to address the problem of multicollinearity by eliminating highly correlated attributes in the models.

PANEL(A) In-Sample: 1 Jan 1997 to 31 Dec 2001; Out-of-Sample: 1 Jan 2002 to 31 Dec 2002				
	Average IC Model	**Grinold IR Model**	**Qian and Hua IR Model**	**Slope *t*-Statistic Model**
Avg IC (In-Sample)	0.2299	0.2291	0.2053	0.1541
Avg IC Score (Out-of-Sample)	0.1109	0.1217	0.1167	0.1173
Grinold IR Score (In-Sample)	1.6933	1.7039	1.5352	1.3922
Grinold IR Score (Out-of-Sample)	0.8153	0.9170	0.8885	1.1887
Qian and Hua IR Score (In-Sample)	1.2350	1.2632	1.5325	0.8553
Qian and Hua IR Score (Out-of--Sample)	0.4355	0.5584	0.6333	0.8763
Slope t-Statistic Score (In-Sample)	4.9802	5.2745	5.9621	6.6670
Slope t-Statistic Score (Out-of-Sample)	1.2845	1.7609	2.2566	2.8962
PANEL (B) In-Sample: 1 Jan 1998 to 31 Dec 2002; Out-of-Sample: 1 Jan 2003 to 31 Dec 2003				
	Average IC Model	**Grinold IR Model**	**Qian and Hua IR Model**	**Slope *t*-Statistic Model**
Avg IC (In-Sample)	0.1464	0.1464	0.1548	0.1801
Avg IC Score (Out-of-Sample)	0.1720	0.1720	0.2183	0.2031
Grinold IR Score (In-Sample)	1.4761	1.4761	1.4966	1.3840
Grinold IR Score (Out-of-Sample)	1.8629	1.8629	2.2707	1.5995
Qian and Hua IR Score (In-Sample)	0.8282	0.8282	0.9295	1.1000
Qian and Hua IR Score (Out-of--Sample)	1.5470	1.5470	1.7464	1.0467
Slope t-Statistic Score (In-Sample)	5.0675	5.0675	6.4753	6.0671
Slope t-Statistic Score (Out-of-Sample)	4.3528	4.3528	5.5305	3.4076
PANEL (C) In-Sample: 1 Jan 1999 to 31 Dec 2003; Out-of-Sample: 1 Jan 2004 to 31 Dec 2004				
	Average IC Model	**Grinold IR Model**	**Qian and Hua IR Model**	**Slope *t*-Statistic Model**
Avg IC (In-Sample)	0.1810 (0.1820)	0.1811	0.1458	0.1757
Avg IC Score (Out-of-Sample)	0.2159 (0.2132)	0.1901	0.2375	0.2679
Grinold IR Score (In-Sample)	1.8586 (1.8635)	1.8655	1.4158	1.7624
Grinold IR Score (Out-of-Sample)	2.3658 (2.3367)	2.0340	2.4938	2.8371
Qian and Hua IR Score (In-Sample)	1.0684 (1.1032)	1.0961	1.2480	1.1196
Qian and Hua IR Score (Out-of--Sample)	1.5109 (1.5307)	1.1714	1.1303	1.2732
Slope t-Statistic Score (In-Sample)	7.1996 (7.4322)	7.0011	8.7875	8.2272
Slope t-Statistic Score (Out-of-Sample)	5.0734 (5.2198)	3.8721	3.6700	4.0775

Table 6.3 The Periodic Model Performance Results - Continued

PANEL (D) In-Sample: 1 Jan 2000 to 31 Dec 2004; Out-of-Sample: 1 Jan 2005 to 31 Dec 2005				
	Average IC Model	**Grinold IR Model**	**Qian and Hua IR Model**	**Slope *t*-Statistic Model**
Avg IC (In-Sample)	0.2123	0.1995	0.1765 (0.1722)	0.1476
Avg IC Score (Out-of-Sample)	0.1650	0.1052	0.1228 (0.1137)	0.0720
Grinold IR Score (In-Sample)	1.9428	2.1144	1.8762 (1.8301)	1.3429
Grinold IR Score (Out-of-Sample)	1.6737	1.1832	1.3822 (1.2805)	0.7258
Qian and Hua IR Score (In-Sample)	1.1095	1.1583	1.2857 (1.3005)	1.1044
Qian and Hua IR Score (Out-of--Sample)	0.8034	0.5638	1.0005 (0.8752)	0.5731
Slope t-Statistic Score (In-Sample)	7.6144	7.4652	7.5744 (7.7808)	8.0513
Slope t-Statistic Score (Out-of-Sample)	2.7658	1.6563	3.4797 (3.0472)	2.1512
PANEL (E) In-Sample: 1 Jan 2001 to 31 Dec 2005; Out-of-Sample: 1 Jan 2006 to 31 Dec 2006				
	Average IC Model	**Grinold IR Model**	**Qian and Hua IR Model**	**Slope *t*-Statistic Model**
Avg IC (In-Sample)	0.2159	0.2134	0.1955	0.2046
Avg IC Score (Out-of-Sample)	0.0514	0.1357	0.1500	0.1701
Grinold IR Score (In-Sample)	2.0658	2.1993	1.9054	2.0389
Grinold IR Score (Out-of-Sample)	0.5310	1.4999	1.6053	1.8560
Qian and Hua IR Score (In-Sample)	1.1020	1.3455	1.4826	1.4386
Qian and Hua IR Score (Out-of--Sample)	0.4584	1.1368	1.6951	1.6005
Slope t-Statistic Score (In-Sample)	7.4818	8.8724	10.1962	9.8246
Slope t-Statistic Score (Out-of-Sample)	1.4342	3.4005	5.1103	4.5163
PANEL (F) In-Sample: 1 Jan 2002 to 31 Dec 2006; Out-of-Sample: 1 Jan 2007 to 31 Dec 2007				
	Average IC Model	**Grinold IR Model**	**Qian and Hua IR Model**	**Slope *t*-Statistic Model**
Avg IC (In-Sample)	0.1852 (0.1861)	0.1938 (0.1982)	0.1718	0.1484
Avg IC Score (Out-of-Sample)	0.1729 (0.1710)	0.2056 (0.1642)	0.1761	0.0829
Grinold IR Score (In-Sample)	1.8617 (1.8693)	2.0353 (2.0780)	1.7564	1.5440
Grinold IR Score (Out-of-Sample)	1.8578 (1.8376)	2.2615 (1.7977)	1.9158	0.8573
Qian and Hua IR Score (In-Sample)	1.1302 (1.1300)	1.3376 (1.3745)	1.2665	1.1265
Qian and Hua IR Score (Out-of--Sample)	1.2711 (1.2517)	1.9808 (1.4664)	1.6964	0.8513
Slope t-Statistic Score (In-Sample)	8.6125 (8.7064)	9.6315 (9.6986)	9.0001	8.9473
Slope t-Statistic Score (Out-of-Sample)	3.1816 (3.1390)	4.6208 (4.1310)	4.2071	2.7897

Figure 6.1 The Periodic In-Sample Model Performance Results

CHART (A) to CHART (D) illustrate the histograms of the in-sample performance scores for the pre-specified models under each of the four performance measurement criteria. The performance scores of the average IC model, the Grinold IR model, the QH IR model and the slope *t*-statistic model in each chart are represented by the red, yellow, blue and green bars respectively to enhance the direct comparisons of the periodic in-sample performance of the pre-specified linear expected return multifactor models.

CHART (A) In-Sample Average IC Scores

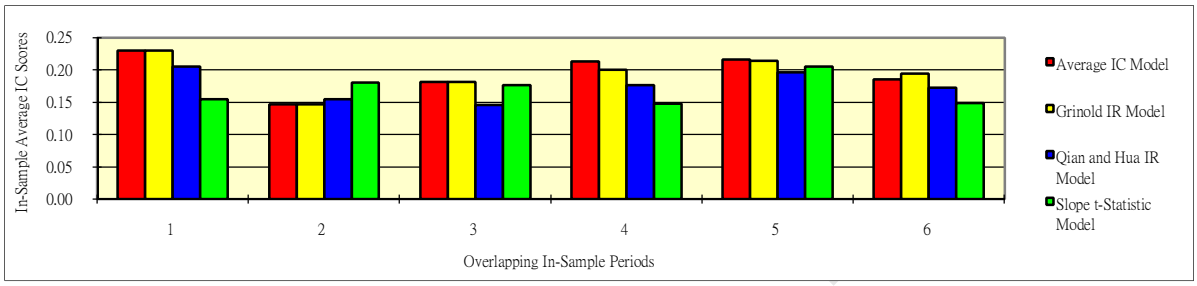


CHART (B) In-Sample Grinold IR Scores

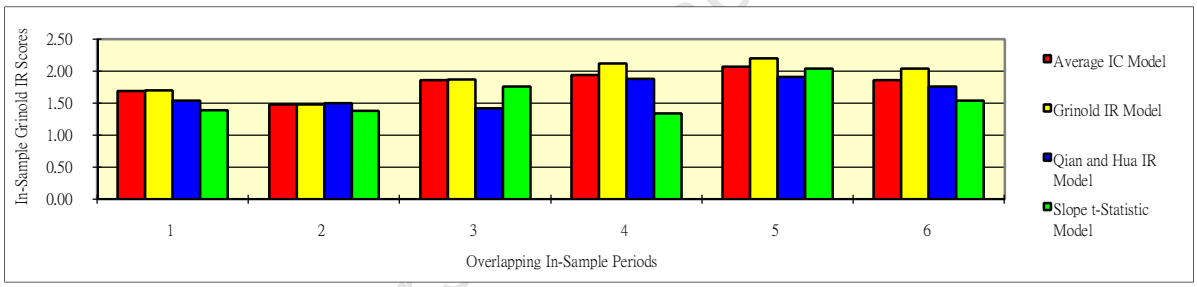


CHART (C) In-Sample Qian and Hua IR Scores

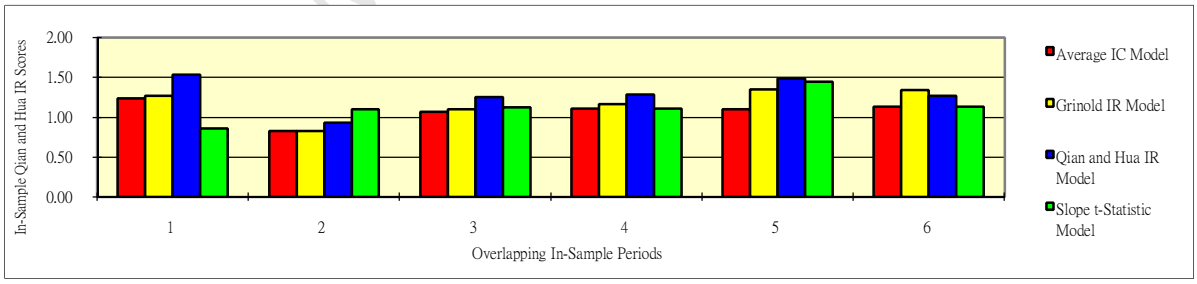


CHART (D) In-Sample Slope *t*-Statistic Scores

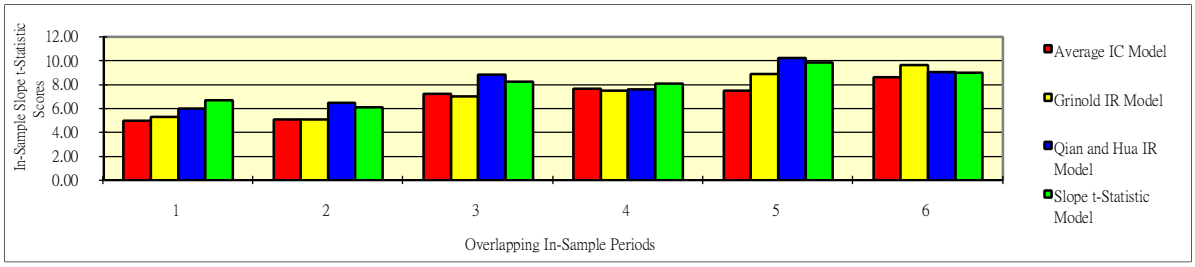


Figure 6.2 illustrates the periodic out-of-sample test results extracted from Table 6.3. Similar to Figure 6.1, the out-of-sample scores of the pre-specified models under each of the four performance measurement criteria are demonstrated in CHART (A) to CHART (D) of Figure 6.2 respectively. In order to detect the time-varying property of the style attributes employed by the linear expected return multifactor models, Figure 6.2 depicts both the performance of the models developed under the dynamic moving-window procedure and that of the models trained within the static period. In contrast to the models trained within the static period from 01 January 1997 to 31 December 2001, the dynamic moving-window procedure updates the identities of the model inputs every 12 months to include the most relevant style attributes in the pre-specified models. The performance scores of the models constructed under the dynamic moving-window procedure are represented by the histograms. On the other hand, the trend lines represent the performance scores of the static models that do not attempt to update the identities of the model inputs.

Comparing the relative performances of the pre-specified linear expected return multifactor models under each of the performance measurement criteria reveals two crucial points for the successful construction of the linear expected return multifactor models. Firstly, contrary to the periodic in-sample results, the pre-specified models do not necessarily achieve relatively better out-of-sample scores for their designated performance criteria over time. For example, the average IC model only achieves the top average IC score in 2005 (see CHART (A) of Figure 6.2). On the other hand, the QH IR model has been the top 2 QH IR performer in 2002, 2003, 2005, 2006 and 2007 (see CHART (C) of Figure 6.2). The Grinold IR model and the slope *t*-statistic model have been the top 2 performers for their respective designated criteria in 3 out of the 6 years of the out-of-sample period. The fact that the pre-specified models are able to outperform competing models in their respective designated criteria over the training periods but fail to do so over the out-of-sample period means that the in-sample performance of a particular model is not a good indication of its out-of-sample performance.

Secondly, the static models that only attempt to update the expected factor payoffs without updating the identities of the inputs over time, in general, underperform the models developed under the dynamic moving-window procedure, with the exception of the slope *t*-statistic model. The out-of-sample performance of the static slope *t*-statistic model is more

or less in line with the performance of the dynamic *t*-statistic model. The style attributes, EY, MOM12-1, CFTP and MOM1, selected by the static slope *t*-statistic model seem to have consistent ability in forecasting the cross-sectional equity returns on the JSE over the period from 01 January 1997 to 31 December 2007. On the other hand, the combinations of the style attributes selected by the other three static models lose their forecasting ability over time. It is also apparent that the out-of-sample underperformance of the static models seems to increase over time. These findings serve as evidence that the relative importance of the identities (in addition to the factor payoffs) of the style attributes are time-varying in explaining the cross-sectional equity returns on the JSE. Therefore, the dynamic moving-window approach should be adapted to update the identities of the style attributes in the linear expected return multifactor models.

The overall periodic in-sample and out-of-sample results suggest that the value attributes (especially BVTP, CFTP and SALESTP) and size attributes (especially LPRICE and LSIZE) are important attributes in explaining and forecasting the cross-sectional equity returns on the JSE. The consolidated out-of-sample performance scores of the dynamic models and that of the static models over the period from 01 January 2002 to 31 December 2007 (72 months) are demonstrated in CHART (A) and CHART (B) of Table 6.4 respectively. The dynamic QH IR model has achieved the best out-of-sample scores for all four performance measurement criteria over the consolidated out-of-sample period. The evaluation of the periodic out-of-sample performance in Figure 6.2 also indicates that the QH IR model achieves consistently good periodic out-of-sample scores for all four performance measurement criteria. Therefore, the QH IR model is the best linear expected return multifactor model for forecasting JSE equity returns over the period from 01 January 2002 to 31 December 2007.

Figure 6.2 The Periodic Out-of-Sample Model Performance Results

CHART (A) to CHART (D) illustrate the periodic out-of-sample performances of the pre-specified models under each of the four performance measurement criteria. The performance scores of the models constructed under the moving-window procedure that updates the model inputs every 12 months are represented by the histograms. On the other hand, the performance scores of the models trained within the static period from 01 January 1997 to 31 December 2001 are represented by the trend lines. The performance of the average IC model, the Grinold IR model, the QH IR model and the slope *t*-statistic model in each chart are highlighted in red, yellow, blue and green respectively to enhance the direct comparisons of the performance of the pre-specified models.

CHART (A) Out-of-Sample Average IC Scores

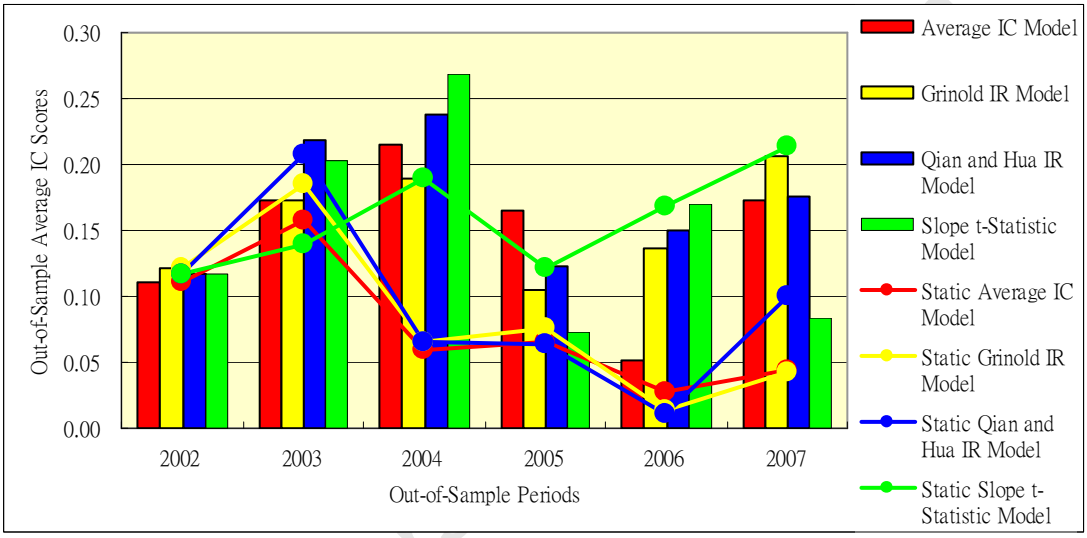


CHART (B) Out-of-Sample Grinold IR Scores

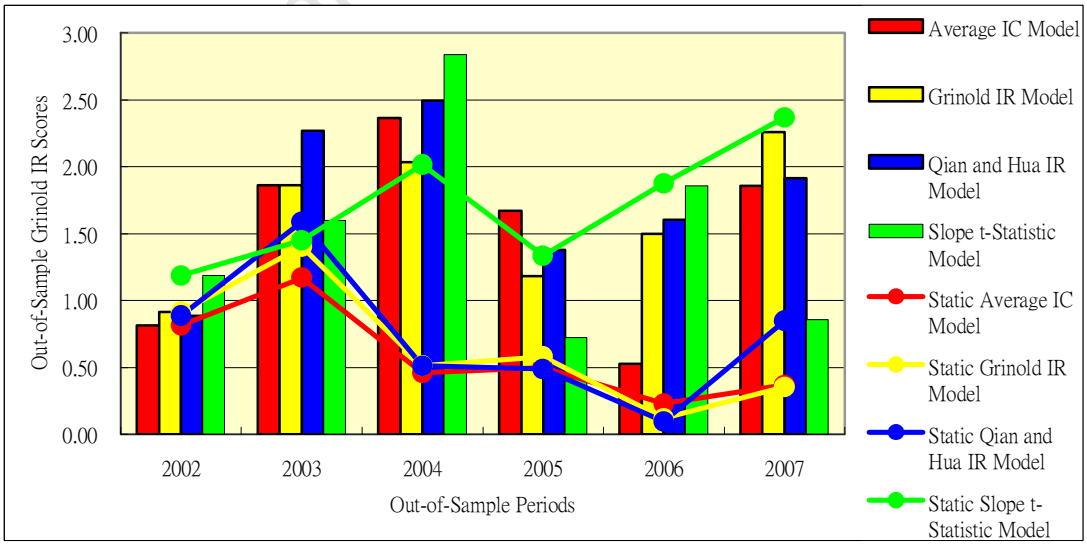


Figure 6.2 The Periodic Out-of-Sample Model Performance Results - Continued

CHART (C) Out-of-Sample Qian and Hua IR Scores

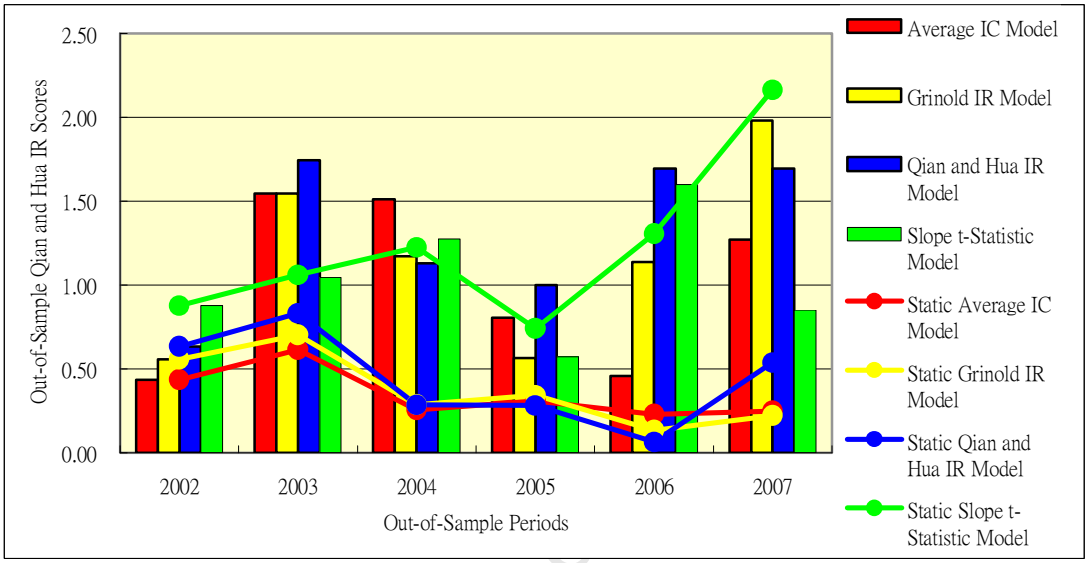


CHART (D) Out-of-Sample Slope *t*-Statistic Scores

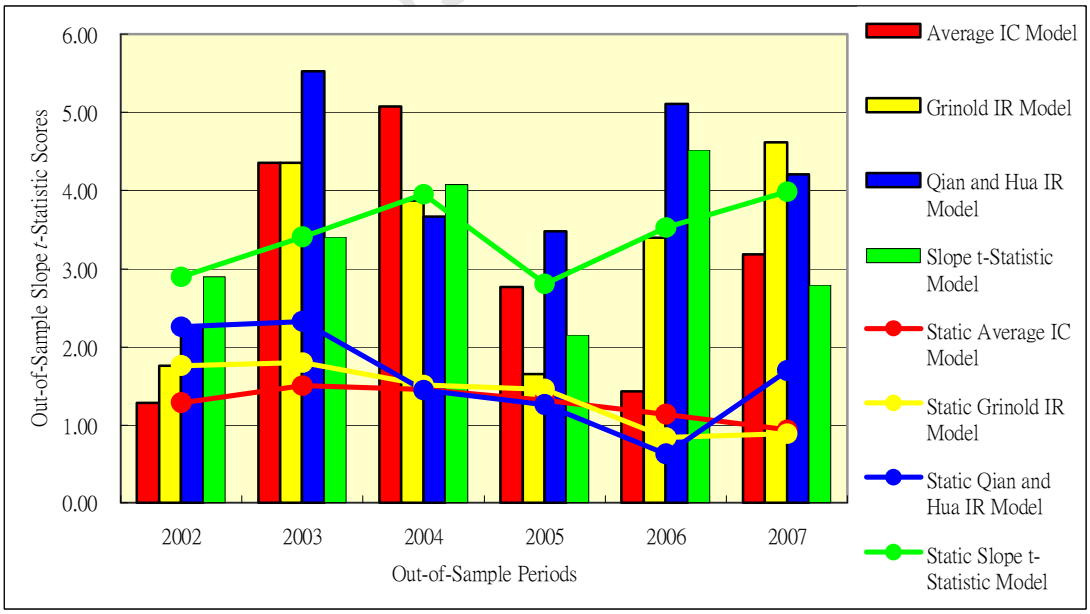


Table 6.4 The Consolidated Out-of-Sample Performance Results

CHART (A) and CHART (B) represent the performance scores of the pre-specified dynamic expected return linear multifactor models and that of the static models respectively over the consolidated out-of-sample period from 1 January 2002 to 31 December 2007. The consolidated out-of-sample period consists of 6 individual out-of-sample sub-periods. The sub-periods are 12 months in length.

CHART (A) Dynamic Moving Window Approach

	The Average IC Model	The Grinold IR Model	The Qian and Hua Model	The Slope t-Statistic Model
Performance Criteria:				
Average IC	0.143	0.150	**0.169**	0.157
Grinold IR	1.461	1.557	**1.737**	1.560
Qian and Hua IR	0.818	0.947	**1.114**	0.950
Slope t-Statistic	5.946	7.044	**8.604**	7.716

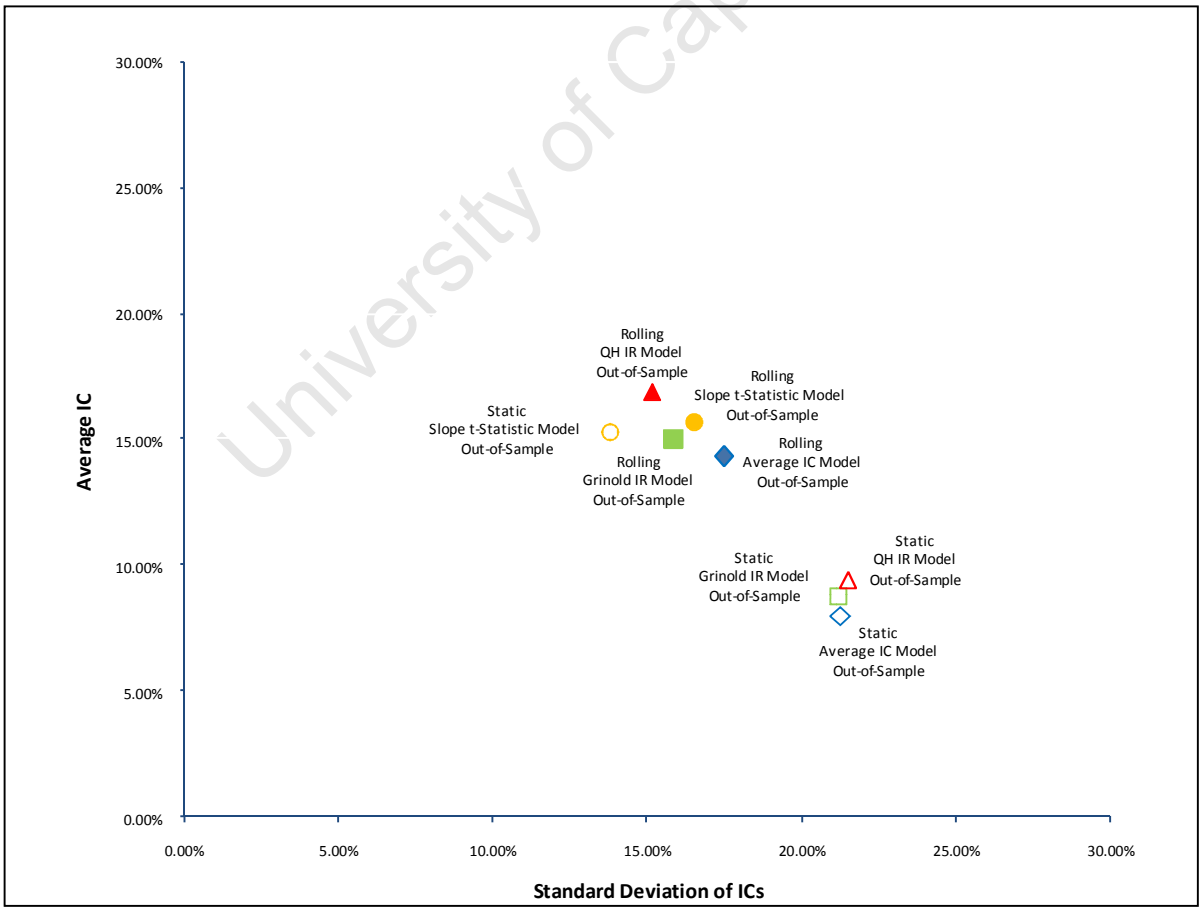
CHART (B) Static Approach

	The Average IC Model	The Grinold IR Model	The Qian and Hua Model	The Slope t-Statistic Model
Performance Criteria:				
Average IC	0.080	0.088	0.094	**0.153**
Grinold IR	0.612	0.676	0.728	**1.646**
Qian and Hua IR	0.376	0.415	0.437	**1.105**
Slope t-Statistic	3.040	3.325	3.876	**8.149**

Figure 6.3 plots the consolidated out-of-sample performances of the static and the rolling linear models based on their respective average IC scores (depicted by the y-axis) and standard deviation of ICs (depicted by the x-axis) over the out-of-sample period from 1 January 2002 to 31 December 2007. With the exception of the static slope *t*-statistic model, the static linear models achieve lower average IC with higher volatility of IC over the out-of-sample period. This serves as evidence that the dynamic updating of the identities of model inputs is a necessary step to build a robust expected return linear model.

Figure 6.3 The Consolidated Out-of-Sample Performance Results

Figure 6.3 plots the static and rolling linear models in the diagram based on their respective consolidated average IC scores and standard deviation of IC over the out-of-sample period from 1 January 2002 to 31 December 2007. The filled labels represent the rolling models while the corresponding static models are represented by empty labels.



The consolidated in- and out-of-sample performance scores for the four linear expected return factor models are tabulated in Table 6.5. The QH IR model achieves the best consolidated in-sample and out-of-sample scores for the QH IR and the slope *t*-statistic performance measurement criteria. Although the QH IR model does not have outstanding in-sample scores for the average IC and the Grinold IR performance measurement criteria, it achieves the highest out-of-sample scores for these two criteria. The in- and out-of-sample performance scores are illustrated in the form of average IC over the standard deviation of ICs in Figure 6.4. Without controlling for the volatility of the forecasting accuracy, the average IC model has the highest out-of-sample forecasting error, and the lowest out-of-sample forecasting accuracy amongst all linear models. On the other hand, the QH IR model has the highest in- and out-of-sample QH IR scores, as depicted by the slope of the line connecting the origin to the position of the model in the diagram, amongst all four linear models.

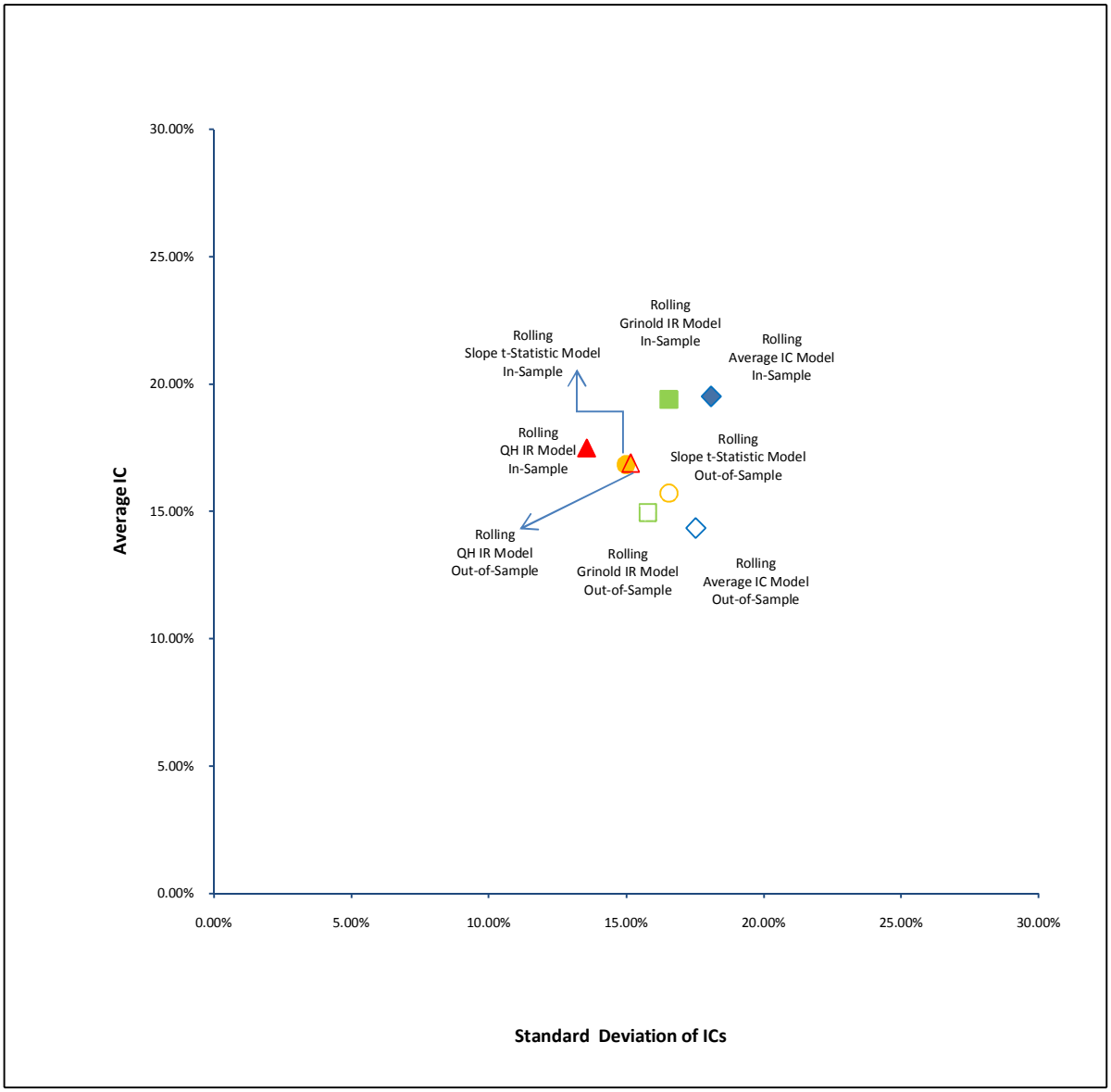
Table 6.5 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Linear Models

Table 6.5 contains the average periodic in-sample and the corresponding out-of-sample performance scores of the rolling expected return factor models developed in this chapter. The best out-of-sample score for each of the four performance measurement criteria is highlighted in bold.

	The Average IC Model	The Grinold IR Model	The Qian and Hua Model	The Slope *t*-Statistic Model
Avg IC (In-Sample)	0.1951	0.1939	0.1750	0.1684
Avg IC (Out-Sample)	0.1435	0.1495	**0.1689**	0.1572
Grinold's IR (In-Sample)	1.8164	1.8991	1.6643	1.5774
Grinold's IR (Out-Sample)	1.4612	1.5570	**1.7367**	1.5603
Qian and Hua's IR (In-Sample)	1.0789	1.1715	1.2908	1.1241
Qian and Hua's IR (Out-Sample)	0.8185	0.9473	**1.1144**	0.9499
Slope t-Statistic (In-Sample)	6.8260	7.2187	7.9993	7.9641
Slope t-Statistic (Out-Sample)	5.9461	7.0441	**8.6043**	7.7160

Figure 6.4 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Linear Models

Figure 6.4 plots the rolling linear expected return factor models in the diagram based on their respective consolidated average IC scores and standard deviation of ICs for the in-sample and out-of-sample periods. The slope of the line connecting the origin to the position of the model in the diagram is essentially the model's QH IR score. The filled labels represent in-sample scores while the corresponding out-of-sample scores are represented by empty labels.



With regard to the comparison between the performances of the models built with and without a regression constant, Appendix B.1 through Appendix B.4 illustrates the periodic performance scores for the average IC model, the Grinold IR model, the QH IR model and the slope *t*-statistic model respectively, with the performance scores for the models built in the absence of the regression constant (the exact Haugen and Baker (1996) model) displayed in the bottom panel in each appendix. Although there is no obvious outperformance for the models built with the regression constant over the models built without the regression constant, in terms of the in-sample scores, the majority of the models built with the regression constant outperform their counterpart models built without the regression constant in the out-of-sample period. This result supports the argument of this research that the performance of the Haugen and Baker (1996) model can be improved by removing the zero regression constant constraint. Incorporating the time-varying regression constant in the moving-window training periods in the expected return forecasting process effectively improves the robustness of the model in forecasting out-of-sample share returns.

6.5 Conclusion

This chapter involves the constructions of a series of expected return multifactor linear models from a pool of 38 style attributes examined in Chapter 5 and the subsequent evaluation of these pre-specified models in explaining and forecasting the cross-sectional equity returns. In doing so, the collective explanatory and forecasting power of the different combinations of the style attributes can be exploited by utlising their complementary properties. Four linear expected return multifactor models are developed. Each of the four models is constructed with an attempt to maximise the in-sample scores of their designated performance measurement criteria, namely, the average IC, Grinold IR, Qian and Hua IR and the slope *t*-statistic.

The average IC score of a model is simply the average monthly correlations between the actual realised returns and the forecasted returns. It is an indication of the extent to which the realised returns and the forecasted returns move in tandem. Being the most basic performance evaluation criterion, the average IC score fails to incorporate the model's breadth, measured by the number of independent projected return, and the volatility of the model's forecasting accuracy, measured by the standard deviation of monthly ICs. As a result, three additional performance measurement criteria examined by Frankish (2004), namely, Grinold IR, QH IR and the slope *t*-statistic are employed to develop the expected return linear multifactor models, and subsequently used to evaluate these models. The performance of a model's forecasting accuracy is improved if the model is able to maintain the forecasting accuracy while increasing the number of forecasts (that is, the breadth of the forecasts). To the other hand, a model with high average IC score is deemed unreliable if the volatility of its average IC score is unreasonably high. Thus, these additional measures serve to adjust the performance scores upwards by the number of forecasts, and/or adjust the performance scores downwards by the volatility of the forecast accuracy.

While Grinold IR incorporates the breadth of the monthly forecasts in the calculation of the information ratio from the average IC score, QH IR adjusts the information ratio downwards by the volatility of the monthly IC of the forecasts. The slope *t*-statistic, on the other hand, rather than using the average IC, measures the forecasting accuracy using the average

monthly slope coefficient obtained by regressing the realised returns on the forecasted returns. The average slope coefficient is subsequently adjusted downwards by the standard error of the average slope. In this way both the breadth of the forecasts and the volatility of the forecasting accuracy are incorporated in the slope *t*-statistic.

In dealing with the method for selecting model inputs, the stepwise regression procedure is adapted to select model inputs where the starting input is the style attribute that achieves the best score for the designated performance measurement criterion and the remaining attributes are added to the factor model in each iteration until the designated performance measurement criterion cannot be improved any further. All 38 style attributes are examined to avoid any important attribute being unfairly filtered out based on their respective univariate scores obtained in the previous chapter. Observing the results of the model inputs selected via the stepwise regression procedure reveals that most of the selected style attributes have significant *t*-statistic for the time-series means of their univariate factor payoffs. This suggests that the technique adapted by Van Rensburg and Robertson (2003), Frankish (2004) and Moore (2008) to narrow down the pool of candidate model inputs based on their respective univariate scores is an effective way of filtering out potentially less important style attributes. Nevertheless, the fact that certain attributes with relatively low univariate scores are selected in the pre-specified models means that the in-sample explanatory power can be improved by including all possible style attributes in the pool of candidate model inputs. On the contrary, not all attributes that produce significant univariate scores are included in the pre-specified models since their explanatory contributions are overlapped with other attributes included in the prior iterations of the stepwise regression procedure.

In addressing the problems of multicollinearity that result in overfitting the pre-specified models, the attribute included in the latter iteration is eliminated from the model in the situation where two attributes have correlation coefficient of greater than the absolute value of 0.70. The study results reveal that the problems of overfitting are likely to occur in the last few iterations of the stepwise regression procedure as the rooms for improvements through further inclusions of model inputs are saturated. This phenomenon is indicated by the fact that highly correlated style attributes are selected by the stepwise regression procedure towards the end of the list of model inputs. It is also found that despite the fact that higher

in-sample scores are achieved by the models that do not attempt to remove highly correlated inputs, the models that attempt to address the multicollinearity problems achieved higher out-of-sample scores in 14 out of 16 cases. This indicates that multicollinearity is a major cause of model overfitting, and the resulting consequence is good in-sample, but poor out-of-sample model performance.

To address the possibility of time-varying factor payoffs identified by Haugen and Baker (1996), this study employs the dynamic moving-window approach to update the actual identities of the model inputs. Under the dynamic moving-window approach, the model inputs are updated every 12 months, resulting in 6 overlapping in-sample training periods, and 6 corresponding non-overlapping out-of-sample periods. When the out-of-sample performance of the dynamic models is compared to that of the static models, the underperformance of the static models increases over time, with the exception of the slope *t*-statistic model. The periodic out-of-sample performance of the static slope *t*-static model is in line with that of the dynamic slope *t*-static model. This provides evidence that the style attributes, EY, MOM12-1, CFTP and MOM1 selected by the static slope *t*-static model possess relatively more consistent forecasting power of the cross-sectional equity returns compared to the model inputs of the competing static models over the period from 01 January 2002 to 31 December 2007. The study results suggest that the relative importance of the identities together with the factor payoffs are time-varying in explaining the cross-sectional equity returns on the JSE. As a result, both the identities and the factor payoffs of the expected return linear multifactor models have to be updated to ensure successful forecasts of equity return.

Analysing the periodic performance of the dynamic expected return linear multifactor models reveal that the out-of-sample performance scores of the pre-specified models, constructed under the dynamic moving-window procedure, are consistent with their respective in-sample performance scores. Further to this, the models constructed under their designated performance measurement criteria have great potential in achieving the best in-sample scores for their respective designated criteria. This suggests that the implications of the weakness of the stepwise regression procedure that bias towards the first selected input suggested by Frankish (2004) are not as severe as suggested. However, contrary to the periodic in-sample

results, the periodic out-of-sample performance of the pre-specified models do not achieve relatively better scores for their respective designated performance criteria than the scores achieved by the competing models. More specifically, the average IC model that constructed with an intention to maximise the in-sample average IC score is only being the top 2 average IC performer in 2004 and 2005 over the entire out-of-sample period from 01 January 2002 to 31 December 2007. On the contrary, the Qian and Hua IR model has been the top 2 Qian and Hua IR performers in 5 out of 6 out-of-sample sub-periods. These findings indicate that in-sample performance of a particular model is not a good indication for the model's out-of-sample performance. This implies that some models are more robust than others, and portfolio managers should focus on identifying the best expected return linear multifactor model that outperforms the competing models in the out-of-sample period, rather than choosing the models based on their in-sample performances.

Analysing the consolidated out-of-sample period from 01 January 2002 to 31 December 2007 shows that the dynamic QH IR model also achieves the best scores over the consolidated out-of-sample period for all four performance measurement criteria. Therefore, it can be concluded that the QH IR model is identified as the best linear expected return multifactor model for forecasting the JSE equity returns in this research.

Finally, the models built with the regression constant are found to be robust compared to the original Haugen and Baker (1996) model. This result indicates that the dynamic updating of the regression constant is a better approach in forecasting equity returns than placing a zero constant constraint in modeling equity returns.

7

Nonlinearities in the Cross-Section of Equity Returns: Backpropagation Artificial Neural Networks

7.1 Introduction

This chapter attempts to explore the existence of nonlinearity in the cross-section of equity returns on the JSE for the period from 01 January 1997 to 31 December 2007. This is achieved by constructing and training artificial neural network (ANN) models to capture these nonlinearities. The cascade-correlation architecture, a supervised learning algorithm, together with a backpropagation learning rule, is used to train the ANN models. Motivation for the use of the backpropagation learning rule lies in the fact that, within the framework of artificial neural network modeling, this algorithm has been the most widely accepted approach for dealing with nonlinear problems in the finance and economics domain to date. In addition to the studies on the equity return predictability employing backpropagation ANN models discussed in Chapter 3, the backpropagation learning algorithm, is also employed by Abhyankar and Wong (1997) and Kanas and Yannopoulos (2001) for the forecasting of stock market index returns. Other applications employing this technique include research conducted by Surkan and Singleton (1993) for bond classification; Dutta and Shekhar (1988, 1993) and Utans and Moody (1991, 1995) for bond rating; Marose (1990) for loan assessment; Qi and Maddala (1995) and White (1995) for option pricing; Sen, Oliver and Sen (1995) for merger prediction and Swanson and White (1995a, 1995b) for macroeconomic forecasting.

In addition to the popularity of the backpropagation learning algorithm, are the fixed-architectures common to most multilayered feed-forward neural networks. A fixed-architecture network entails determining the appropriate input neurons, hidden layers and hidden neurons in advance. Of particular importance to neural network modeling, the hidden neuron problem could severely impact the accuracy, robustness and generalisation ability of the model. Evidence has proven that for a fixed-architecture network, a single

hidden layer is enough to approximate any continuous function (Cybenko (1989) and Hornick, Stinchcombe and White (1989)). Still, there lies a problem of determining the number of hidden neurons that will occupy that hidden layer. This will vary from application to application and is largely determined experimentally, using heuristics or by trial and error. A danger of specifying too many hidden neurons is that of network overtraining. The network trains both relevant and irrelevant information in the training set, thus negatively impacting on the generalisation ability of the model. On the other hand, too few hidden neurons could seriously impact the ability of the network to adequately learn the training data. In order to overcome the problems associated with these traditional fixed-architecture backpropagation networks and the hidden layer problem, the cascade-correlation algorithm, was developed by Fahlman and Lebiere (1990). This topology-modifying architecture which includes adding hidden nodes one at a time to the network, trains many different networks until the optimal network is selected. An important feature of using the cascade-correlation algorithm is that "*it removes the requirement to predefine the architecture of the neural network prior to network training*" (Kovalishyn, Tetko, Luik, Kholodovych, Villa and Livingstone, 1998: 651).

Selecting the set of significant input variables that will serve as the best predictor variables when combined together is imperative in ANN modeling. Choosing too many variables as inputs in the network could impact on the efficiency of the network training, possibly resulting in the overfitting of the data. The model may contain relevant variables, as well redundant variables. Mao and Billings (1999: 351) draw attention to the fact that, particularly with nonlinear modeling, increasing the number of (redundant) variables dramatically increases the complexity of the models "*because the size of nonlinear system models increases dramatically with the number of variables*". On the other hand, training too few variables could contribute to the problem of underfitting, as too little information is provided to the network to map the optimal solution. Both problems, overfitting and underfitting, contributes to the poor generalisation ability of the neural network. This research employs a genetic algorithm (Holland (1975) and Goldberg (1989)) to determine the best subset of variables to be used as inputs for the ANN models. Financial application of genetic algorithms includes research conducted by Mahfoud and Mani (1995) for stock selection,

Rutan (1993) for portfolio selection and Kingdom and Feldman (1995) for bankruptcy prediction.

This chapter seeks to examine the effectiveness of combining the cascade-correlation algorithm with an embedded backpropagation learning rule to explore the existence of nonlinearities in the cross-section of JSE equity returns. Several empirical studies, to some degree, have provided ground for research into the nonlinearity of equity returns on the JSE. For example, Fraser and Page (2000) study the momentum effect on the JSE over the period of 1977 to 1997. They find that although momentum shares with relatively higher prior returns earn higher future returns in general, the negative momentum shares earn higher returns than the shares placed in the fourth momentum quintile. Van Rensburg and Robertson (2004) find that the firms in the fifth (smallest) size and value quintiles earn higher returns than the firms in other quintiles over the period from July 1990 to June 2000. However, the average monthly returns earned by the respective quintiles do not appear to be linear. More specifically, the shares in the third size quintile actually earn lower returns than shares in the second size quintile. On the other hand, the average monthly returns are progressively distributed to the shares in the higher value quintile (that is, higher indications of value). If these nonlinearities are found to be apparent, of more relevance then is to determine whether these nonlinearities are robust over time, and how to capture and utilise them in forecasting equity returns and subsequent stock selection.

The models developed in this chapter are all constructed within the cascade-correlation architecture and trained using an embedded backpropagation learning algorithm. These models differ in terms of network architecture, variable selection technique and whether they are trained under the static or moving-window (rolling) procedure. The details of these specifications are explained in Section 7.2. Each model developed is evaluated based on the following performance measurement criteria, namely, the average information coefficient (average IC), Grinold information ratio (Grinold IR), Qian and Hua information ratio (QH IR) and the slope *t*-statistic. The objective of specifying these models is to determine their relative robustness in equity return forecasting.

7.2 Descriptive Statistics and Methodology

Three areas of concern are addressed before the training of the ANN models. These areas include selecting and specifying the data sets, input variable selection and neural network construction and training. These specifications are discussed in Section 7.2.1, Section 7.2.2 and Section 7.2.3, respectively.

7.2.1 Model Design and Specifications

There are two procedures under which the multilayer perceptron (MLP) generates expected returns, namely, the static procedure and the moving-window (rolling) approach developed in this research. The application of the static procedure means that the entire data in the in-sample period (01 January 1997 to 31 December 2001) are trained collectively to predict the out-of-sample equity returns (01 January 2002 to 31 December 2007). The advantage of this procedure is its simplicity and the fact that it is less time-consuming. However, this procedure does not incorporate the dynamic updating of the identities of model inputs after the training period. In addition, the major shortcoming of this procedure is that all data in the entire in-sample period are used to train the model, which may incorporate out-dated data/information in the earlier months. For the purpose of this study, with 72 months of returns to be forecasted in the out-of-sample period, the static procedure is deemed inappropriate.

On the other hand, the dynamic moving-window procedure applied by Kaastra and Boyd (1996), Qi (1999), Quah and Srinivasan (1999) and Eakins and Stansell (2003) rectifies this problem by updating new data/information and discarding the older data/information as time proceeds. The disadvantage of this method is that it is more time-consuming since a series of training and forecasting windows have to be implemented. This study employs the moving-window procedure whereby 60 months data are trained to forecast the immediate 12-month returns following the training period (60-for-12). Under this design, the dataset spanning the

period from 01 January 1997 to 31 December 2001 represents the first training dataset to forecast the 12 months returns in 2002. The second training window covers the dataset spanning the period from 01 January 1998 to 31 December 2002. This training window discards the first 12 months data (that is, the data in 1997) and incorporates new information that becomes available in 2002 to forecast the 12-month returns in 2003. This overlapping procedure is then repeated a further 4 times until all returns in the out-of-sample period is forecasted.

This research first takes on all 38 candidate attributes as potential inputs in the initial training set. These 38 candidates serve as potential inputs in the neural network. The same evaluation criteria of the linear models, namely, the average information coefficient (average IC), Grinold information ratio (Grinold IR), Qian and Hua information ratio (QH IR) and the slope *t*-statistic, implemented in Chapter 6 are used to evaluate the performances of the ANN models. This allows direct comparisons to be made between the performances of the linear expected return factor models and that of the ANN models. As discussed in Chapter 5, selecting model inputs based on their univariate score fails to incorporate the complementary power of the combinations of the attributes in the model. As a result, simplifying the input dataset in this manner shall not be able to improve forecasting results since some important attributes might be filtered out. The detailed description of the ANN models, as well as the data set specifications are described in Part (a) and Part (b) below.

(a) Specifying the Artificial Neural Network (ANN) Models

Employing the dynamic moving-window approach, four ANN models are constructed. These models differ in terms of their variable selection techniques and network architectures (that is, whether there is a direct connection from their input layer neurons to their output layer neuron). The models are presented in (1) to (4).

(1) *Backprop_Direct (SVS)* is a representation of an augmented model, similar to the representation suggested by Gonzales (2000) in the Chapter 2.[1] A linear constraint is placed in that there exists a direct connection from the input layer neurons to the output layer neuron (in addition to the hidden layer connections). This is depicted later in Figure 7.2. The model uses a genetic algorithm to select the best variables as inputs in the neural network. This method of selecting variables is referred to as a standard variable selection *(SVS)* technique.

(2) *Backprop_Direct (CVS)* is also a representation of an augmented model with a direct connection from the input layer neurons to the output layer neuron (in addition to the hidden layer connections), depicted later in Figure 7.2. In terms of variable selection, the model uses a cascade variable selection *(CVS)* technique. This technique includes a "cascade pre-selection" before the standard variable selection step. This cascade pre-selection is similar to the univariate test in Chapter 5 were the variable set is pruned/reduced before the standard variable selection so that only the best variables are combined in the genetic algorithm. The objective is to determine the probability that a variable will be included in an optimum solution and based on this probability, eliminate those variables with low probability, while retaining those with high probability. Specifically, using different initialisation values, a series of runs of the standard variable selection is initiated. Those variables that consistently appear amongst the top variable sets in each run are retained for later runs, while the remainder eliminated. This process is repeated iteratively until no further changes to the variable set takes place. The initial variable set is thus reduced to a smaller more optimal subset of variables to which the standard variable selection is applied.

Both models, *Backprop_Direct (SVS)* and *Backprop_Direct (CVS)* are identical in architecture but differ with respect to variable selection, with a more stringent, time-consuming technique been applied to the latter model. The objective of this is to

[1] Gonzales (2000) suggests an augmented model for a fixed-architecture network. This chapter suggests the use of the augmented model for the cascade-correlation architecture.

determine the validity of applying a more stringent variable selection technique in ANN modeling.

(3) *Backprop_No Direct (SVS)* is a representation where the linear constraint (present in (1) and (2) above) is removed from this model, and thus the direct connection from the input layer neurons to the output layer neuron is absent in this regard. The objective of this is to remove the involuntary imposition of the linear constraint. This is in line with studies conducted by Qi and Maddala (1998), who report that when the linear constraint is removed from the models, the forecasting ability of the ANN models improves. The model uses a genetic algorithm to select the best variables as inputs in the neural network. This standard variable selection *(SVS)* technique is consistent with (1).

(4) *Backprop_No Direct (CVS)* is also a representation where the linear constraint is removed from the model, and thus the direct connection from the input layer neurons to the output layer neuron is absent in this regard. Unlike (3), this model employs a cascade variable selection *(CVS)* technique.

Again, both models, *Backprop_No Direct (SVS)* and *Backprop_No Direct (CVS)* are identical in architecture in that the linear constraint is removed from both models, but differ with respect to variable selection, with a more stringent, time-consuming technique been applied to the latter model.

Comparison can also be drawn between *Backprop_Direct (SVS)* and *Backprop_No Direct (SVS)*. These models have identical variable selection techniques but different network architectures. The same comparison is to be drawn between *Backprop_Direct (CVS)* and *Backprop_No Direct (CVS)*.

(b) Train, Test and Validation Set Selection

When developing a model to project the expected return of each share over the in-sample period, the objective is to maximise the accuracy of the estimation over the in-sample period. As is with the factor selection procedure of the linear factor models, nonlinear models with differential style factors are constructed. Haugen and Baker (1996: 408) emphasise that "*in building an expected return factor model, we must estimate the tendency for stocks with differing exposures to different factors to produce differing returns*". To satisfy this condition, the factor payoffs to style attributes have to be identical across all sample shares in any given month. This will ensure that shares with identical values of the style attributes employed in the model will produce identical expected returns from the model. The cross-sectional factor weights are estimated by training the MLP (multilayer perceptron) of the ANN models. The lagged values of the cross-sectional style attributes are used as the model inputs that correspond to the cross-sectional realised returns of the sample shares (that is, the "target" outputs).

Three data sets are specified, namely, the training set (in-sample set), the test set (in-sample set) and the validation set (out-of-sample set). The purpose of the training set is to map the model parameters. The network weights are thus updated using the training set. The generalisation ability of the learning process is assessed by the training set. The test set is indirectly used in building the model. The test set monitors model performance by observing how well the model interpolates. A technique is adopted whereby training set records are selected from a secondary working set at regular intervals. For all models, this procedure is such that 70% of the in-sample data set is used as the training set while 30% is selected as the test set at regular intervals. The validation set is completely independent of the model building process. The out-of-sample data, which represents the validation set, is manually set aside for verification of the network.

7.2.2 Input Variable Selection

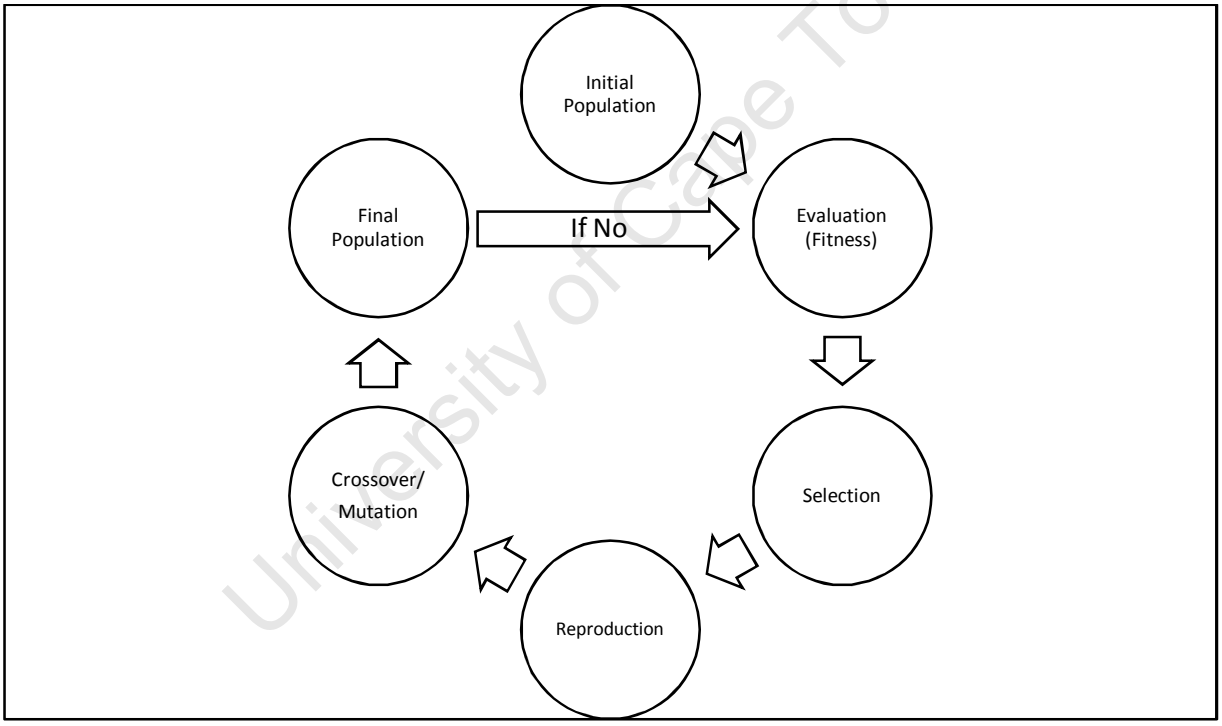
One of the most important steps in building a good ANN model is that of the implementation of a good variable selection technique. This entails selecting those set of input variables that act most synergistically when combined together. No standard variable selection procedure is apparent or consistently used in share return prediction. For example some studies employ stepwise regression, backward stepwise regression (Motiwalla and Wahab, 2000), principal component analysis, as well as selecting variables based on a set of *a priori* assumptions. Mao and Billing (1999) argue that although several available methods are applicable for linear modeling, these approaches cannot be simply extended to nonlinear modeling. For example, the authors attribute the inapplicability of forward selection and backward elimination algorithms to the weight structure of an MLP neural network. They explain, "*these algorithms are in fact term selection and/or deletion methods rather than variable selection and/or deletion algorithms in the nonlinear system case"* (Mao and Billings, 1999: 352). To find the optimal subset of input variables to use for model prediction, a genetic algorithm is employed.

Inspired by the empirical studies in genetics and Charles Darwin's principle of evolution, genetic algorithms are (evolutionary) search procedures which have the capability of modeling complex systems for which there are many solutions. According to the rules of natural selection in Darwinian evolution, only the fittest members of a group will survive (concept popularly known as "survival of the fittest"). These fit members will genetically recombine with other fit members to produce offspring. In this way, good/successful characteristics are passed on to the next generation. Less fit members are discarded, and as a result, the characteristics which they carry will not pass on to the new/next generation (Holland (1975) and Goldberg (1989)). The standard genetic algorithm begins with the simulation of a random population consisting of subsets of input variables. Subsets of input variables in each of the four ANN models are referred to as "individuals" in the evolutionary process. The next step involves each member of the population being evaluated and graded according to a fitness function with the objective of extracting a proportion of the fittest

individuals from the current (known as "parent") population. Through the process of reproduction (crossover and mutation) individuals are selected for the next generation until a particular stopping criterion is attained.

Figure 7.1 The Genetic Algorithm Cycle

The standard genetic algorithm begins with the simulation of a random population consisting of subsets of input variables. Subsets of input variables in each of the four ANN models are referred to as "individuals" in the evolutionary process. During successive iterations, called generations, the initial individuals advance towards fitter individuals by reproduction among members of the previous generation. New generations are created via three genetic operators: selection, crossover and mutation. Only the fittest individuals can crossover or mutate. Individuals are selected for the next generation until a particular stopping criterion is attained.



Fitness is judged for each individual by determining how good the models are (in terms of accuracy and robustness) generated by the combination of variables (Sofge, 2002). The

fitness of each subset of input variables is assessed by employing a linear variable selection model where the objective function is to minimise the root mean power error. This model is discussed below.

(a) Variable Selection Model Specifications

The objective of the genetic algorithm is to determine the optimal subset from the full set of the total variable set to use as inputs in the neural network. Fitness of the individuals is computed via a linear variable selection model. A logistic multiple regression function selects the model inputs. As mentioned in Section 7.2.1, the research employs two types of variable selection techniques for input variable selection. The first type employs the standard variable selection algorithm, while the second type is referred to as a cascade variable pre-selection technique. The cascade variable pre-selection technique is more exhaustive than the standard variable selection technique, since it includes a pre-selection of variables prior to the standard variable selection step. The object of employing a pre-selection step is to reduce the search space by eliminating variables which are unlikely to be included in the optimal solution. Genetic algorithma are encoded in bit strings of "1" and "0" whereby "1" indicates that a specific variable is included and "0" indicates that a variable is eliminated.

(b) Fitness and Selection

The fitness function is the most critical step in genetic algorithms. With the exception of all elite individuals new individuals are either copied from a single fit parent (current individual) to the next generation or are the product of two fit parents that are recombined.[2] The research computes probabilities of the first parent, Parent 1, according to rank, and the second parent, Parent 2, is selected randomly according to a uniform distribution. Computing probabilities according to fitness is generally problematic in generating an optimal solution and thus not implemented. Wallet, Marchette, Solka and Wegman (1996: 3) explain, "*early in the search, when there is a great deal of variety in the population, fitness proportionate weightings tend to heavily favour the few best answers or "super individuals". This leads to an incomplete*

[2] Elite individuals are not involved in selection and reproduction but are automatically carried into the next generation.

search and can often result in convergence to a local maximum. Late in the search, when the goal is to get the last bit of fine tuning, fitness proportional probabilities lack the ability to distinguish between individuals who are very close in fitness". In determining probabilities according to rank, individuals are first arranged according to relative fitness, and selected according to a linear distribution, favouring the higher ranked individuals. A linear rank bias (ranging between 1 and 2) of 1.4 is specified for all four ANN models. The higher this value is, the more bias there is towards choosing the fitter individuals as parents.

(c) Reproduction

Reproduction ensures that only the highest scoring parents will propagate their characteristics to the next generation. The most important genetic operators are the crossover and mutation operators. The crossover operator takes two parent individuals and combines their characteristics to produce new individuals (children/offspring). A crossover probability of 0.7 is specified in the research, specifying the likelihood that the next generation will be as a result of two parents from the previous generation combining and not as a result of replication from a single parent in the previous generation. A way of re-introducing lost variables back into the population is by specifying a mutation factor. The mutation operator is applied after the crossover operator and excludes any elite individuals. A mutation is a random addition or deletion of a characteristic in an individual, and is governed by a preset mutation factor. A mutation factor of 1 percent is specified. Another technique used in this research is the application of an "elitist factor", which determines what fraction of the fittest members of the population will survive to be continued into the next generation via replication. An elitist factor of 0.05 is specified.

(d) Stopping Criterion

Jarvis and Goodacre (2007: 862) suggest three ways that the stopping criterion can be defined. This includes setting the maximum number of iterations (known as generations) that the genetic algorithm must run for, specifying a maximum target outcome value for fitness, or specifying a set number of generations for which the fitness value for the fittest individual remains constant. In specifying the convergence conditions for the genetic algorithm, the

maximum iterations allowed for this procedure is 50 generations for Backprop_Direct (SVS) and Backprop_No Direct (SVS), while maximum iterations specified for Backprop_Direct (CVS) and Backprop_No Direct (CVS) is 800 generations (due to the pre-selection step). Tolerance per generation is used to determine whether the average fitness has improved, while patience places a limit on the number of iterations allowed during variable selection. Under the condition where the average fitness of the population has not improved within the specified tolerance factor (0.001) for 50 iterations, the algorithm is halted.

All 38 variables are selected as potential input variables which may or may not be used as actual inputs in the neural network depending upon the information provided by the genetic algorithm. The proper selection of inputs cannot only reduce the size and complexity of the network, but also lessen the chance of overfitting, thus improving the generalisation of the models.

7.2.3 Neural Network Construction and Training

The structural design of the ANN architecture as well as the training rule influences the manner in which the network is able to learn. The network is constructed within the cascade-correlation architecture. The training procedure, known as cascade learning, entails the employment of an embedded backpropagation learning algorithm. This is explained further in Part (a), Part (b) and Part (c) below.

(a) Cascade-Correlation Architecture

The structural design of the ANN architecture as well as the training rule influences the manner in which the network is able to learn. The neural network is constructed within the cascade-correlation architecture which was introduced by Fahlman and Lebiere (1990). Cascade-correlation serves a two-fold purpose. Firstly, it establishes the cascade-architecture where hidden neurons are added one by one to the network. Once they have been added to the network, they do not change. Secondly, it is a feed-forward, supervised learning algorithm for

training artificial neural networks. It is the job of the learning algorithm to install the hidden neurons. Many different networks are trained until the best one is selected. No *a priori* network architecture needs to be specified in advance. Thus, the advantage of using cascade-correlation is that "*it learns very quickly and the network determines its own size and topology"* (Fahlman and Lebiere (1990) and Diamtopoulou (2005)). The objective of constructing ANN models is to determine the existence of essential relationships in data. More complex patterns would thus require more hidden layers. With cascade-correlation networks, "*there can be any number of hidden nodes, depending on (and increasing with) the complexity of the pattern in the input data"* (Coats and Fant, 1993: 144). Initially there are no hidden nodes in the hidden layer. Hidden nodes are then added to the network one at a time. *"For each new hidden unit, we attempt to maximize the magnitude of the correlation between the new unit's output and the residual signal we are trying to eliminate"* (Fahlman and Lebiere, 1990: 3). The structure of the cascade network is depicted in Figure 7.2.

Figure 7.2 Cascade-Correlation Architectures

The neural network is constructed within the cascade-correlation architecture which was introduced by Fahlman and Lebiere (1990). Referring to CHART (A), input neurons are directly connected to the output neuron as well as hidden neurons. The direct connections capture the linear effects in the input variables. Hidden neurons, on the other hand, produce nonlinear interaction effects among the input variables that are necessary to connect inputs to outputs. Initially there are no hidden nodes in the hidden layer. Hidden nodes are then added to the network one at a time, incrementally growing the network. Referring to CHART (B), the network starts with a minimal configuration (with no hidden layer). Hidden nodes are then added to the network one at a time, incrementally growing the network (and hence the number of hidden layers).

CHART (A) Cascade-Correlation Architecture with Direct Connection

(Backprop_Direct (SVS) / Backprop_Direct (CVS))

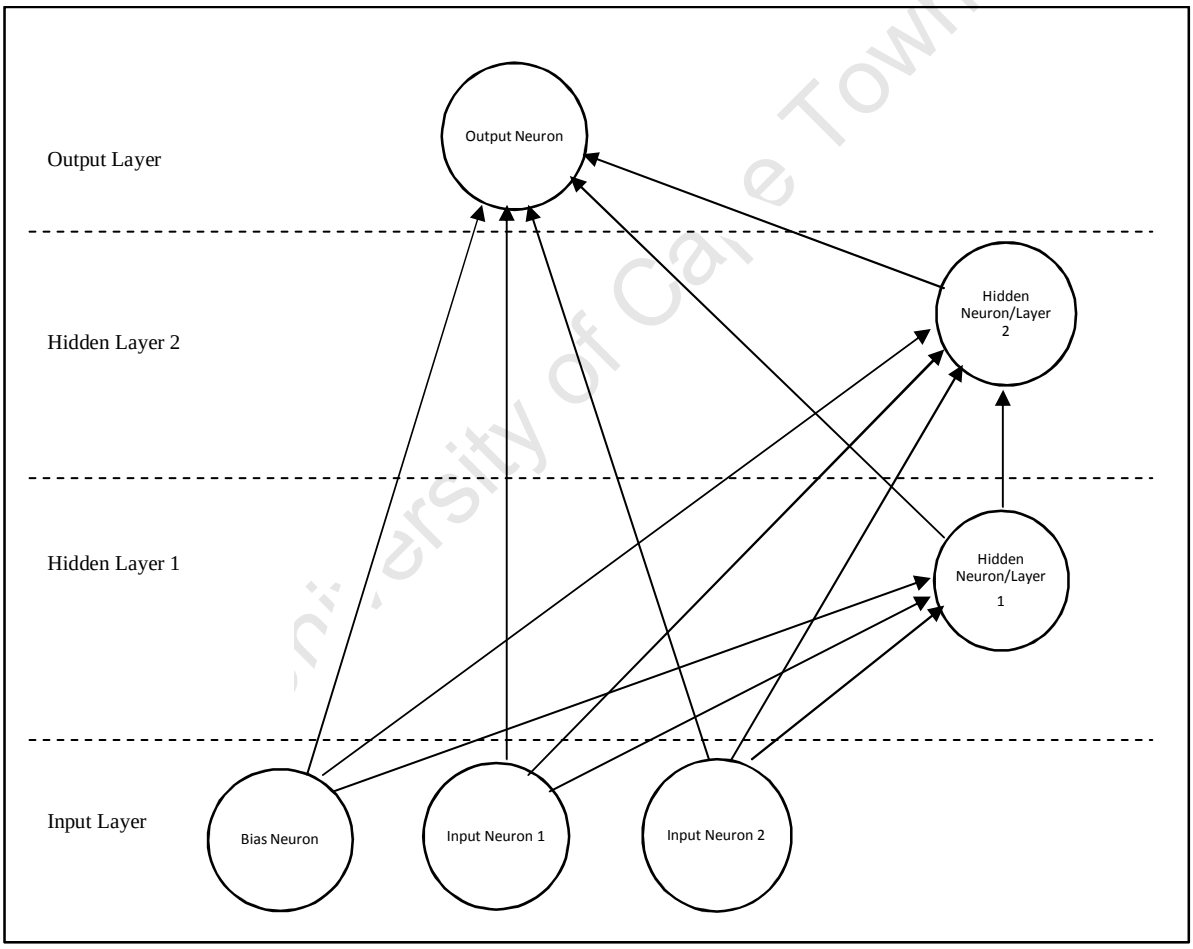
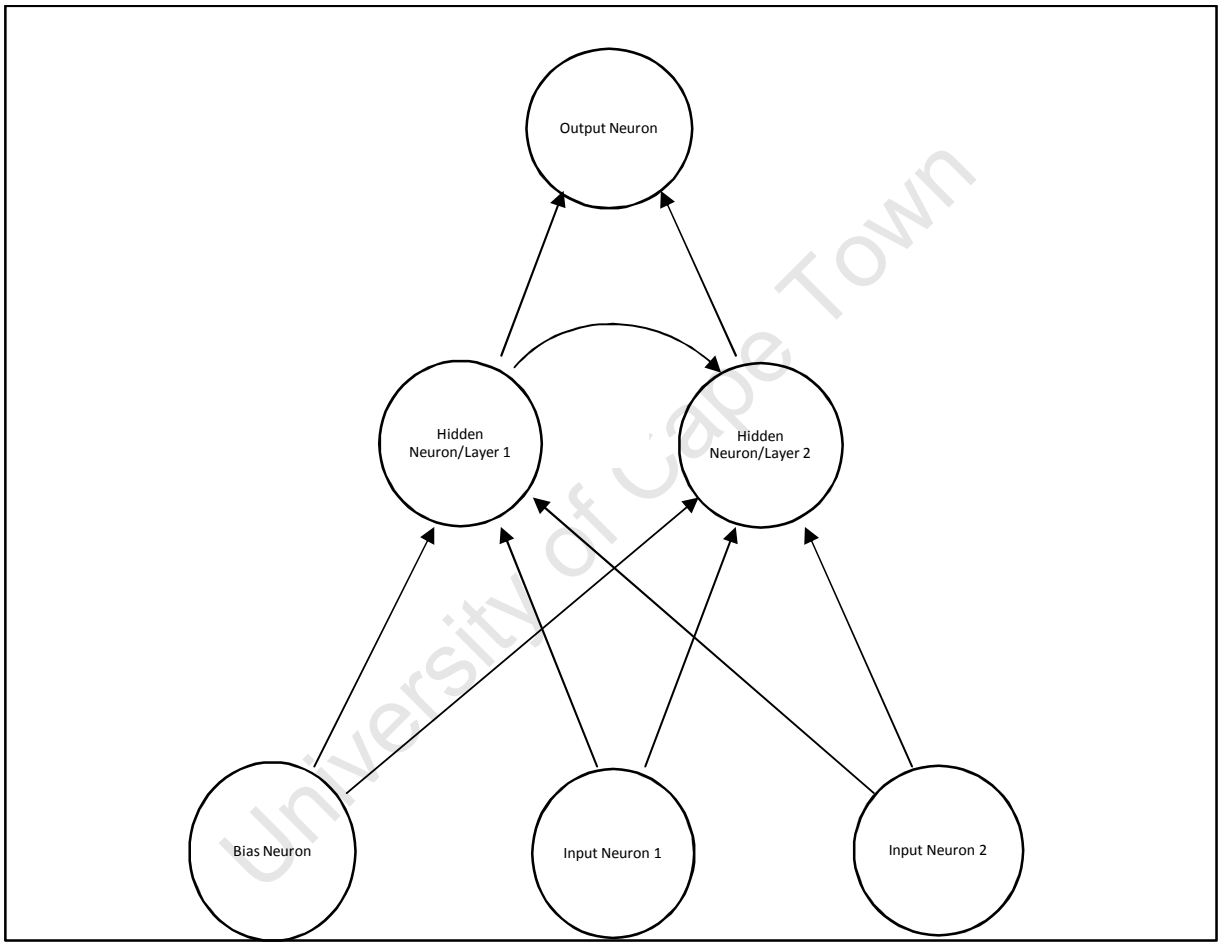


Figure 7.2 Cascade-Correlation Architectures - Continued

CHART (B) **Cascade-Correlation Architecture with No Direct Connection**
(Backprop_No Direct (SVS) / Backprop_No Direct (CVS))



It is important to note that each hidden neuron is fully connected from both input neurons and pre-existing hidden neurons to all existing output neurons. Thus, with a constructive cascaded architecture, the objective is *"to "incrementally" capture the knowledge in the unique structure of hidden nodes and connection weights"* (Coats and Fant, 1993: 145-147). This

cascaded architecture also ensures that the neural network determines the actual number of hidden nodes and connection weights. All four ANN models are trained via the cascade learning algorithm. This algorithm trains the weights using an embedded backpropagation learning rule. Part (b) describes the backpropagation method which will be used within cascade learning to train the network weights.

(b) Backpropagation

The objective of supervised learning is to determine the set of network weights which minimises the mismatch between the network output and the target/desired output. Feed-forward ANN models are generally trained using an algorithm based on standard optimisation techniques, such as backpropagation and Widrow-Hoff learning rules. During training, the network uses a learning rule to adjust the weights so as to be able to more correctly predict the target output of a given set of input data (weights adjusted to minimise the discrepancy between the network output and target values). Backpropagation, which is employed in the research, is a gradient descent algorithm in which the network weights are moved along the negative of the gradient of the performance function.

Artificial neural networks are organised in layers and send the signal "forward". The network error (difference between the network output and the target output) are then propagated "backwards" into the network using the backpropagation algorithm until the network error is minimised. Training begins with random values, between -1 and +1 being assigned to the weights. The objective is to adjust these weights until the error is minimised (determine a global minimum). The output computed by the network depends on the network weights, hidden neurons and the activation function. The activation function (net_k) of the artificial neurons in ANNs can be represented as the sum of the product of the inputs (x_j) and their respective weights (w_{kj}):

$$net_k(\bar{x}, \overline{w}) = \sum_{j=1}^{n} x_j w_{kj} \tag{7.1}$$

This is then passed through a transfer function. The output function employed in the research is the sigmoid function represented as follows:

$$y_k = \frac{1}{e^{-net_k}} \tag{7.2}$$

An important requirement for the activation function for use in the backpropagation algorithm is that it is differentiable. The partial derivative of the sigmoid function with respect to its argument, for example *z*, is derived and represented as follows:

$$\frac{\partial sigmoid\,(z)}{\partial z} = sigmoid\,(z)\big(1 - sigmoid(z)\big) \tag{7.3}$$

To this end, Rojas (1996: 153) explains *"a differentiable activation function makes the function computed by a neural network differentiable, since the network itself computes only function compositions. The error function also becomes differentiable"*. Examining Equation 7.1 and Equation 7.2, it is thus evident that output (y_k) is dependent on the activation function (which in turn depends on the inputs and their respective weights). The weights are thus adjusted at a rate reflective of the curvature of the sigmoid function.

The objective of training the network is to determine a target output from a set of given inputs. The error signal is the difference between the actual activation of the output (y_k) and the target ('desired') activation (t_k) for the neuron. The error is dependent on the weights, thus the weights need to be adjusted in order to minimise the error. This error (cost function) can be represented as follows:

$$E_k = \frac{1}{2}\sum_k (y_k - t_k)^2 \tag{7.4}$$

The goal is to reduce the error by adjusting the weights, thus the partial derivative of the error with respect to the weights is computed (change in the error with respect to the weights or the gradient of the error function). This gradient of *E* in respect of the weights *w* is:

$$\nabla E_k(w_{kj}) = \frac{\partial E}{\partial w_{kj}} \tag{7.5}$$

When applying the backpropagation technique, knowledge is accumulated through the learning process. According to Kuan and White (1994), knowledge is incarnated in these weights. The network weights are updates as follows:

$$w_{kj+1} = w_{kj} + \Delta w_{kj} \tag{7.6}$$

Δw_{kj} (representing the training process/adjustment of each weight) is the vector of all weights of the network.

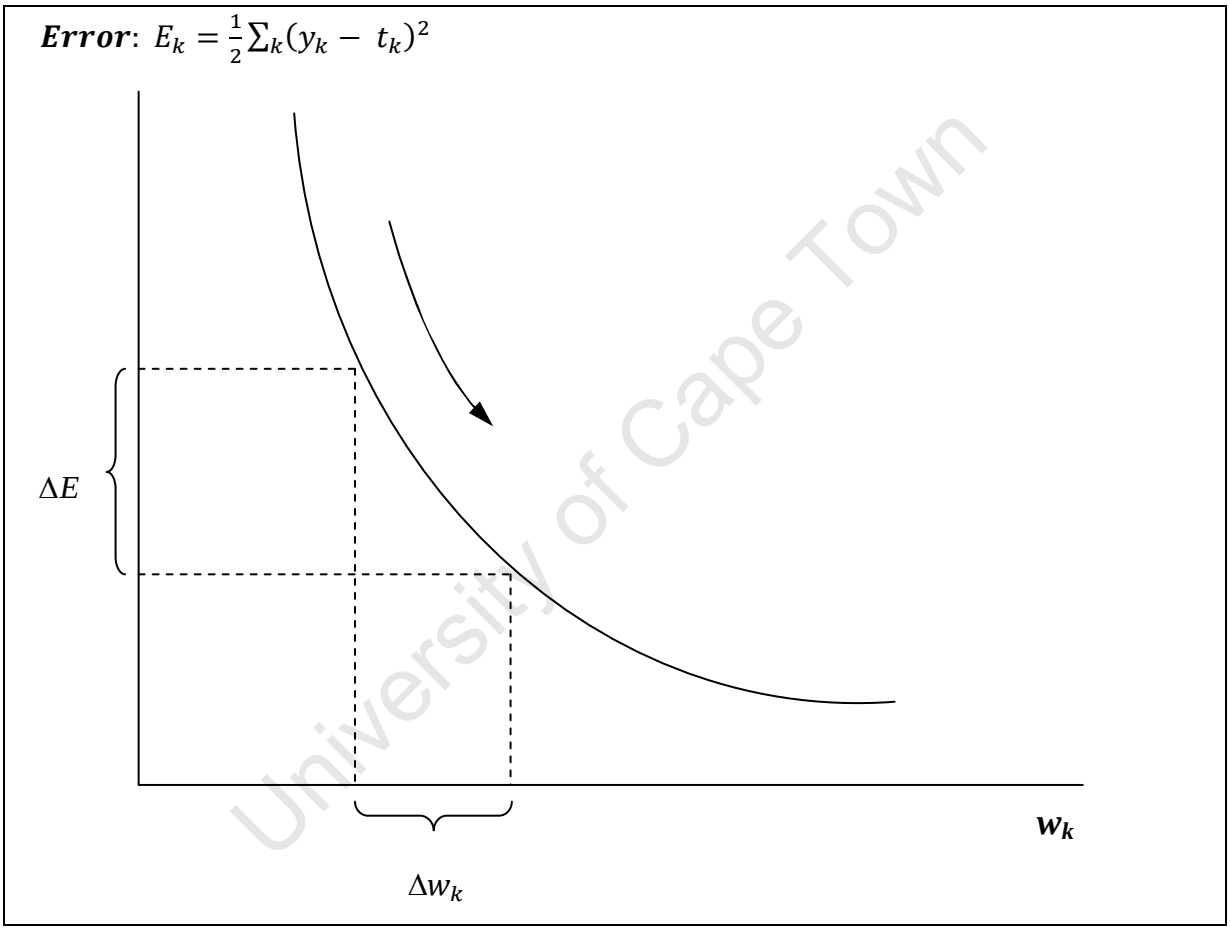
The backpropagation algorithm adjusts the weights using the method of *gradient descent:*

$$\Delta w_{kj} = -\eta \frac{\partial E}{\partial w_{kj}} \tag{7.7}$$

The learning rate (η) determines how the weights change at each step length of each iteration. If η is too small, the algorithm will take too long to converge. On the other hand, if η is too large, the increment jumps may end up way beyond the error surface and the algorithm may not converge at all. The learning rate must lie between 0 and 1 ($0 \leq \eta \leq 1$). The size of the weight adjustment is thus largely dependent on the learning rate and on the contribution of the weight to the error function. The weights (Equation 7.7) are adjusted until the appropriate weights, which minimise the error function, are determined. The negative sign indicates that the weight changes are in the direction of the decrease in error. This is depicted in Figure 7.3.

Figure 7.3 The Backpropagation Algorithm

The weigh adjustment process is an application of the backpropagation algorithm/method of gradient descent (going downhill on the error surface). The network error *(E)*, which is a function of the weights *(w)* decreases in the direction of negative gradient. ΔE represents a change in the quadratic error for a unit change in the weights,Δw_k. If the slope of $E(w_k)$ is negative, the more the weight w_k would need to increase. On the other hand, if the slope of $E(w_k)$ is positive, the more the weight w_k would need to decline. This is controlled by the learning rate η, which determines how much the weight changes at each step.



Backpropagation thus depends on the learning rate, the sum of squared error and the activation function. Any network can be trained as long as its weights, its net inputs and the transfer function are differentiable. The objective being, to find the derivative of E with

respect to w_{kj}, backwards. In an attempt to find this, the first step is to determine the extent to which the error depends on the output. Thus referring to Equation 7.4, the partial derivative of the error with respect to the output can be represented as follows:

$$\frac{\partial E}{\partial y_k} = (y_k - t_k) \tag{7.8}$$

Next, expanding the partial derivative $\frac{\partial y_k}{\partial w_{kj}}$ by the chain rule, determine how much the output (y_k) depends on the activation (net_k) and how this activation depends on the weights. Referring to Equation 7.1 and Equation 7.2:

$$\frac{\partial y_k}{\partial w_{kj}} = \frac{\partial y_k}{\partial net_k}\frac{\partial net_k}{\partial w_{kj}} = y_k(1 - y_k)x_j \tag{7.9}$$

Thus, combining Equation 7.8 and Equation 7.9, the partial derivative of the error with respect to the weights can be represented as follows:

$$\frac{\partial E}{\partial w_{kj}} = \frac{\partial E}{\partial y_k}\frac{\partial y_k}{\partial w_{kj}} = (y_k - t_k)y_k(1 - y_k)x_j \tag{7.10}$$

Rewriting Equation 7.7, the adjustment to each weight is now derived as follows:

$$\Delta w_{kj} = -\eta(y_k - t_k)y_k(1 - y_k)x_j \tag{7.11}$$

(c) Cascade Learning

Cascade learning can be divided into two repetitive phases. The first phase involves the cascade-architecture, while the second phase involves the recruitment of a "candidate" neuron to the growing network architecture. In the first phase, the network starts with an input and an output layer and no hidden neurons (layers). All the connections (connecting input to output layer) are trained using an embedded backpropagation learning algorithm (described in Part (b) above). This training continues until the error of the network no longer declines. If no significant error reduction has occurred within a specified number of epochs (training cycles), and there exists some residual error that needs to be further reduced, a hidden neuron (layer) may be added to the network. Once a hidden neuron is installed (activated), its input connections are frozen and the neuron is in no way altered/modified. The network is trained again and if no further reduction in the error occurs, another hidden neuron (in a new hidden layer) is installed, with this new neuron connected to the previous hidden neuron, resulting in the cascaded architecture. This process continues until the global minimum is established or a final architecture is obtained. After each hidden neuron is installed, training of the output weight resumes.

An important consideration is the determination of the hidden neuron (described in the first phase) to be installed in the hidden layer. Before a neuron is "recruited" for installation in the hidden layer, there may exist a number of "candidate" neurons, each possessing the potential to qualify as a hidden neuron. The second phase thus involves the analysis of a candidate neuron and can be divided into two parts. The first part is referred to as *input-to-hidden-layer training*, while the second part of this phase is referred to as *hidden-to-output-layer training*. *Input-to-hidden-layer training* starts with the generation of the candidate neurons. Each candidate neuron is connected with all input neurons and all existing hidden neurons by trainable input connections. At this stage, there exist no connection weights between the candidate neurons and the output neurons. The weights (on the input side) of the candidate neurons are trained using the backpropagation algorithm, with the objective of maximising the correlation between the residual error of the network and the activation of the candidate neurons. Training halts when no further improvements to the

correlation score results. The candidate unit with the maximum correlation is installed in the network as a hidden neuron. The goal of the adjustment is to maximise *S*, the sum over all input units o of the magnitude of the correlation *V*, the candidate unit's value, and E_o, the residual output error observed at unit *o*. *S* is defined as follows (Fahlman and Lebiere, 1990) :

$$S = \sum_{o} \left| \sum_{p} (V_p - \bar{V})\,(E_{p.o} - \overline{E_o}) \right| \tag{7.12}$$

Where *o* is the network output at which the error is measured and *p* is the training pattern. The quantities $\bar{V}$ and $\overline{E_O}$ are the values of *V* and E_o averaged over all patterns.

Training is stopped if the correlation ceases to improve or a predefined number of cycles are exceeded. The final step of the second phase is the installment of the candidate neuron as a hidden neuron in the network. The newly trained weights are frozen and no further adjustments or modifications made to these weights. Freezing the input weights ensures that it will always be able to track the aspects of the network error that it was trained to track. The new hidden neuron now has trainable weights connecting it with the output neurons. These weights now undergo training (*hidden-to-output-layer training*). The addition of the hidden neurons continues until the learning process is complete.

With regard to all four ANN models constructed in this chapter, in constructing the hidden layer, a minimum increment is set to 1, while the maximum increment is set to 2. The minimum increment indicates the minimum number of hidden nodes to add at a time while the maximum increment indicates the maximum number of hidden neurons to add at a time. In the quest for a more accurate model, a sigmoid output layer function is chosen over a linear output layer function.

7.3 Results

Table 7.1 presents the periodic variable selection process of the backpropagation models for each of the six overlapping training (in-sample) periods. The backpropagation models that undergo the same variable selection technique identify the same set of variables regardless of whether they have the direct connections in their network architecture.

When a variable is included by the algorithm as a selected input in the neural network, it is assigned with a letter "I". By contrast, variables that fail to pass the standard variable selection procedure are assigned with a letter "V". For the models that pre-select variables using the cascade variable pre-selection technique, variables eliminated in the pre-selection step are assigned with a letter "C" without going through the standard variable selection procedure.

Over the six overlapping training periods spanning from 01 January 1997 to 31 December 2006, book value-to-price (BVTP) and cash flow-to-price (CFTP) are selected by all backpropagation models in all training periods. As expected, the cascade pre-selection technique drastically reduces the number of variables selected by the models compared to the standard variable selection technique over the six training periods.

Table 7.2 demonstrates the number of training periods in which each of the 38 attributes are selected by the backpropagation models with and without using the cascaded pre-selection technique developed in this chapter. Besides BVTP and CFTP mentioned above, the 1-month lagged 11-month return momentum (MOM12-1), 24-month return momentum (MOM24), log value of market capitalisation (LSIZE) and dividend yield (DY) are commonly included by both variable selection techniques. By contrast, interest coverage before tax (ICBT), together with EARNREV, is excluded by both variable selection techniques. In summary, attributes from the value and momentum categories dominate the variables selected by all backpropagation models.

Table 7.1 Periodic Variable Selections of the Backpropagation Models

Table 7.1 documents the complete variable selection process of the backpropagation models for each of the six overlapping training periods over the period spanning from 01 January 1997 to 31 December 2006. The backpropagation models are divided into models with or without using the cascade pre-selection procedure, in addition to the existing standard variable selection procedure. When a variable is included by the algorithm as a selected input in the neural network, it is assigned with a letter "I". By contrast, variables that fail to pass the standard variable selection procedure are assigned with a letter "V". For the models that pre-select variables using the cascade pre-selection technique, variables eliminated in the pre-selection stage are assigned with a letter "C" without going through the standard variable selection procedure.

In-Sample Period:	**1/1/97 ~ 31/12/01**		**1/1/98 ~ 31/12/02**		**1/1/99 ~ 31/12/03**		**1/1/00 ~ 31/12/04**		**1/1/01 ~ 31/12/05**		**1/1/02 ~ 31/12/06**	
	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection
BVTP	/	/	/	/	/	/	/	/	/	/	/	/
CFTCURRLIABS	I	V	V	C	I	V	V	C	V	V	I	C
CFTDEBT	V	V	I	V	I	V	V	C	V	C	V	C
CFTP	/	/	/	/	/	/	/	/	/	/	/	/
CURRENTRATIO	I	I	V	C	I	V	V	C	I	V	V	V
DEBTTBVE	I	I	V	V	V	V	V	C	I	V	V	V
DEBTTMVE	I	V	V	C	V	C	I	C	V	C	V	C
DY	I	I	I	V	V	I	I	C	I	V	V	C
EARNREV	V	V	V	V	V	V	V	V	V	V	V	V
EG1	V	C	V	C	I	C	I	V	V	V	V	I
EY	I	V	I	C	I	V	V	I	V	V	I	C
FOREY1	V	I	V	C	V	C	V	I	I	V	V	V
FOREY2	V	I	V	V	I	C	V	C	V	C	V	V
G12MCPS	V	V	V	C	V	C	V	V	V	C	I	V
G12MDPS	I	V	I	I	V	I	V	V	I	C	V	C
G12MEPS	I	V	I	V	I	C	V	C	I	I	I	I
G12MGPMARGIN	I	I	V	V	V	C	V	C	V	C	V	C
G12MNPMARGIN	V	C	I	V	V	V	V	C	V	C	I	V
G12MSALES	V	V	V	C	V	C	V	I	V	C	V	C
G24MEPS	V	V	V	C	I	C	I	V	V	V	V	V
GPMARGIN	V	C	I	C	V	C	V	C	V	C	I	C
GROWTH	V	I	V	C	I	C	V	C	V	V	V	C
ICBT	V	C	V	C	V	V	V	C	V	V	V	V
LAGLPRICE	I	C	I	V	V	C	I	V	I	V	V	V
LPRICE	V	C	I	C	V	C	V	V	I	V	V	V
LSIZE	I	C	V	C	I	V	V	V	I	I	I	I
MOM1	I	I	I	V	V	I	V	C	V	C	I	I
MOM12	V	C	V	I	V	V	V	V	I	V	I	V
MOM12-1	I	I	I	V	I	I	I	V	V	I	V	I
MOM24	I	V	V	I	I	V	I	I	I	I	V	V
MOM3	V	V	V	C	V	C	I	C	V	C	V	C
MOM6	I	C	V	V	I	I	V	V	I	V	I	V
NPMARGIN	V	V	I	V	V	C	I	V	V	C	V	V
PAYOUT	V	V	I	I	I	V	V	V	V	C	V	C
ROA	V	V	I	C	I	C	V	V	V	C	V	C
ROE	I	C	V	C	V	C	V	V	V	I	V	C
SALESTP	V	I	I	V	V	V	V	V	I	C	V	V
TATURNOVR	I	V	I	C	I	C	V	C	I	V	V	V
Total:	**19**	**12**	**18**	**6**	**18**	**7**	**11**	**6**	**16**	**7**	**12**	**7**

Table 7.2 Frequency of Inclusion for the Candidate Style Attributes by the Standard and the Cascade Pre-Selection Techniques in the Backpropagation Models

Table 7.2 documents the number of training periods in which each of the candidate style attributes is included by the backpropagation models with and without using the cascade pre-selection technique respectively. The variables are arranged in descending order according to their frequency of inclusions in the respective models.

Standard Variable Selection Technique		**Cascade Pre-Selection Technique**	
Firm-Specific Attributes	**Frequency of Inclusion (Out of 6)**	**Firm-Specific Attributes**	**Frequency of Inclusion (Out of 6)**
BVTP	***6***	***BVTP***	***6***
CFTP	***6***	***CFTP***	***6***
G12MEPS	5	MOM12-1	4
DY	4	MOM24	3
EY	4	MOM1	3
LAGPRICE	4	G12MEPS	2
LSIZE	4	DY	2
MOM12-1	4	LSIZE	2
MOM24	4	G12MDPS	2
MOM6	4	FOREY1	2
TATURNOVER	4	EY	1
CFTCURRLIABS	3	MOM6	1
CURRENTRATIO	3	CURRENTRATIO	1
G12MDPS	3	DEBTTBVE	1
MOM1	3	EG1	1
CFTDEBT	2	MOM12	1
DEBTTBVE	2	PAYOUT	1
DEBTTMVE	2	SALESTP	1
EG1	2	FOREY2	1
G12MNPMARGIN	2	G12MGPMARGIN	1
G24MEPS	2	GROWTH	1
GPMARGIN	2	ROE	1
LPRICE	2	G12MSALES	1
MOM12	2	LAGLPRICE	0
NPMARGIN	2	TATURNOVER	0
PAYOUT	2	CFTCURRLIABS	0
ROA	2	CFTDEBT	0
SALESTP	2	DEBTTMVE	0
FOREY1	1	G12MNPMARGIN	0
FOREY2	1	G24MEPS	0
G12MCPS	1	GPMARGIN	0
G12MGPMARGIN	1	LPRICE	0
GROWTH	1	NPMARGIN	0
MOM3	1	ROA	0
ROE	1	G12MCPS	0
EARNREV	0	MOM3	0
G12MSALES	0	EARNREV	0
ICBT	0	ICBT	0

Table 7.3 demonstrates the detailed periodic performances over the six overlapping training periods covered by the in-sample and the corresponding out-of-sample periods for the respective backpropagation models. There are two models constructed under each of the standard variable selection procedure and the cascade pre-selection procedure: a backpropagation model with direct connection from the input layer to the output layer in its network architecture and a backpropagation model that does not enforce a direct connection from its inputs to its output.

A common observation in Table 7.3 across all four backpropagation models is that the out-of-sample performance scores in the first three years of the out-of-sample period from 01 January 2002 to 31 December 2004 are more or less in line with their corresponding in-sample scores (refer to PANEL (A) to PANEL (C) in Table 7.3). Although the models that use the cascade pre-selection technique (model 2 and model 4) have lower in-sample performance scores compared to their standard variable selection counterparts (model 1 and model 3), they outperform the models with standard variable selection over the first three years of the out-of-sample period from 01 January 2002 to 31 December 2004. The fact that the models using the cascade pre-selection technique achieves better out-of-sample results with small set of variables serves as evidence that the large number of inputs selected by the standard variable selection technique have severely overfitted model 1 and model 3, which results in the loss of robustness in the models' forecasting power.

By contrast, the out-of-sample performance scores for the second half of the out-of-sample period from 01 January 2005 to 31 December 2007 are drastically lower than their corresponding in-sample scores (refer to PANEL (D) to PANEL (F) in Table 7.3). The models using the cascade pre-selection technique are also less robust during this period.

With regard to the choice of direct connection in the neural network architecture, no evidence is found in support of Qi and Maddala (1998)'s argument that removing the direct connection constraint can effectively improve the model's forecasting power.

Table 7.3 Periodic Performance Scores of the Backpropagation Models

The performance scores of the ANN backpropagation models for the six overlapping training periods and their corresponding out-of-sample performances are displayed in PANEL (A) through PANEL (F) respectively. The backpropagation models using the cascade pre-selection technique are assigned with CVS. On the other hand, the models applying the standard variable selection technique are assigned with SVS. The best out-of-sample scores in each sub-period are highlighted in bold.

PANEL (A) In-Sample: 1 Jan 1997 to 31 Dec 2001; Out-of-Sample: 1 Jan 2002 to 31 Dec 2002

	(1) Backprop_ Direct (SVS)	(2) Backprop_ Direct (CVS)	(3) Backprop_ No Direct (SVS)	(4) Backprop_ No Direct (CVS)
Avg IC Score (In-Sample)	0.1378	0.1345	0.1488	0.1578
Avg IC Score (Out-Sample)	0.0644	**0.0786**	0.0337	0.0731
Grinold's IR Score (In-Sample)	1.4955	1.4677	1.6158	1.7092
Grinold's IR (Out-Sample)	0.7297	**0.8900**	0.3845	0.8272
Qian and Hua's IR Score (In-Sample)	0.9056	0.7912	0.9969	0.9446
Qian and Hua's IR (Out-Sample)	0.5230	**0.9330**	0.2050	0.7848
Slope t-Statistic Score (In-Sample)	6.2410	4.8212	7.4503	6.3231
Slope t-Statistic (Out-Sample)	2.0082	**3.1267**	0.9236	2.6833

PANEL (B) In-Sample: 1 Jan 1998 to 31 Dec 2002; Out-of-Sample: 1 Jan 2003 to 31 Dec 2003

	(1) Backprop_ Direct (SVS)	(2) Backprop_ Direct (CVS)	(3) Backprop_ No Direct (SVS)	(4) Backprop_ No Direct (CVS)
Avg IC Score (In-Sample)	0.1466	0.1452	0.1495	0.1423
Avg IC Score (Out-Sample)	0.1626	0.1647	0.1560	**0.1677**
Grinold's IR Score (In-Sample)	1.6121	1.5928	1.6448	1.5612
Grinold's IR (Out-Sample)	1.8741	1.8983	1.7992	**1.9331**
Qian and Hua's IR Score (In-Sample)	0.9679	0.9623	0.9502	0.9472
Qian and Hua's IR (Out-Sample)	**1.5812**	1.4631	1.5273	1.5330
Slope t-Statistic Score (In-Sample)	5.9580	6.1664	5.9567	5.9517
Slope t-Statistic (Out-Sample)	**5.5937**	4.9740	5.2204	5.3190

PANEL (C) In-Sample: 1 Jan 1999 to 31 Dec 2003; Out-of-Sample: 1 Jan 2004 to 31 Dec 2004

	(1) Backprop_ Direct (SVS)	(2) Backprop_ Direct (CVS)	(3) Backprop_ No Direct (SVS)	(4) Backprop_ No Direct (CVS)
Avg IC Score (In-Sample)	0.1888	0.1811	0.1901	0.1910
Avg IC Score (Out-Sample)	0.1787	**0.2000**	0.1579	0.1896
Grinold's IR Score (In-Sample)	2.1023	2.0136	2.1160	2.1224
Grinold's IR (Out-Sample)	2.0893	**2.3411**	1.8464	2.2190
Qian and Hua's IR Score (In-Sample)	1.3640	1.3245	1.3420	1.1527
Qian and Hua's IR (Out-Sample)	1.0449	1.1626	0.9379	**1.2435**
Slope t-Statistic Score (In-Sample)	9.8169	9.1077	9.4149	7.5992
Slope t-Statistic (Out-Sample)	3.7584	3.9306	3.3928	**4.5603**

Table 7.3 Periodic Performance Scores of the Backpropagation Models - Continued

PANEL (D) In-Sample: 1 Jan 2000 to 31 Dec 2004; Out-of-Sample: 1 Jan 2005 to 31 Dec 2005

	(1) Backprop_ Direct (SVS)	(2) Backprop_ Direct (CVS)	(3) Backprop_ No Direct (SVS)	(4) Backprop_ No Direct (CVS)
Avg IC Score (In-Sample)	0.2212	0.2366	0.2200	0.2398
Avg IC Score (Out-Sample)	0.1175	0.1036	**0.1177**	0.07926
Grinold's IR Score (In-Sample)	2.5100	2.6857	2.4954	2.7183
Grinold's IR (Out-Sample)	1.3932	1.2275	**1.3957**	0.9390
Qian and Hua's IR Score (In-Sample)	1.5705	1.5587	1.5313	1.4982
Qian and Hua's IR (Out-Sample)	**0.9121**	0.5532	0.8336	0.42157
Slope t-Statistic Score (In-Sample)	11.8852	10.1732	11.8123	10.3992
Slope t-Statistic (Out-Sample)	**3.3727**	1.7709	2.9159	1.2685

PANEL (E) In-Sample: 1 Jan 2001 to 31 Dec 2005; Out-of-Sample: 1 Jan 2006 to 31 Dec 2006

	(1) Backprop_ Direct (SVS)	(2) Backprop_ Direct (CVS)	(3) Backprop_ No Direct (SVS)	(4) Backprop_ No Direct (CVS)
Avg IC Score (In-Sample)	0.2032	0.2112	0.2105	0.2150
Avg IC Score (Out-Sample)	0.0678	0.0900	0.0772	**0.09215**
Grinold's IR Score (In-Sample)	2.3317	2.4261	2.4174	2.4713
Grinold's IR (Out-Sample)	0.8231	1.0775	0.9361	**1.11601**
Qian and Hua's IR Score (In-Sample)	1.2166	1.3305	1.4249	1.3391
Qian and Hua's IR (Out-Sample)	0.3816	0.6084	0.7443	**0.7679**
Slope t-Statistic Score (In-Sample)	9.1518	9.7059	10.7882	9.6092
Slope t-Statistic (Out-Sample)	1.0253	1.6781	**2.4151**	2.2068

PANEL (F) In-Sample: 1 Jan 2002 to 31 Dec 2006; Out-of-Sample: 1 Jan 2007 to 31 Dec 2007

	(1) Backprop_ Direct (SVS)	(2) Backprop_ Direct (CVS)	(3) Backprop_ No Direct (SVS)	(4) Backprop_ No Direct (CVS)
Avg IC (In-Sample)	0.1861	0.2327	0.1849	0.2080
Avg IC Score (Out-Sample)	0.0767	0.0514	**0.0837**	0.0794
Grinold's IR Score (In-Sample)	2.1686	2.7120	2.1547	2.4244
Grinold's IR (Out-Sample)	0.9500	0.6419	**1.0372**	0.9867
Qian and Hua's IR Score (In-Sample)	1.1679	1.2745	1.1698	1.4675
Qian and Hua's IR (Out-Sample)	0.5218	0.3179	0.5726	**0.6260**
Slope t-Statistic Score (In-Sample)	7.8232	8.2905	7.9857	10.2135
Slope t-Statistic Score (Out-Sample)	1.2055	0.6186	1.3504	**1.40922**

Figure 7.4 provides the periodic in-sample scores achieved by the four backpropagation models for each of the performance measurement criteria as illustrated in CHART (A) to CHART (D). The in-sample performance scores for the backpropagation models are competitive of each other. Periodically, the in-sample scores for all of the performance criteria improve consistently over the first four overlapping training periods, but deteriorate thereafter for the last two overlapping training periods. Overall, the backpropagation models that apply the cascade pre-selection technique without direct connection in the network architecture (green histograms) achieves the most consistent scores for all of the performance measurement criteria over the overlapping training periods.

Figure 7.5 compares the periodic performance scores of the backpropagation models constructed under the dynamic (or rolling) moving-window procedure (refer to the histograms) to the performance scores of the backpropagation models trained under the static procedure (refer to the trend lines). The comparisons of the out-of-sample performances between the rolling approach and the static approach serves to determine whether updating the identities of the model inputs periodically can effectively improve the robustness of the forecasting power of the backpropagation models. Examining the periodic out-of-sample performances of the rolling backpropagation models indicates that their out-of-sample average IC scores (refer to CHART (A) of Figure 7.5) and Grinold IR scores (refer to CHART (B) of Figure 7.5) improve over time and peak in 2004 (the third year of the out-of-sample period). On the other hand, the out-of-sample QH IR scores (refer to CHART (C) of Figure 7.5) and the slope *t*-statistics (refer to CHART (D) of Figure 7.5) peak in 2003 (the second sub-period) and deteriorate afterwards. The declining in the forecasting accuracy of the rolling backpropagation models in the second half of the out-of-sample period is in direct contrast to their outstanding in-sample scores during the same period. On the other hand, the performance scores achieved by the static models are inconsistent over time across all performance measurement criteria. However, there is no evidence that the rolling backpropagation models outperform their static counterparts. This might be due to the general poor forecasting power of the backpropagation models in the second half of the out-of-sample period, which makes the comparison between the rolling approach and the static approach less meaningful.

Figure 7.4 Periodic In-Sample Performance Scores of the Backpropagation Models Trained under the Dynamic Moving-Window Procedure

CHART (A) to CHART (D) illustrate the in-sample performance scores for the ANN backpropagation models trained under the dynamic (rolling) moving-window procedure for each of the four performance measurement criteria. The performance scores of the ANN models in each chart are represented by the red, yellow, blue and green bars in the form of histograms.

CHART (A) In-Sample Average IC Scores

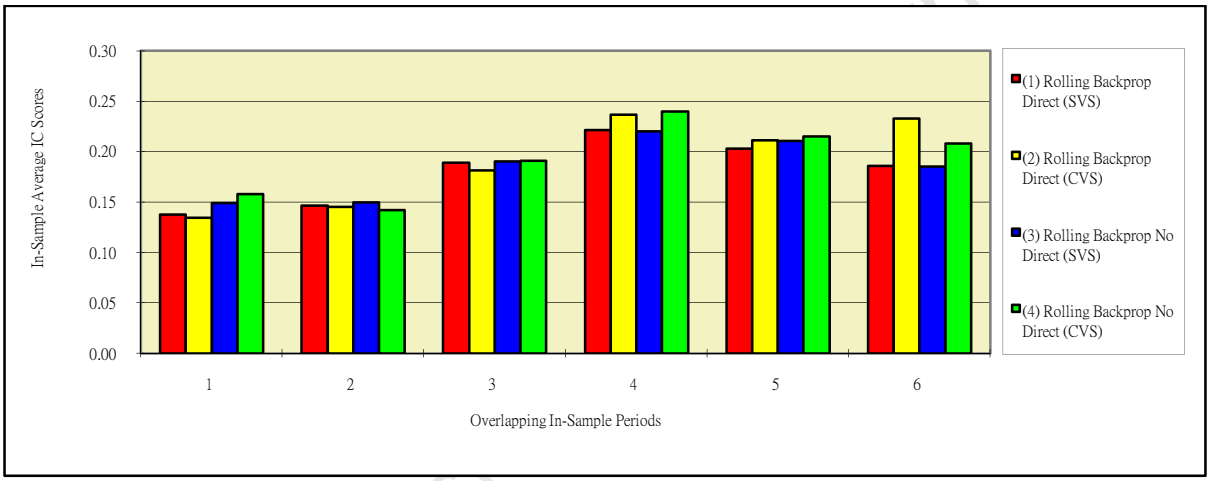


CHART (B) In-Sample Grinold IR Scores

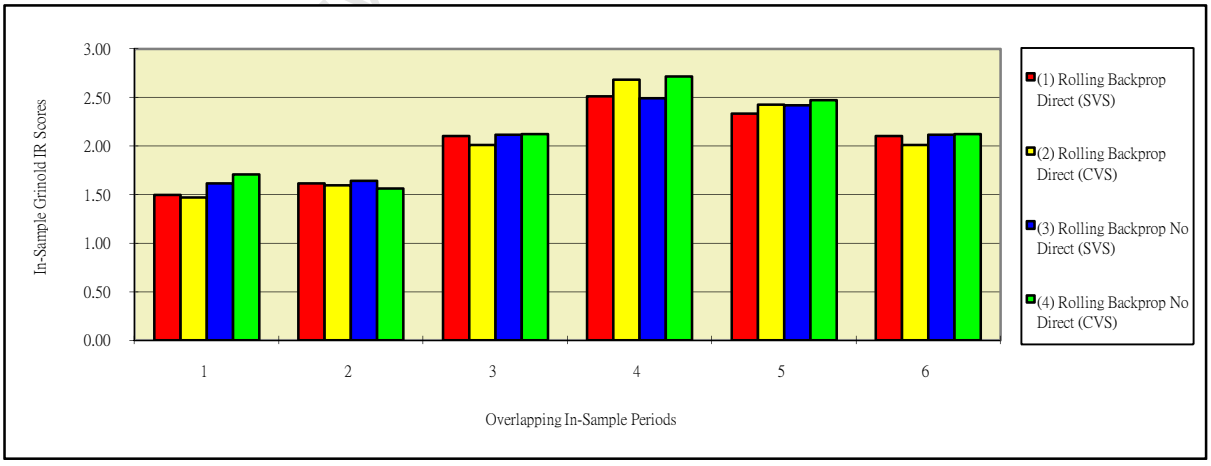


Figure 7.4 Periodic In-Sample Performance Scores of the Backpropagation Models Trained under the Dynamic Moving-Window Procedure - Continued

CHART (C) In-Sample Qian and Hua IR Scores

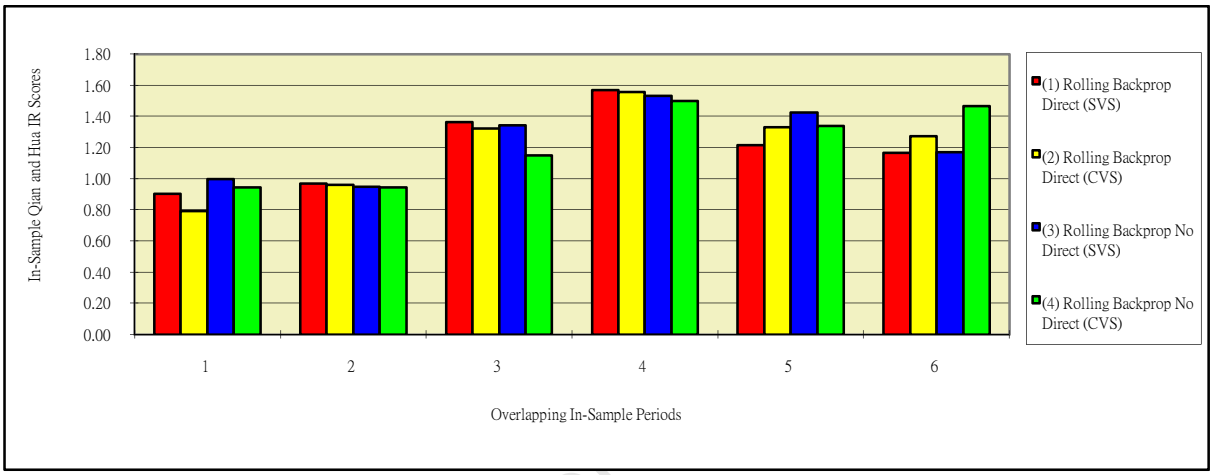


CHART (D) In-Sample Slope *t*-Statistic Scores

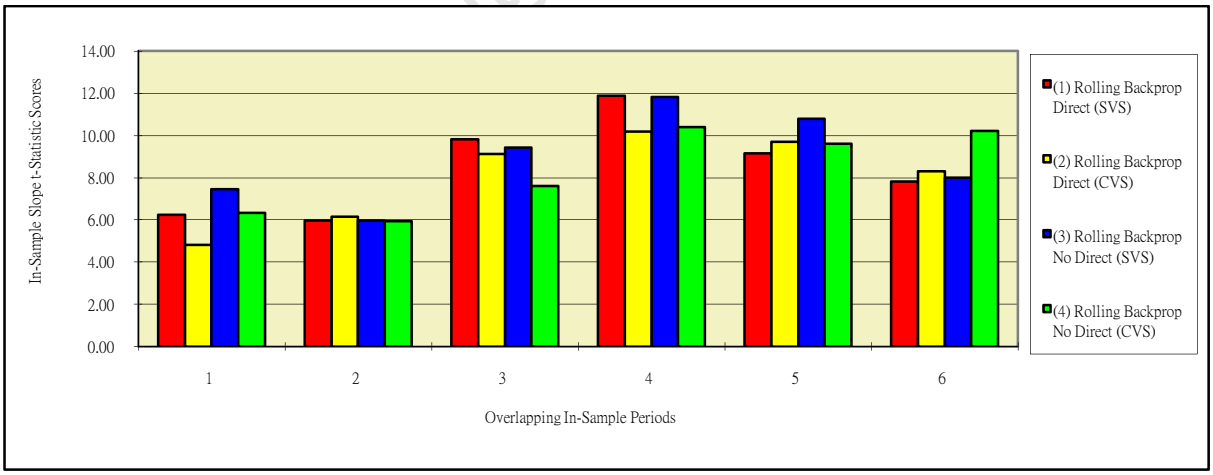


Figure 7.5 Periodic Out-of-Sample Performance Scores of the Backpropagation Models

CHART (A) to CHART (D) illustrate the periodic out-of-sample performances of the ANN backpropagation models under each of the four performance measurement criteria. The performance scores of the models constructed under the moving-window procedure that updates the model inputs every 12 months are represented by the histograms. On the other hand, the performance scores of the models trained within the static period from 01 January 1997 to 31 December 2001 and forecasted over the remainder of the examination period (01 January 2002 to 31 December 2007) are represented by the trend lines.

CHART (A) Out-of-Sample Average IC Scores

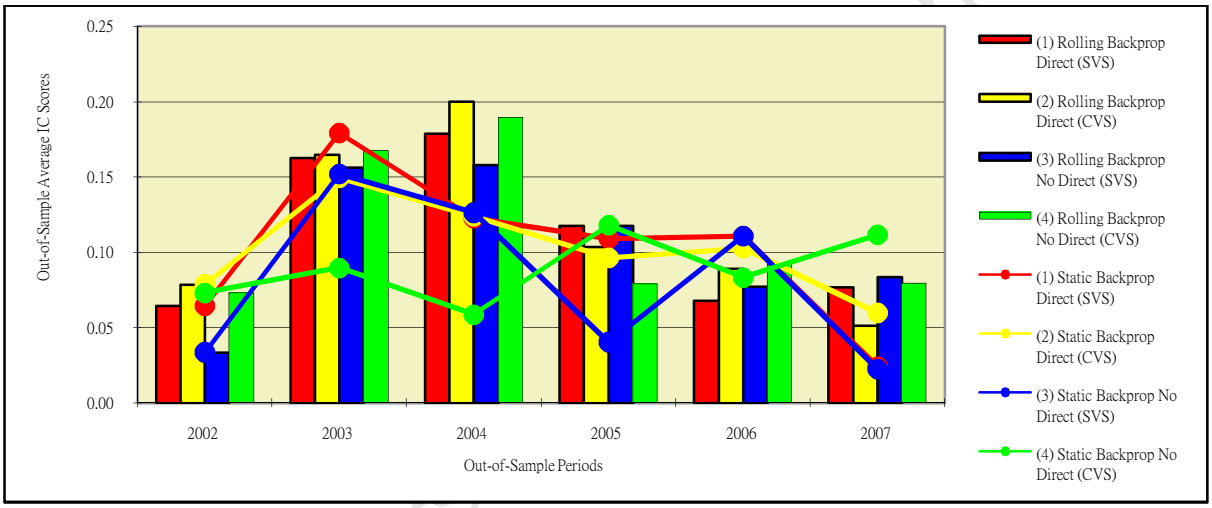


CHART (B) Out-of-Sample Grinold IR Scores

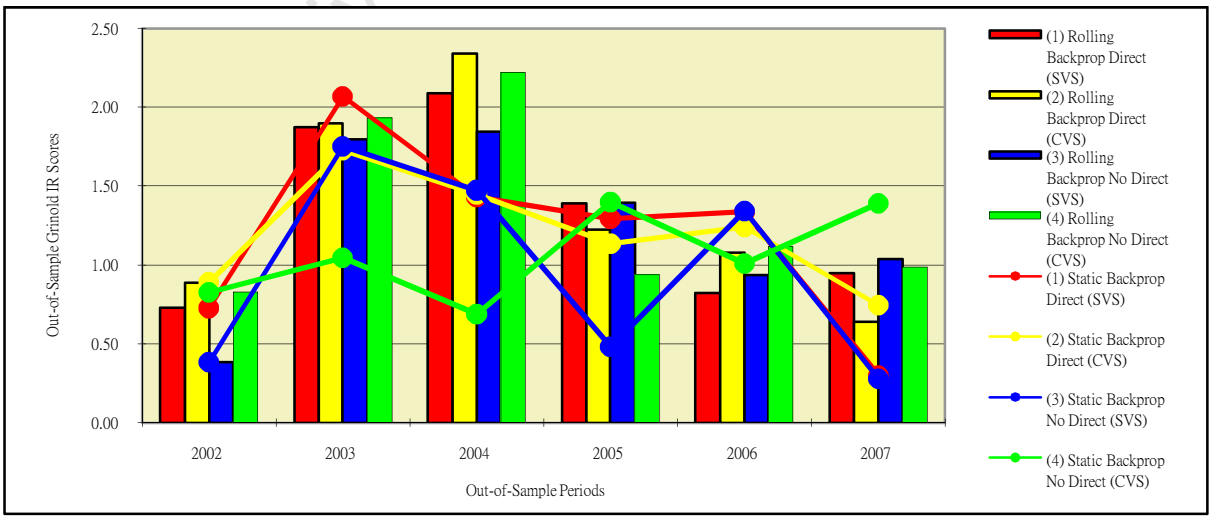


Figure 7.5 Periodic Out-of-Sample Performance Scores of the Backpropagation Models - Continued

CHART (C) Out-of-Sample Qian and Hua IR Scores

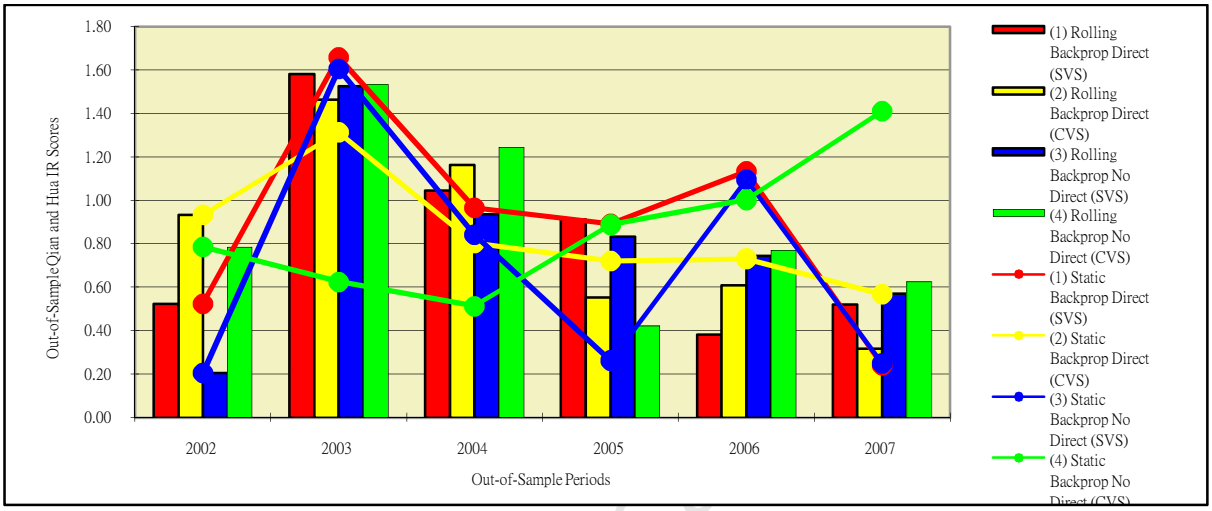


CHART (D) Out-of-Sample Slope *t*-Statistic Scores

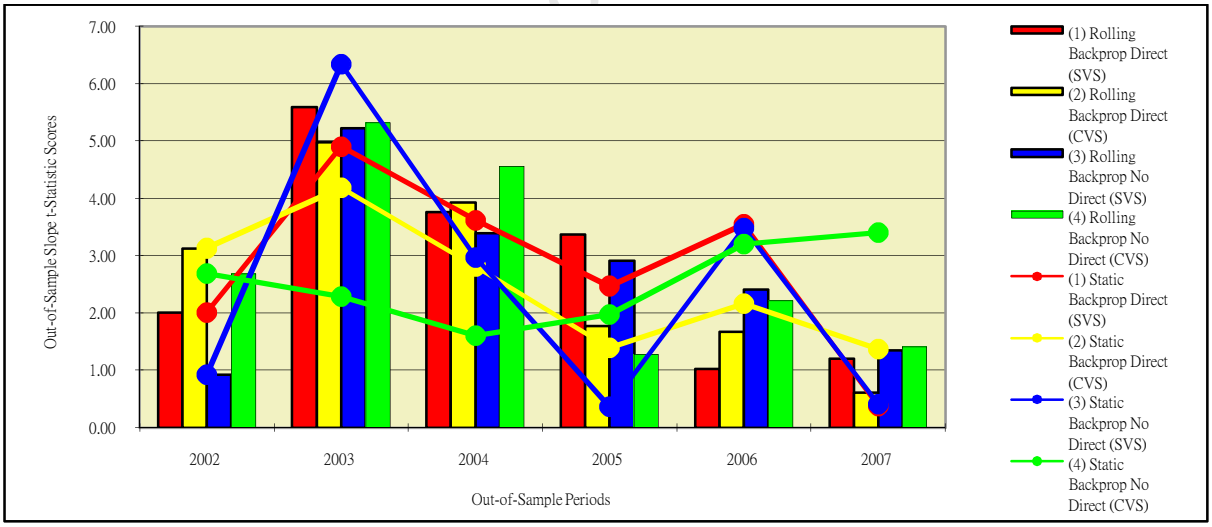


Table 7.4 demonstrates the consolidated in-sample and out-of-sample performance scores for the rolling backpropagation models constructed in this chapter. The consolidated in-sample and out-of-sample performance scores for each of the performance criteria are calculated as the periodic average scores over the six sub-periods.

The best in-sample average IC and QH IR scores are achieved by the backpropagation model that applies the cascade pre-selection technique with no direct connection from inputs to output in the network architecture (refer to model 4). This model also achieved the second best Grinold's IR, the best QH IR and the best slope *t*-statistic scores in the out-of-sample period. The other cascade pre-selection backpropagation model with direct connection from the input layer to the output layer (refer to model 2) has the second best in-sample average IC score and achieved the best average IC, the best Grinold's IR, the second best QH IR and the second best slope *t*-statistic in the out-of-sample period. These results indicate that backpropagation models that apply the cascade pre-selection technique are more robust in generalising pricing relationships than the models applying the standard variable selection technique.

Figure 7.6 plots the rolling backpropagation models in the diagram based on their respective consolidated average IC scores and standard deviation of ICs for the in-sample and out-of-sample periods. The filled labels represent in-sample scores while the corresponding out-of-sample scores are represented by empty labels. The slope of the line connecting the origin to the position of the model in the diagram is essentially the model's QH IR score. Examining the relative locations of the backpropagation models based on their in-sample performances reveals that the four models have more or less the same in-sample QH IR. Although the models applying the cascade pre-selection technique (refer to model 2 and model 4 in the diagram) achieve relatively higher in-sample average IC, their in-sample standard deviations of ICs are also relatively higher than the standard variable selection counterparts (refer to model 1 and model 3 in the diagram). However, while the cascade pre-selection model with the direct connection constraint (refer to model 2 in the diagram) achieves relatively higher out-of-sample average IC score, its forecasting accuracy is hampered by the relatively higher standard deviation of its forecasting accuracy. On the other hand, the cascade

pre-selection model with no direct connection constraint (refer to model 4 in the diagram) is able to maintain high forecasting accuracy with substantially lower standard deviation for its forecasting accuracy.

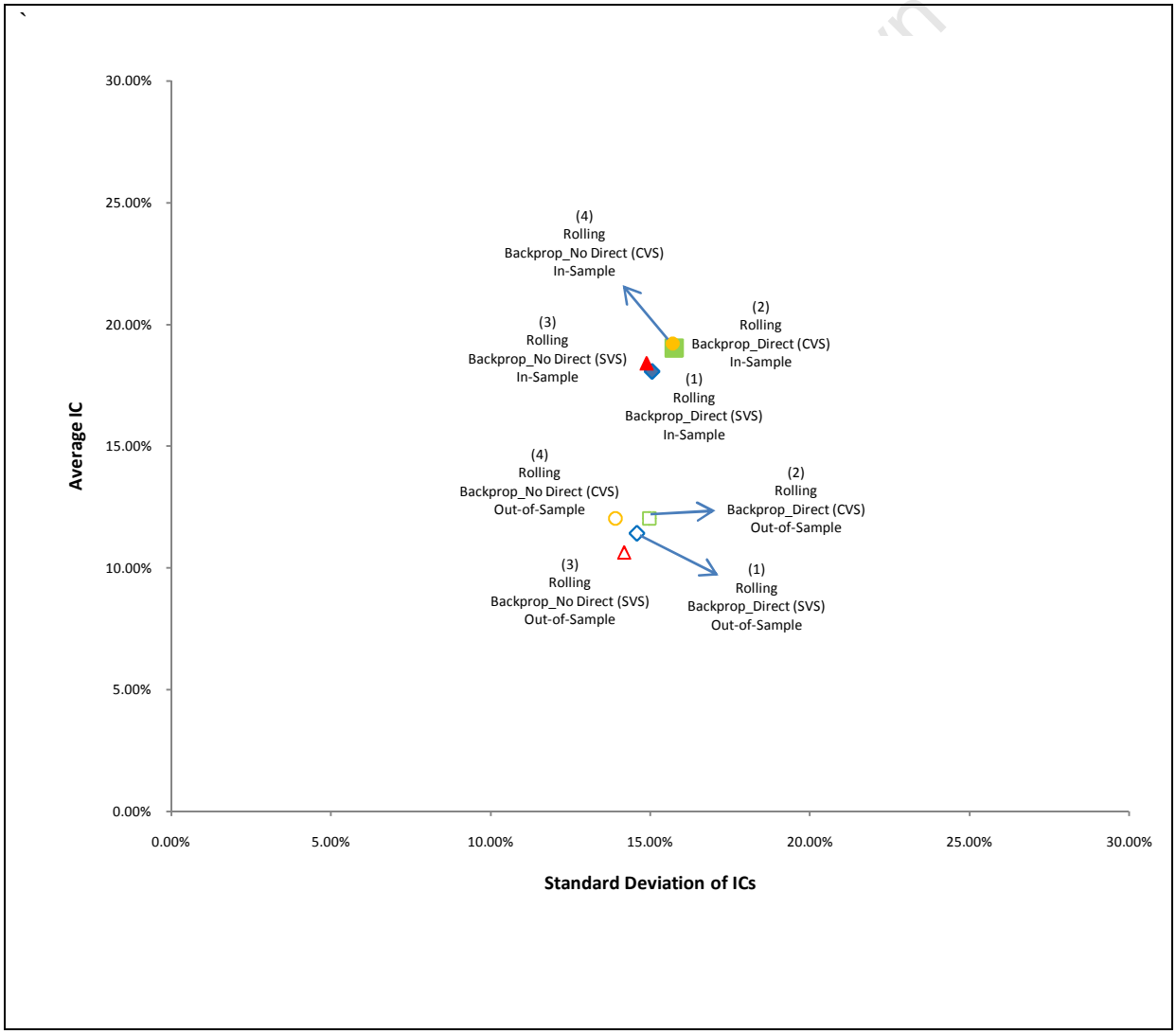
Table 7.4 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Backpropagation Models

This table contains the average periodic in-sample and the corresponding out-of-sample performance scores of the rolling backpropagation models developed in this chapter. The models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and output. The backpropagation models applying the cascade pre-selection technique are assigned with CVS. On the other hand, the models applying standard variable selection technique are assigned with SVS. The best out-of-sample score is highlighted in bold.

	(1) Rolling Backprop Direct (SVS)	(2) Rolling Backprop Direct (CVS)	(3) Rolling Backprop No Direct (SVS)	(4) Rolling Backprop No Direct (CVS)
Avg IC (In-Sample)	0.1806	0.1902	0.1840	0.1923
Avg IC (Out-Sample)	0.1144	**0.1203**	0.1062	0.1166
Grinold's IR (In-Sample)	2.0367	2.1496	2.0740	2.1678
Grinold's IR (Out-Sample)	1.3426	**1.4101**	1.2511	1.3687
Qian and Hua's IR (In-Sample)	1.1987	1.2070	1.2358	1.2249
Qian and Hua's IR (Out-Sample)	0.7839	0.8029	0.7491	**0.8375**
Slope t-Statistic (In-Sample)	8.4794	8.0442	8.9014	8.3493
Slope t-Statistic (Out-Sample)	6.1700	6.3250	6.2431	**6.7726**

Figure 7.6 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Backpropagation Models

Figure 7.6 plots the rolling backpropagation models in the diagram based on their respective consolidated average IC scores and standard deviation of ICs for the in-sample and out-of-sample periods. The slope of the line connecting the origin to the position of the model in the diagram is essentially the model's QH IR score. The filled labels represent in-sample scores while the corresponding out-of-sample scores are represented by empty labels. The four backpropagation models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and outputs. The backpropagation models applying the cascade pre-selection technique are assigned with CVS. On the other hand, the models applying standard variable selection technique are assigned with SVS.



Appendix C documents the performance of the backpropagation models that select inputs from the style attributes that exhibit greater than 1.0 (*t*-stats > 1.0) for the absolute values of their mean factor payoffs in each sub-period. The periodic variable selections, in- and out-of-sample scores and the consolidated results are presented in Table C.1, Figure C.1, Figure C.2, Table C.2 and Figure C.3 respectively.

Comparing the variables selected by the backpropagation models (*t*-stats > 1.0) in Table C.1 to the variables selected by the backpropagation models that choose inputs from the entire pool of all 38 attributes in Table 7.1 reveals that book value-to-price and cash flow-to-price remain the two most favourite attributes being selected by all models across all sub-periods. Although no significant differences between the periodic in-sample performance scores are identified (refer to the comparison between Figure 7.4 and Figure C.1), the consolidated in-sample scores for the backpropagation models (*t*-stats > 1.0) are concentrated in one area since they are built from a smaller set of variables. In addition, the periodic out-of-sample scores are drastically lower for these models (refer to Figure C.2) compared to the models that select variables from the entire pool of all 38 variables (refer to Figure 7.5) in the first (2002) and the last (2007) sub-periods. With the exception of the backpropagation model using the cascade pre-selection technique with direct connection, the other three models that select inputs from variables with greater than 1.0 for the absolute values of their mean factor payoffs have much lower out-of-sample QH IR compared to that of the models that select inputs from the entire pool of variables. Thus, it is concluded that narrowing down the initial pool of variables based on their respective significance of mean factor payoffs is an ineffective way of improving the forecasting power of the backpropagation models.

7.4 Conclusion

There are in total four rolling backpropagation models developed in this chapter, of which each model is unique in terms of its variable selection technique and network architecture. In the variable selection process, book value-to-price and cash flow-to-price are the most important variables being selected by all models across all training periods. In general, attributes from the value and momentum categories dominate the variables selected by all backpropagation models. As expected, the cascade pre-selection technique effectively reduces the number of variables selected by the models compared to the standard variable selection technique.

Over the out-of-sample period from 2002 to 2007, the out-of-sample scores of the backpropagation models are more or less in line with their in-sample scores in the first three years. During this period, the backpropagation models applying the cascade pre-selection technique outperform their standard variable selection counterparts with lower in-sample scores and substantially less number of model inputs. This finding suggests that the large number of inputs selected by the standard variable selection technique significantly overfit the backpropagation models, which hampers the robustness of the model performance in the out-of-sample period. On the other hand, the out-of-sample performance scores for the last 3 sub-periods from 2005 to 2007 are not consistent with the performance scores obtained from their corresponding in-sample periods. The models applying the cascade pre-selection technique are also less robust during this period.

Overall, this study finds that the variables selected by backpropagation models that apply the cascade pre-selection technique to be more robust in generalising pricing relationships than the models applying the standard variable selection technique. In addition, the comparisons of the out-of-sample performances between the moving-window (rolling) approach and the static approach reveals no evidence that the rolling models are more robust than their static counterparts. In addition, the backpropagation models that select inputs from the entire pool

of 38 attributes are found to be more robust than the alternative models that select inputs from only those variables with *t*-statistic of greater than 1.0 for their mean factor payoffs.

With regard to the choice of direct connection in the neural network architecture, the examination of the periodic performance scores finds no evidence in support of the Qi and Maddala (1998)'s argument that removing the direct connection constraint can effectively improve the model's forecasting power. However, this argument is reinforced when the consolidated performance results reveal that the cascade pre-selection model with no direct connection constraint is able to achieve relatively higher forecasting accuracy represented by the average IC score with substantially lower standard deviation of the forecasting accuracy compared to its direct connection counterparts. In conclusion, the rolling backpropagation model that applies the cascade pre-selection technique prior to its standard variable selection procedure with no direct connection constraint is the most robust expected return forecasting backpropagation model identified in this research.

8

Nonlinearities in the Cross-Section of Equity Returns: Extended Kalman Filter Artificial Neural Networks

8.1 Introduction

Financial data is often viewed as been noisy which impacts negatively on the generalisation ability of the artificial neural network (ANN) models. For this reason, it could be argued that perhaps a learning rule, other than backpropagation is more suitable for this noisy regression data. Thus a weight estimation procedure that specifically accounts for noise, such as the extended Kalman filter learning rule, requires consideration.

The objective of this chapter is to explore the existence of nonlinearity in the cross-section of equity returns on the JSE for the period 01 January 1997 to 31 December 2007. As in Chapter 7, ANN models are constructed and trained within the cascade learning architecture proposed by Fahlman and Lebiere (1990). A major problem with fixed-topology networks, highlighted in Chapter 7, is that the number of layers and neurons comprising each layer are determined prior to training the network. An advantage of employing the cascade-correlation algorithm is that the network topology evolves dynamically as hidden neurons are added to the network, one at a time. The cascade learning procedure adopted in Chapter 7 is unique in that it entails employing an embedded backpropagation (first-order stochastic gradient descent method) learning rule to aid the training process (weight adjustment). Adopting the methodology of Chapter 7, this chapter constructs and trains ANN models, but instead of employing the backpropagation rule, the extended Kalman filter (EKF) learning rule is employed. Backpropagation is a technique used to assign weights in a multilayer perceptron (MLP) network. An extended Kalman filter (second-order descent method), a nonlinear version of the popular Kalman filter proposed by Kalman (1960) is used to estimate nonlinear systems via linearisation about the current parameter estimates. Inspired by systems identification and nonlinear filtering, the extended Kalman filter can be used in ANN modeling to estimate the network weights by assuming that the weights are the states to be estimated.

The cascade-correlation network is combined with the extended Kalman filter training rule in the learning process. The extended Kalman filter is chosen as it is a fast converging alternative to the backpropagation algorithm and specifically adjusts for the presence of noise in financial data (Nechyba and Xu, 1997). Singhal and Wu (1989) were the first to test the extended Kalman filter as a training technique in neural networks. They find the extended Kalman filter to be superior to backpropagation for training multilayered networks. An alternate interpretation proposed by Ruck, Rogers, Kabrisky, Maybeck and Oxley (1992) is that backpropagation is simply a degenerate form of the extended Kalman filter.

The models developed in this chapter are all constructed within the cascade-correlation architecture and trained using an embedded extended Kalman filter learning algorithm. These models differ in terms of network architecture, variable selection technique and whether they are trained under the static or moving-window (rolling) procedure. The models are evaluated under the same criteria used in Chapter 7, namely, the average information coefficient (average IC), Grinold information ratio (Grinold IR), Qian and Hua information ratio (QH IR) and the slope *t*-statistic. The objective of specifying these models is to determine their relative robustness in equity return forecasting.

8.2 Descriptive Statistics and Methodology

Following the methodology used in Chapter 7, the dynamic moving-window (rolling) procedure whereby 60 months data are trained to forecast the immediate 12-month returns following the training period (60-for-12) is employed. Six overlapping in-sample periods comprising 60 months each are specified. There are six corresponding 12-month out-of-sample periods, respectively. As in Chapter 7, the first in-sample training set spans the period from 01 January 1997 to 31 December 2001. The corresponding out-of-sample period is from 01 January 2002 to 31 December 2002. The second training window spans the period from 01 January 1998 to 31 December 2002, to forecast the following 12 months, 01 January 2003 to 31 December 2003. This overlapping procedure is then repeated a further four times until all returns in the out-of-sample period are forecasted.

8.2.1 Model Design and Specifications

Four ANN models differing in terms of model architecture and in terms of variable selection are constructed according to specifications (1) to (4) below.

(1) *Kalman_Direct (SVS)* represents an augmented model, which, in addition to its nonlinear structure, has a direct connection from the input layer neurons to the output layer neuron. A standard variable selection *(SVS)* is employed where a genetic algorithm selects the best input variables for the network.

(2) *Kalman_Direct (CVS)* represents an augmented model (with direct connections from the input layer neurons to the output layer neurons, in addition to the hidden layer connections) with a cascade variable selection *(CVS)* technique. The mechanics of this pre-selection procedure before the standard variable selection is described in Chapter 7.

Kalman_Direct (SVS) and *Kalman_Direct (CVS)* have identical architectures. They differ with respect to variable selection, with a pre-selection (similar to the univariate test of Chapter 5) being applied to the latter model.

(3) *Kalman_No Direct (SVS)* represents a nonlinear model with the linear constraint that is present in (1) and (2) above, being removed. The model uses a genetic algorithm to select the best variables as inputs in the neural network. This standard variable selection *(SVS)* technique is consistent with (1).

(4) *Kalman_No Direct (CVS)* represents a nonlinear model where the linear constraint is removed from the model and as a result the direct connection from the input to the output layer is absent. The model employs a cascade variable selection *(CVS)* technique. The mechanics of this pre-selection procedure before the standard variable selection is described in Chapter 7.

Kalman_No Direct (SVS) and *Kalman_No Direct (CVS)* have identical architectural structures, but a more stringent variable selection technique being applied to *Kalman_No Direct (CVS)*. Comparison can also be drawn between *Kalman_Direct (SVS)* and *Kalman_No Direct (SVS)*. These models have identical variable selection techniques but different network architectures. The same comparison is to be drawn between *Kalman_Direct (CVS)* and *Kalman_No Direct (CVS)*.

As in Chapter 6 (linear factor models) and Chapter 7 (nonlinear factor models), the lagged values of the cross-sectional style attributes are used as the model inputs to forecast the expected returns of the sample shares. The in-sample set consists of a training set and a test set. This procedure is such that 70% of the in-sample data set is used as the training set while 30% is selected as the test set at regular intervals. The out-of-sample data serves as the validation set and is completely independent of the model building process. All 38 candidate attributes serve as potential inputs in the ANN models. The final set of input variables are selected via a genetic algorithm. Consistent with Chapter 6 and Chapter 7, the performance of the ANN models are evaluated according to the following evaluation criteria namely, the average information coefficient (average IC), Grinold information ratio (Grinold IR), Qian and Hua information ratio (QH IR) and the slope *t*-statistic.

8.2.2 Input Variable Selection

To select the optimal set of input variables for training the ANN models, a genetic algorithm is employed. A linear variable selection model is used to determine the fitness of the subset of input variables ("individuals") in Chapter 7. A necessary requirement for the extended Kalman filter rule is that the number of input variables presented to network for training is reduced. Thus the linear variable selection model, as employed in Chapter 7, is deemed inappropriate in this chapter. Instead, the genetic algorithm will employ a nonlinear variable selection model. A neural network (called a "secondary neural network") is employed as the variable selection model to determine the optimal subset of variables to be included in the 4 specified ANN models (which will be referred to as "primary neural network" for comparison purposes). A neural network selects the model inputs, while a sigmoid output function is used to shape the model.

In the case of variable selection the goal is to find a small subset of variables that are most significant for a regression. With the secondary neural network, evaluation starts with the encoding of the individuals (chromosomes) into neural networks whereby "1" indicates that a specific variable is used and "0" indicates that a variable is not used by the network. The training of the secondary neural network is consistent with the training process for the primary neural network as described in Chapter 7. All 38 variables are selected as potential input variables which may or may not be used as actual inputs in the neural network depending upon the information provided by the genetic algorithm. It is imperative that a reduced number of variables are selected and presented to the network, where the weights are trained via the embedded extended Kalman filter within the cascade learning architecture.

8.2.3 Neural Network Construction and Training

This chapter combines the cascade-correlation architecture, which dynamically adjusts as the network is trained and the extended Kalman filter learning algorithm, which can be viewed as a faster converging alternative to the backpropagation algorithm that effectively suppresses noises in behavioural problems (such as expected return determination).

Part (a) and Part (b) below provides a discussion of ANN construction and training.

(a) Cascade Learning

As described in Chapter 7, cascade learning begins with an input and an output layer. Initially there are no hidden layers, just the linear connections linking the input to the output layer. This is the case with Kalman_Direct (SVS) and Kalman_Direct (CVS) in which a direct connection from input to output layer exists. In the case of Kalman_No Direct (SVS) and Kalman_No Direct (CVS), in which there are no direct connections from input to output layer, a hidden neuron is installed. The best candidate hidden neuron is selected after all candidates have been trained using the extended Kalman filter learning rule embedded in the cascade architecture with different random initial weights. Once a new hidden neuron is installed, the hidden neuron input weights are frozen, while the weights to the output are retrained. The process is repeated until the algorithm succeeds in reducing the error sufficiently for the training set or the number of hidden layer reaches a specified maximum number. In this research a maximum of 30 hidden layers are specified, with a minimum increment of 1 and a maximum increment of 2 hidden neurons added at a time.

The extended Kalman learning rule is acceptable to regression type problems in which the number of inputs is not too large. Since it is not always possible or desirable to measure every variable that one wish to control, a Kalman filter is used, because it provides a means for inferring the missing information from indirect and noisy measurements (Grewal and Andrews, 2001). The filter combines all available measurement data in order to produce an estimate of the target variables in such a way that the error is statistically minimised. Part (b) below provides a discussion of the extended Kalman filter learning rule for training artificial neural networks.

(b) The Extended Kalman Filter in ANN Training

The extended Kalman filter learning rule is embedded within the cascade-correlation architecture to train the weights on the input-side and output-side of the candidate hidden neuron during cascade learning.

State estimation is a procedure in which the state of a process needs to be estimated, given the

presence of noisy and inaccurate measurements from the process. According to Orderud (1999: 1) "*state-space models actually consists of two separate models: the process model, which describes how the state propagates in time based on external influences, such as input and noise; and the measurement model, which describes how measurements z are taken from the process, typically simulating noisy and/or inaccurate measurements*". In ANN modeling, the network weights represent the state that need to be estimated in the training process. The extended Kalman filter is the nonlinear extension of the Kalman filter, which is an optimal estimator for linear systems. The EKF provides a convenient measure of prediction accuracy through the covariance matrix. The Kalman filter is obtained through the linearisation of the nonlinear state and observation equations. The strength of using an EKF is that it provides a covariance matrix that is indicative of the uncertainty in the prediction. Singhal and Wu (1989) were the first to put forward an argument in favour of the extended Kalman filter algorithm for training neural networks. Their argument is based on the fact that because a multilayered feed-forward neural network can be viewed as a static nonlinear dynamic system whose state is a vector containing all its synaptic weights, ANN training can thus be viewed as a state estimation problem for a stationary nonlinear system. Hence, if the neural network is formulated in terms of state-space concepts similar to those of a static nonlinear dynamic system, then the best conditional mean and covariance of the synaptic weight vector can be found by employing an extended Kalman filter. Thus, through the covariance matrix (which indicates prediction uncertainty), the extended Kalman filter serves as a confidence measure of prediction accuracy. The extended Kalman filter has two distinct phases: predict next state (before the measurements are taken) and measurement update (update state after the measurements are taken). Thus process and observation (measurement) models are specified in extended Kalman filtering. In state estimation form, the training of the neural network can be formulated as a filtering problem, assuming that the data are generated by the following noisy signal model (Haykin, 2001):

$$w_k = w_{k-1} + u_k \tag{8.1}$$

$$t_k = h(w_k, x_k) + v_k \tag{8.2}$$

Where:

$h(w_k, x_k)$	=	the value of the function is the ANN output;
t_k	=	is the known output target (actual) output vector;
v_k	=	is the measurement noise vector that is assumed to a Gaussian-distributed random variable with zero mean and constant covariance;
x_k	=	is the input vector;
w_k	=	the weights of the ANN model at k; and
u_k	=	is the process noise vector that is assumed to a Gaussian-distributed random variable with zero mean and constant covariance.

The following assumptions are made:

u_k is a white process noise with $E[u_j u_k^T]$ = $\delta_{jk} Q_k$ covariance matrix

v_k is a white measurement noise with $E[v_j v_k^T]$ = $\delta_{jk} R_k$ covariance matrix

$E[u_j v_k^T]$ = 0, for all j, k.

Relating artificial neural networks to state estimation, the state of the system is given by the network's weight parameter value w_k. Equation 8.1 specifies that the ideal neural network is characterised as a stationary process corrupted by process noise u_k. It represents the process (predict phase) model, which uses estimates from the previous step to produce and estimate of the current state. Equation 8.2 represents the network's desired response vector t_k as a nonlinear function of the input vector x_k and the weight parameter w_k (Haykin 2001). Observation of the weights (w_k), is made according to Equation 8.2, the measurement phase. The nonlinear function h thus relates the state to the measurement t_k.

State estimation using the EKF entails determining/estimating the weights $\hat{w}$ that minimises the network error. Thus, the EKF solution to finding $\hat{w}$ can be obtained by using the following recursion (Haykin, 2001; Lary and Mussa, 2004):

$$\hat{w}_k = \hat{w}_{k-1} + K_k(t_k - \hat{y}_k) \tag{8.3}$$

Where:

$\hat{w}$ = weights (the state vector) to be estimated;

$\hat{w}_{k-1}$ = previous step weight estimate;

$t_k - \hat{y}_k$ = vector contains the prediction errors (innovations); and

$\hat{y}_k$ = is the prediction (=$h(w_{k-1}, x_k)$).

K_k represents a weighting factor known as the Kalman gain matrix at step k. It is the gain or blending factor that minimises the error covariance, where P_k is defined as the estimate error covariance (conditional mean covariance matrix). P_k represents a measure of the estimated accuracy of the state estimate. The Kalman gain matrix can thus be represented as follows:

$$K_k = P_{k-1}H_k\,(R_k + H_k^T P_{k-1} H_k)^{-1} \tag{8.4}$$

Where:

P_{k-1} = the estimate error covariance at the previous iteration; and

R_k = the measurement noise covariance at iteration *k*.

Observing Equation 8.4, it is evident that as P_{k-1}, approaches zero, the Kalman gain weights the residual error/innovation less heavily, but as R_k approaches zero, the Kalman gain weights the residual error more heavily. P_k is thus represented as follows:

$$P_k = P_{k-1} - K_k H_k^T P_{k-1} + Q_k \tag{8.5}$$

Where:

H_k is the matrix of the partial derivatives h_k with respect to all elements of $\hat{w}_{k-1}$, that is:

$$H_k = \frac{[\partial h(\hat{w}_{k-1}, x_k)]}{\partial \hat{w}_{k-1}} \tag{8.6}$$

$Q_k =$ the process noise covariance, assumed to be zero

An important feature of the EKF is that the function of H_k (the Jacobian matrix), is to correctly propagate the relevant component of the measurement component.

8.3 Results

Table 8.1 illustrates the variable selection process of the ANN Kalman models. Letter "I" denotes that the attribute is chosen by the algorithm as model inputs. Letter "V" denotes that the attribute is eliminated in the standard variable selection stage. Letter "C" denotes that the attribute is eliminated in the cascade variable pre-selection stage before the standard variable selection stage.

The number of variables selected by the ANN Kalman models is substantially less than the number of variables selected by the ANN backpropagation models in the previous chapter. In addition, the ANN Kalman models that apply the cascade pre-selection technique do not necessarily have fewer variables than the models that apply standard variable selection technique.

Table 8.2 provides an indication as to the number of training periods in which each of the candidate attributes are selected by the standard variable selection procedure and the variable selection procedure that applies the cascade pre-selection technique respectively. As shown in Table 8.2, cash flow-to-price (CFTP) is once again being included by both variable selection procedures in all overlapping in-sample periods as is for the backpropagation models. Book value-to-price (BVTP) that enjoys frequent inclusions by the backpropagation models also remains a favourite model inputs for both variable selection procedures. Debt-to-book value (DEBTTBVE) that is included in three out of six training periods by the standard variable selection procedure is completely excluded by the cascade pre-selection models. On the other hand, 6-month return momentum (MOM6) and earnings yield (EY) that are frequently included by the cascade pre-selection technique are not considered important model inputs by the standard variable selection technique. In summary, while attributes from the value and momentum categories continue to dominate the model inputs selected by the cascade pre-selection approach, momentum attributes are considered less significant by the standard variable selection procedure.

Table 8.1 Periodic Variable Selections of the Kalman Models

This table documents the complete variable selection process of the Kalman models developed in this chapter for each of the six overlapping in-sample periods over the period from 1 January 1997 to 31 December 2007. The Kalman models are divided into models with or without the cascade pre-selection procedure in addition to the existing standard variable selection procedure. When a variable is included by the algorithm as a selected input in the neural network, it is assigned with a letter "I". By contrast, variables that failed to pass the standard variable selection procedure are assigned with a letter "V". For the models that pre-select variables using the cascade variable selection technique, variables eliminated in the pre-selection step are assigned with a letter "C" without going through the standard variable selection procedure.

In-Sample Period:	1/1/97 ~ 31/12/01		1/1/98 ~ 31/12/02		1/1/99 ~ 31/12/03		1/1/00 ~ 31/12/04		1/1/01 ~ 31/12/05		1/1/02 ~ 30/12/06	
	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection
BVTP	V	V	I	I	I	I	I	I	I	I	V	I
CFTCURRLIABS	V	C	V	V	V	V	V	C	V	C	V	C
CFTDEBT	V	V	V	C	V	V	V	C	V	V	V	V
CFTP	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**	**I**
CURRENTRATIO	V	C	V	V	V	V	V	V	V	C	V	C
DEBTTBVE	V	V	I	V	V	V	I	C	V	V	I	V
DEBTTMVE	V	V	V	V	V	V	V	C	V	V	V	V
DY	V	V	V	V	I	V	V	V	V	C	V	V
EARNREV	V	V	V	V	V	V	V	V	V	V	V	V
EG1	V	C	V	V	V	V	V	V	I	I	V	V
EY	V	I	V	V	V	I	V	C	V	I	V	C
FOREY1	V	C	V	C	V	V	V	C	V	I	V	V
FOREY2	V	C	V	V	V	V	V	V	V	V	V	V
G12MCPS	V	V	V	V	V	C	I	V	V	V	V	C
G12MDPS	I	I	V	C	V	I	V	C	V	V	V	V
G12MEPS	V	C	V	C	V	C	I	C	I	V	V	C
G12MGPMARGIN	V	C	V	V	V	C	V	V	V	C	V	V
G12MNPMARGIN	V	V	V	C	V	V	I	V	I	I	V	C
G12MSALES	V	C	V	C	V	V	V	V	V	C	V	C
G24MEPS	V	V	V	I	V	C	I	C	V	C	V	V
GPMARGIN	V	V	V	V	V	V	V	C	V	C	V	I
GROWTH	V	C	V	V	V	V	V	C	V	C	V	V
ICBT	V	C	V	V	V	V	V	C	V	C	V	V
LAGLPRICE	V	V	I	V	V	V	V	C	I	V	V	I
LPRICE	V	V	V	V	V	C	V	C	V	V	I	V
LSIZE	V	C	V	C	V	C	V	C	V	I	V	V
MOM1	V	I	V	C	V	V	V	V	V	C	V	V
MOM12	V	C	I	V	V	V	V	V	V	V	V	I
MOM12-1	V	C	V	V	I	V	V	V	V	I	V	V
MOM24	V	I	V	V	V	V	I	I	V	C	V	V
MOM3	V	V	V	V	V	C	V	C	V	C	V	C
MOM6	V	I	V	C	V	I	V	C	V	I	I	I
NPMARGIN	V	V	V	V	I	C	V	C	V	V	V	I
PAYOUT	I	I	V	I	V	C	V	C	V	V	V	C
ROA	V	V	V	C	V	V	V	I	I	V	V	I
ROE	V	C	V	I	V	V	V	V	V	C	V	C
SALESTP	V	V	V	V	V	V	V	C	V	V	V	C
TATURNOVR	V	C	V	V	I	V	V	V	I	C	V	V
Total:	**3**	**7**	**5**	**5**	**6**	**5**	**8**	**4**	**8**	**9**	**4**	**8**

Table 8.2 Frequency of Inclusion for the Candidate Style Attributes by the Standard and Cascade Pre-Selection Techniques in the Kalman Models

This table documents the number of in-sample periods in which each of the candidate style attributes is included by the Kalman models with and without the cascade pre-selection technique respectively. The variables are arranged in descending order according to their frequency of inclusions in the models.

Standard Variable Selection Technique		**Cascade Pre-Selection Technique**	
Firm-Specific Attributes	**Frequency of Inclusion (Out of 6)**	**Firm-Specific Attributes**	**Frequency of Inclusion (Out of 6)**
CFTP	***6***	***CFTP***	***6***
BVTP	4	BVTP	5
DEBTTBVE	3	MOM6	4
G12MEPS	2	EY	3
G12MNPMARGIN	2	G12MDPS	2
LAGLPRICE	2	MOM24	2
TATURNOVER	2	PAYOUT	2
DY	1	ROA	2
EG1	1	EG1	1
G12MCPS	1	FOREY1	1
G12MDPS	1	G12MNPMARGIN	1
G24MEPS	1	G24MEPS	1
LPRICE	1	GPMARGIN	1
MOM12	1	LAGLPRICE	1
MOM12-1	1	LSIZE	1
MOM24	1	MOM1	1
MOM6	1	MOM12	1
NPMARGIN	1	MOM12-1	1
PAYOUT	1	NPMARGIN	1
ROA	1	ROE	1
CFTCURRLIABS	0	CFTCURRLIABS	0
CFTDEBT	0	CFTDEBT	0
CURRENTRATIO	0	CURRENTRATIO	0
DEBTTMVE	0	DEBTTBVE	0
EARNREV	0	DEBTTMVE	0
EY	0	DY	0
FOREY1	0	EARNREV	0
FOREY2	0	FOREY2	0
G12MGPMARGIN	0	G12MCPS	0
G12MSALES	0	G12MEPS	0
GPMARGIN	0	G12MGPMARGIN	0
GROWTH	0	G12MSALES	0
ICBT	0	GROWTH	0
LSIZE	0	ICBT	0
MOM1	0	LPRICE	0
MOM3	0	MOM3	0
ROE	0	SALESTP	0
SALESTP	0	TATURNOVER	0

The periodic in-sample and out-of-sample performance scores for each of the ANN Kalman models trained under each of the six sub-periods are demonstrated in Table 8.3. Large discrepancies between the in-sample and out-of-sample performance scores are documented over the six sub-periods. In addition, examining the periodic performances of the ANN Kalman models gives no indication as to whether removing the direct connection from the inputs to the output will enhance the performances of the models.

Cross-examining the effectiveness of the cascade pre-selection technique to the standard variable selection technique reveals that the ANN Kalman models that apply the cascade pre-selection technique (refer to model 2 and model 4 in the table) significantly underperform their standard variable selection counterparts (refer to model 1 and model 3) in numerous occasions in the in-sample and out-of-sample periods. This finding contradicts to the results obtained from the comparison of the performance scores between the two alternative variable selection techniques for the ANN backpropagation models.

Table 8.3 Periodic Performance Scores of the Kalman Models

The performance scores of the ANN Kalman models for the six overlapping training periods and their corresponding out-of-sample performances are displayed in PANEL (A) through PANEL (F) respectively. The ANN Kalman models using the cascade pre-selection technique are assigned with CVS. On the other hand, the models applying standard variable selection technique are assigned with SVS.

PANEL (A) In-Sample: 1 Jan 1997 to 31 Dec 2001; Out-of-Sample: 1 Jan 2002 to 31 Dec 2002

	(1) Kalman_ Direct (SVS)	(2) Kalman_ Direct (CVS)	(3) Kalman_ No Direct (SVS)	(4) Kalman_ No Direct (CVS)
Avg IC Score (In-Sample)	0.1128	0.1111	0.1177	0.1324
Avg IC Score (Out-Sample)	**0.1268**	0.0916	0.0947	0.0351
Grinold's IR Score (In-Sample)	1.2330	1.2093	1.2849	1.4399
Grinold's IR (Out-Sample)	**1.4300**	1.0339	1.0674	0.4018
Qian and Hua's IR Score (In-Sample)	0.6215	0.6735	0.6729	0.8512
Qian and Hua's IR (Out-Sample)	**0.9245**	0.6285	0.6413	0.2577
Slope t-Statistic Score (In-Sample)	4.1093	4.3821	4.4081	5.9284
Slope t-Statistic (Out-Sample)	**2.9998**	2.1058	2.1546	1.1148

PANEL (B) In-Sample: 1 Jan 1998 to 31 Dec 2002; Out-of-Sample: 1 Jan 2003 to 31 Dec 2003

	(1) Kalman_ Direct (SVS	(2) Kalman_ Direct (CVS)	(3) Kalman_ No Direct (SVS)	(4) Kalman_ No Direct (CVS
Avg IC Score (In-Sample)	0.1323	0.1224	0.1781	0.1207
Avg IC Score (Out-Sample)	**0.1994**	0.1558	0.1933	0.1788
Grinold's IR Score (In-Sample)	1.4534	1.3443	1.9615	1.3282
Grinold's IR (Out-Sample)	**2.2972**	1.7967	2.2224	2.0599
Qian and Hua's IR Score (In-Sample)	0.8737	0.7021	1.2391	0.6840
Qian and Hua's IR (Out-Sample)	1.8488	1.8166	1.6037	**1.9391**
Slope t-Statistic Score (In-Sample)	5.6319	4.6811	8.1944	4.5195
Slope t-Statistic (Out-Sample)	**5.7604**	5.2843	5.4288	5.7056

PANEL (C) In-Sample: 1 Jan 1999 to 31 Dec 2003; Out-of-Sample: 1 Jan 2004 to 31 Dec 2004

	(1) Kalman_ Direct (SVS	(2) Kalman_ Direct (CVS)	(3) Kalman_ No Direct (SVS)	(4) Kalman_ No Direct (CVS)
Avg IC Score (In-Sample)	0.1759	0.1644	0.1738	0.1641
Avg IC Score (Out-Sample)	0.2099	0.2123	**0.2125**	0.2083
Grinold's IR Score (In-Sample)	1.9595	1.8299	1.9358	1.8267
Grinold's IR (Out-Sample)	2.4542	**2.4843**	2.4839	2.4369
Qian and Hua's IR Score (In-Sample)	1.1998	1.1024	1.1647	1.1040
Qian and Hua's IR (Out-Sample)	1.1365	1.3917	1.2026	**1.4348**
Slope t-Statistic Score (In-Sample)	8.3401	7.7036	8.1597	7.7619
Slope t-Statistic (Out-Sample)	3.9195	4.5469	4.1461	**4.6915**

Table 8.3 Periodic Performance Scores of the Kalman Models - Continued

PANEL (D) In-Sample: 1 Jan 2000 to 31 Dec 2004; Out-of-Sample: 1 Jan 2005 to 31 Dec 2005	(1) Kalman_ Direct (SVS)	(2) Kalman_ Direct (CVS)	(3) Kalman_ No Direct (SVS)	(4) Kalman_ No Direct (CVS)
Avg IC Score (In-Sample)	0.1956	0.1981	0.1978	0.2067
Avg IC Score (Out-Sample)	0.0911	0.0823	**0.0916**	0.1069
Grinold's IR Score (In-Sample)	2.2217	2.2479	2.2463	2.3481
Grinold's IR (Out-Sample)	1.0784	0.9754	1.0849	**1.2666**
Qian and Hua's IR Score (In-Sample)	1.3255	1.3235	1.3011	1.4682
Qian and Hua's IR (Out-Sample)	0.5264	0.4484	0.6020	**0.6267**
Slope t-Statistic Score (In-Sample)	9.7886	9.7389	9.6414	10.7977
Slope t-Statistic (Out-Sample)	1.5713	1.2767	1.8954	**1.9656**

PANEL (E) In-Sample: 1 Jan 2001 to 31 Dec 2005; Out-of-Sample: 1 Jan 2006 to 31 Dec 2006	(1) Kalman_ Direct (SVS)	(2) Kalman_ Direct (CVS)	(3) Kalman_ No Direct (SVS)	(4) Kalman_ No Direct (CVS)
Avg IC Score (In-Sample)	0.2030	0.2026	0.2093	0.1976
Avg IC Score (Out-Sample)	0.0820	0.1082	0.0796	**0.1048**
Grinold's IR Score (In-Sample)	2.3289	2.3260	2.4044	2.2693
Grinold's IR (Out-Sample)	0.9937	**1.3092**	0.9631	1.2679
Qian and Hua's IR Score (In-Sample)	1.3116	1.2586	1.3574	1.1806
Qian and Hua's IR (Out-Sample)	**0.9851**	0.9707	0.9035	0.8228
Slope t-Statistic Score (In-Sample)	9.1985	9.2277	9.6337	8.7089
Slope t-Statistic (Out-Sample)	**2.8563**	2.8042	2.6710	2.3202

PANEL (F) In-Sample: 1 Jan 2002 to 31 Dec 2006; Out-of-Sample: 1 Jan 2007 to 31 Dec 2007	(1) Kalman_ Direct (SVS)	(2) Kalman_ Direct (CVS)	(3) Kalman_ No Direct (SVS)	(4) Kalman_ No Direct (CVS)
Avg IC Score (In-Sample)	0.1451	0.1769	0.1726	0.1526
Avg IC Score (Out-Sample)	**0.1179**	0.0800	0.1179	0.0553
Grinold's IR Score (In-Sample)	1.6867	2.0581	2.0073	1.7704
Grinold's IR (Out-Sample)	**1.4660**	0.9924	1.4658	0.6890
Qian and Hua's IR Score (In-Sample)	1.0035	1.3353	1.4572	0.9276
Qian and Hua's IR (Out-Sample)	1.3200	0.5992	**1.4294**	0.3934
Slope t-Statistic Score (In-Sample)	7.3631	9.1298	10.3417	6.5437
Slope t-Statistic (Out-Sample)	4.1183	1.4696	**4.1222**	0.7507

The periodic in-sample scores achieved by the four ANN Kalman models for each of the performance measurement criteria are illustrated in the form of histograms in Figure 8.1. The in-sample performance scores of the ANN Kalman models improve over time as old data are replaced by newly arrived data in the rolling process. Cross-examining the in-sample scores of the four ANN Kalman models reveals that the ANN Kalman model that applies the standard variable selection technique with no direct connection between its inputs and output (refer to model 3 in the diagram) is ranked as the top two achievers across all performance measurement criteria over the six overlapping in-sample periods.

Figure 8.2 demonstrates the out-of-sample performance scores of the four rolling ANN Kalman models from 1 January 2002 to 31 December 2007 in the form of histograms. The out-of-sample performance scores of the corresponding static ANN Kalman models that are trained within a fixed 5-year period from 01 January 1997 to 31 December 2001 to forecast the entire out-of-sample returns are represented by trend lines. Without updating the variables and the networks periodically, the static ANN Kalman models with standard variable selection (refer to model 1 and model 3 represented by the red and blue trend lines) lost their predictive power completely and turn into the negative territory in 2007. By contrast, the static ANN Kalman models that apply the cascade pre-selection technique (refer to model 2 and model 4 represented by the yellow and green trend lines) maintain their robustness throughout the out-of-sample period.

On the other hand, the out-of-sample scores of the rolling ANN Kalman models do not match their corresponding in-sample scores. The relatively poorer in-sample training in the second and the third sub-periods produce excellent out-of-sample scores in 2003 and 2004, while the excellent in-sample training in the fourth and fifth sub-periods indeed produce relatively lower out-of-sample performances in 2005 and 2006.

Figure 8.1 Periodic In-Sample Performance Scores of the Kalman Models Trained under the Dynamic Moving-Window Procedure

CHART (A) to CHART (D) illustrate the periodic in-sample performance scores of the ANN Kalman models under each of the performance measurement criteria. The performance scores of the Kalman models in each chart are represented by the red, yellow, blue and green bars in the form of histograms.

CHART (A) In-Sample Average IC Scores

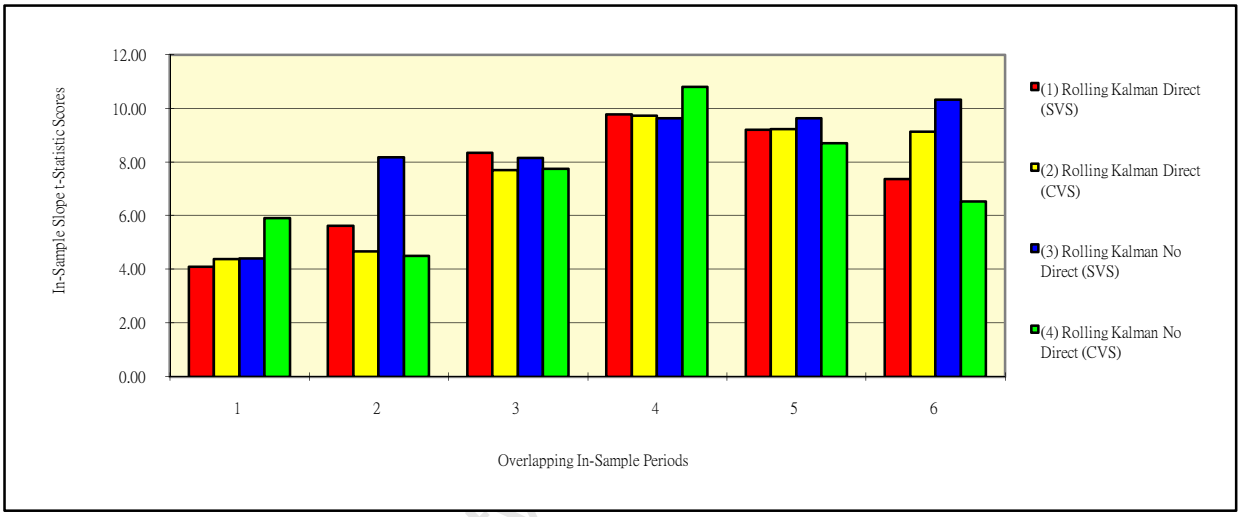


CHART (B) In-Sample Grinold IR Scores

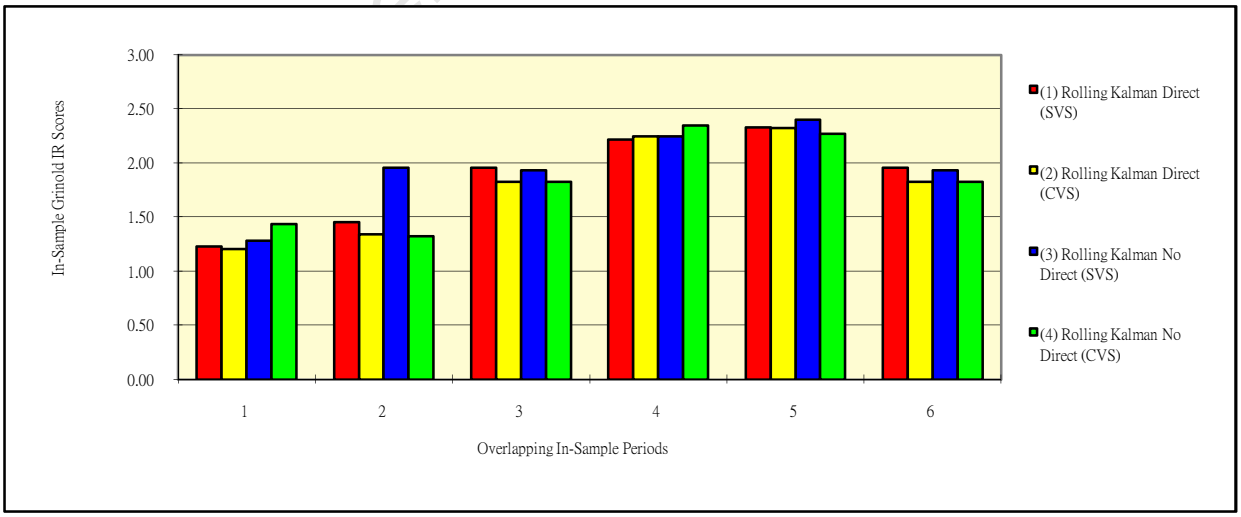


Figure 8.1 Periodic In-Sample Performance Scores of the Kalman Models Trained under the Dynamic Moving-Window Procedure - Continued

CHART (C) In-Sample Qian and Hua IR Scores

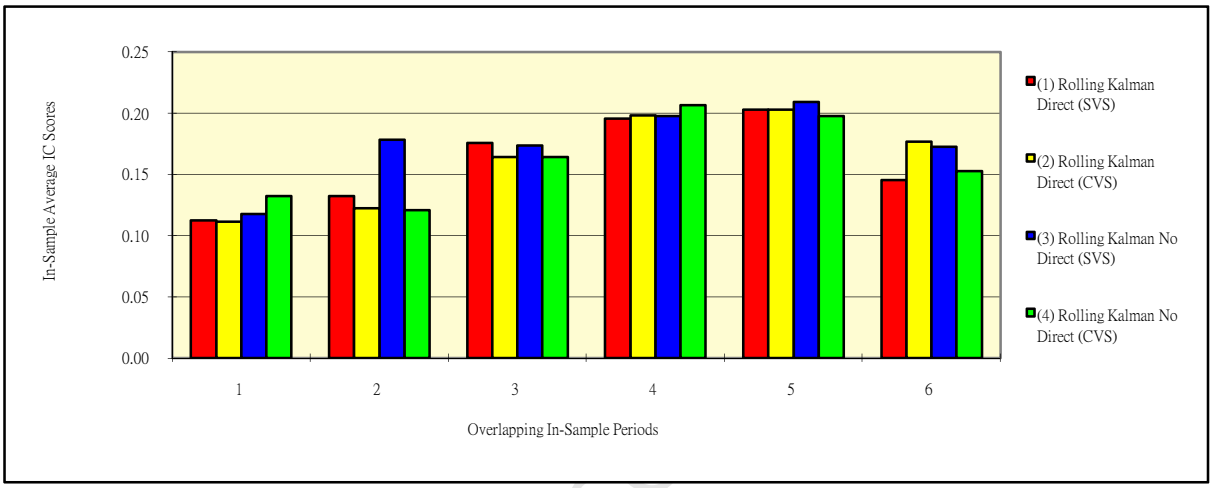


CHART (D) In-Sample Slope *t*-Statistic Scores

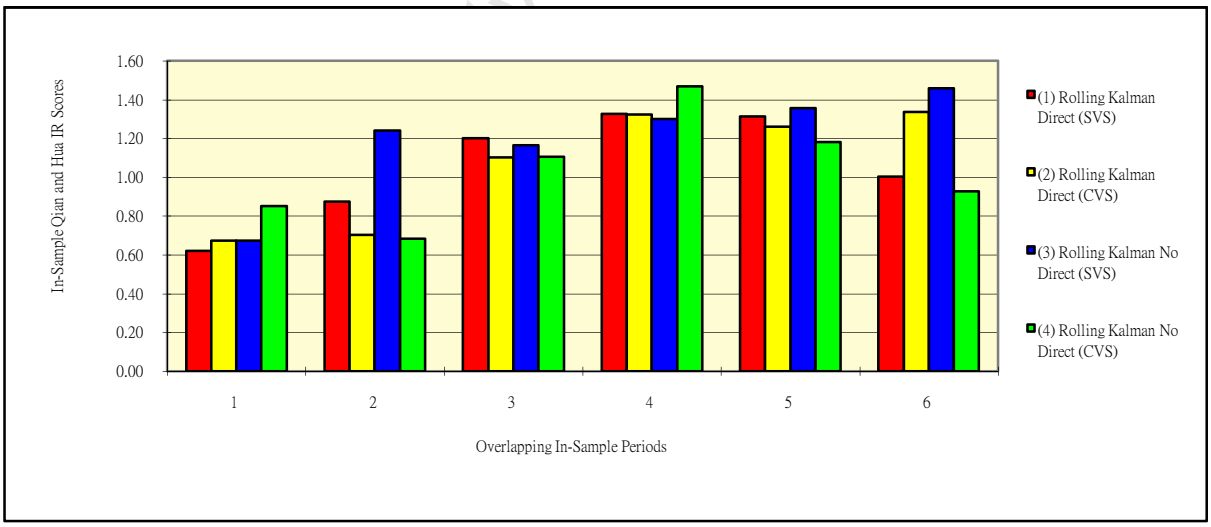


Figure 8.2 Periodic Out-of-Sample Performance Scores of the Kalman Models

CHART (A) to CHART (D) illustrate the periodic out-of-sample performance scores of the ANN Kalman models under each of the performance measurement criteria. The performance scores of the models constructed under the moving-window procedure that updates the model inputs every 12 months are represented by the histograms. On the other hand, the performance scores of the models trained within the static period from 01January 1997 to 31 December 2001 are represented by the trend lines. The ANN models in each chart are highlighted in red, yellow, blue and green respectively to enhance the direct comparisons of the respective performance scores of the pre-specified models.

CHART (A) Out-of-Sample Average IC Scores

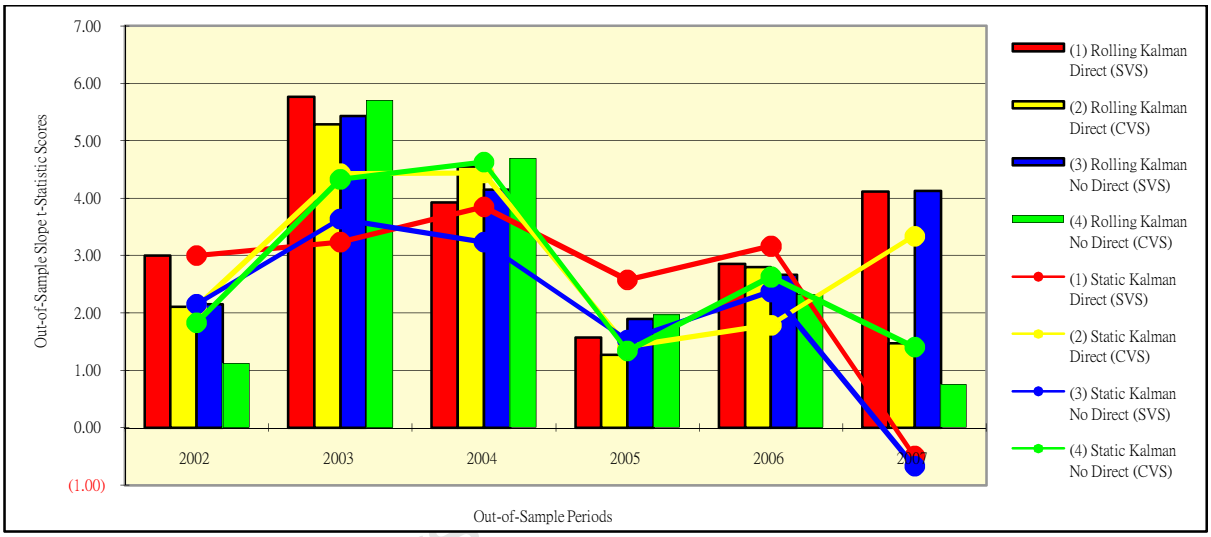


CHART (B) Out-of-Sample Grinold IR Scores

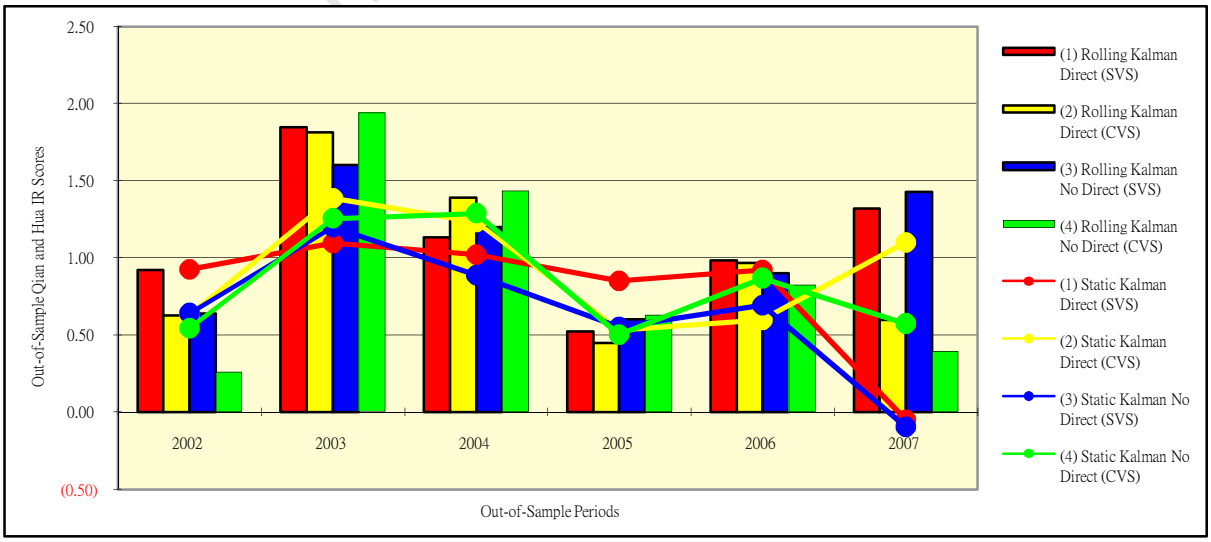


Figure 8.2 Periodic Out-of-Sample Performance Scores of the Kalman Models - Continued

CHART (C) Out-of-Sample Qian and Hua IR Scores

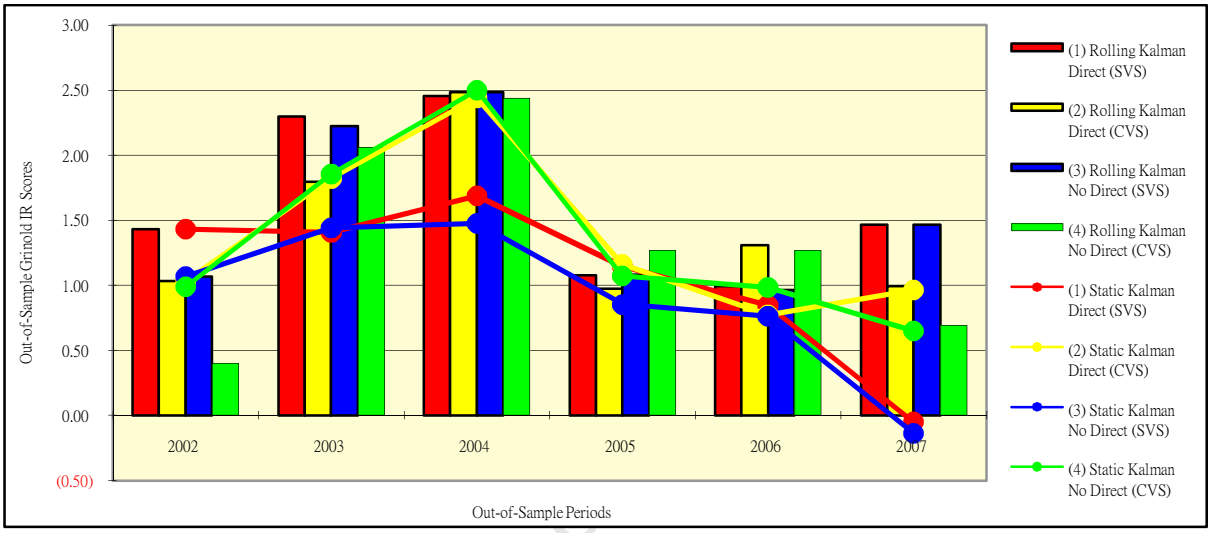


CHART (D) Out-of-Sample Slope *t*-Statistic Scores

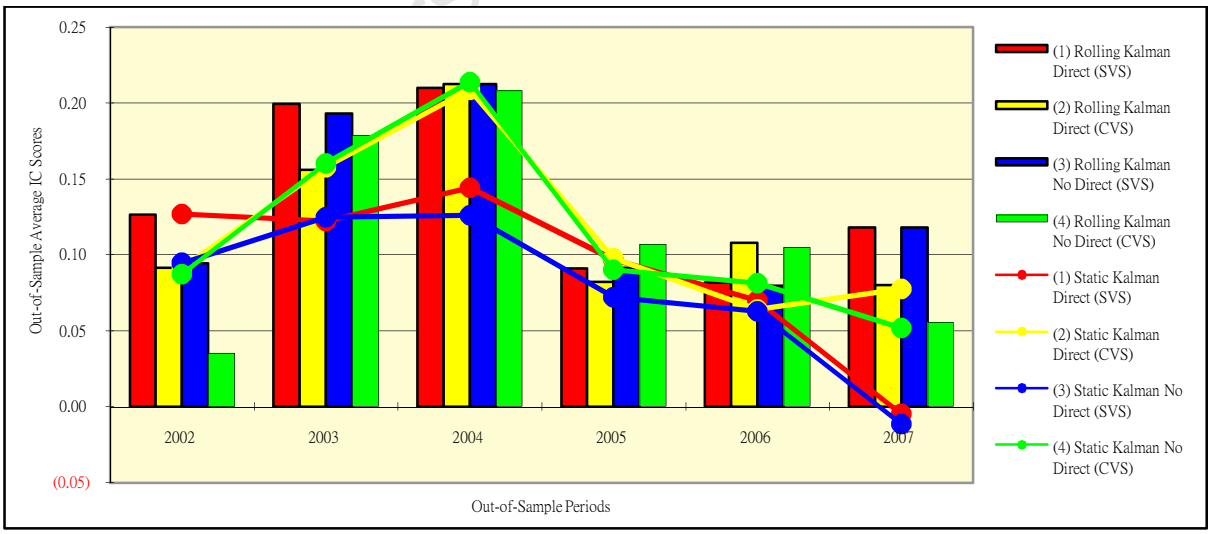


Table 8.4 demonstrates the consolidated in-sample and out-of-sample performance scores for the rolling ANN Kalman models developed in this chapter. The consolidated in-sample and out-of-sample performance scores for each of the performance criteria are calculated as the periodic average scores from the six sub-periods.

As shown in Table 8.4, the ANN Kalman model that applies the standard variable technique with no direct connection (refer to model 3 in Table 8.4) achieves the highest consolidated in-sample scores for all performance measurement criteria. However, its consolidated out-of-sample scores are outperformed by its direct connection counterpart (refer to model 1 in Table 8.4) for all performance measurement criteria. These two Kalman models that apply the standard variable selection technique turn out to be the top performers in the out-of-sample period compared to their cascade pre-selection counterparts (refer to model 2 and model 4 in Table 8.4).

Figure 8.3 plots the ANN Kalman models based on their average IC and standard deviation of ICs over the in-sample and out-of-sample periods respectively. The standard variable selection Kalman model with no direct connection (refer to model 3 in the diagram) achieves the highest average IC with significantly lower standard deviation of ICs compared to the other three ANN Kalman models over the in-sample period. Over the out-of-sample period, both of the ANN Kalman models that apply the standard variable selection technique (refer to model 1 and model 3) are competitive of each other and outperform their cascade pre-selection counterparts in terms of the QH IR scores. The model that has the lowest QH IR score in the out-of-sample period is the ANN Kalman model that applies the cascade pre-selection technique with direct connection in its network (refer to model 2 in the diagram).

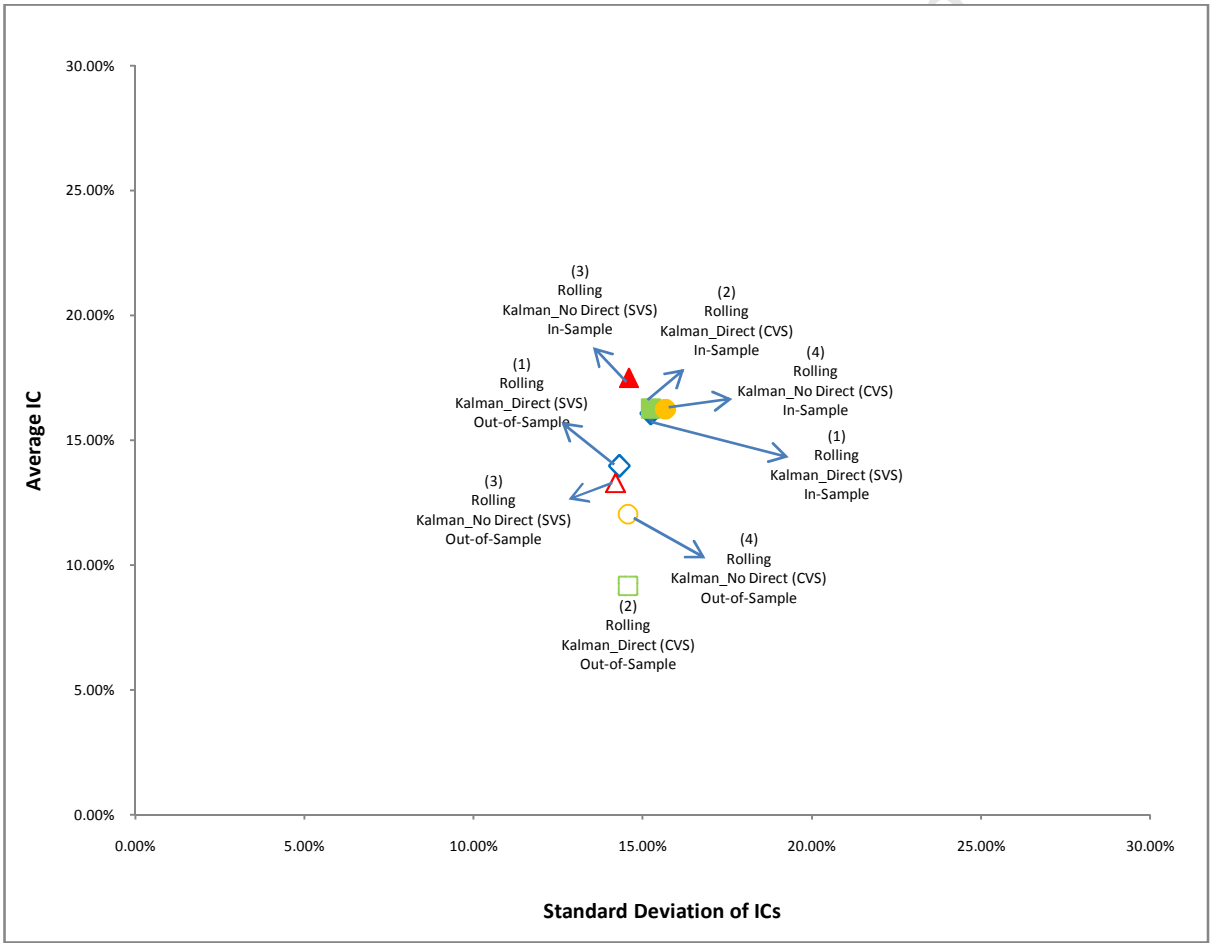
Table 8.4 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Kalman Models

This table contains the average periodic in-sample and the corresponding out-of-sample performance scores of the moving-window Kalman models developed in this chapter. The models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and output. The Kalman models applying the cascade variable selection technique are assigned with CVS. On the other hand, the models applying the standard variable selection technique are assigned with SVS.

	(1) Rolling Kalman Direct (SVS)	(2) Rolling Kalman Direct (CVS)	(3) Rolling Kalman No Direct (SVS)	(4) Rolling Kalman No Direct (CVS)
Avg IC (In-Sample)	0.1608	0.1626	0.1749	0.1624
Avg IC (Out-Sample)	**0.1397**	0.0916	0.1329	0.1203
Grinold's IR (In-Sample)	1.8139	1.8359	1.9734	1.8305
Grinold's IR (Out-Sample)	**1.6340**	1.0339	1.5554	1.4141
Qian and Hua's IR (In-Sample)	1.0560	1.0659	1.1998	1.0360
Qian and Hua's IR (Out-Sample)	**0.9762**	0.6285	0.9357	0.8261
Slope t-Statistic (In-Sample)	7.4053	7.4772	8.3965	7.3767
Slope t-Statistic (Out-Sample)	**7.6379**	2.1058	7.242	6.4101

Figure 8.3 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Kalman Models

Figure 8.3 plots the rolling ANN Kalman models in the diagram based on their respective consolidated average IC scores and standard deviation of ICs for the in-sample and out-of-sample periods. The slope of the line connecting the origin to the position of the model in the diagram is essentially the model's QH IR score. The filled labels represent in-sample scores while the corresponding out-of-sample scores are represented by empty labels. The four ANN Kalman models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and output. The ANN Kalman models applying the cascade variable selection are assigned with CVS. On the other hand, the models applying the standard variable selection technique are assigned with SVS.



Appendix D documents the performances of the Kalman models that select inputs from the style attributes that exhibit greater than 1.0 (*t*-stats > 1.0) for the absolute values of their mean factor payoffs in each sub-period. The periodic variable selections, in- and out-of-sample scores and the consolidated results are presented in Table D.1, Figure D.1, Figure D.2, Table D.2 and Figure D.3 respectively.

Comparing the variable selected by the Kalman models (*t*-stats > 1.0) in Table C.1 to the variables selected by the Kalman models that choose inputs from the entire pool of all 38 attributes in Table 8.2 reveals that book value-to-price and cash flow-to-price remain the two most favourite attributes being selected by all the models across most sub-periods. In addition, the total number of inputs chosen by models under both approaches remains low. No significant differences between the periodic in-sample performance scores are identified (refer to the comparison between Figure 8.1 and Figure D.1). Similar to the observations in the consolidated in-sample scores of the backpropagation models, the consolidated in-sample scores for the Kalman models (*t*-stats > 1.0) are concentrated in one area due to the smaller set of variables that are available. The periodic out-of-sample scores are compatible for the Kalman models that select variables from different pools of attributes with the exception of their slope *t*-statistic scores, in which the Kalman models with a smaller set of input selection underperform their counterpart models significantly (refer to the comparison of the out-of-sample scores between Table 8.4 and Table D.2).

Overall, mild underperformances of the models that choose inputs from variables with greater than 1.0 for the absolute values of their mean payoffs are documented. Thus, it is concluded that narrowing down the initial pool of variables based on their respective significance of mean factor payoffs does not improve the forecasting power of the Kalman models.

8.4 Conclusion

This chapter constructs four moving-window (rolling) ANN Kalman models to forecast the sample share returns from 01 January 2002 to 31 December 2007. In terms of the variable selection process, the number of variables selected by the ANN Kalman models is substantially less than the number of variables selected by the ANN backpropagation models examined in the previous chapter. In addition, the cascade pre-selection technique does not result in fewer variables selected for the ANN Kalman models compared to the number of variables selected by the standard variable selection technique.

Cash flow-to-price (CFTP) and book-value-to-price (BVTP) remain favourite choices of the ANN Kalman models as they are for the ANN backpropagation models. While the 6-month return momentum (MOM6) remain an important variable under the cascade pre-selection procedure, the attributes from the momentum category are considered less important by the ANN Kalman models applying the standard variable selection technique.

While the static ANN Kalman models applying the cascade pre-selection technique maintain their robustness over the entire out-of-sample period from 2002 to 2007, the performance scores of the static ANN Kalman models that adopt standard variable selection technique turn into negative territory in 2007. With regard to the rolling ANN Kalman models, their out-of-sample scores do not match their corresponding in-sample scores. More specifically, the relatively poorer in-sample training in the second and the third sub-periods produce excellent out-of-sample scores in 2003 and 2004, while the excellent in-sample training in the fourth and fifth sub-periods produce relatively lower out-of-sample performances in 2005 and 2006. In addition, there is no evidence supporting Qi and Maddala (1998)'s argument that the direct connection in the network architecture negatively affects the performances of the ANN Kalman models.

Overall, the ANN Kalman model that applies the standard variable selection technique with no direct connection achieves the highest consolidated in-sample scores for all of the performance measurement criteria. This model is also the second best performer in the out-of-sample period

for all of the performance measurement criteria. The best achiever for the consolidated out-of-sample scores is the ANN Kalman model that applies the standard variable selection technique with direct connection in its network structure, which achieves best out-of-sample scores for all of the performance measurement criteria. The consolidated in-sample and out-of-sample performance scores indicate that the cascade variable selection technique does not serve to improve the forecasting power of the ANN Kalman models as they do for the ANN backpropagation models.

9

Stock Selection Model Optimisation

9.1 Introduction

This chapter serves to consolidate the performance results for the multifactor linear models developed in Chapter 6, the artificial neural network (ANN) backpropagation models developed in Chapter 7 and the ANN Kalman models developed in Chapter 8. The performances of the best linear model, the ANN backpropagation model and the ANN Kalman model are compared with and contrasted against each other to provide the relative strength and weakness analysis of the respective expected return forecasting models.

The best performance of the linear models is achieved by the model that is trained under the moving-window stepwise procedure to maximise the Qian and Hua information ratio (QH IR) in Chapter 6. This result is in line with the finding of Frankish (2004) in the South African and the U.S. equity markets. As a performance measurement criterion, QH IR provides adequate evaluation of the forecasting accuracy for a factor model since it focuses on the robustness of the model by discounting the forecasting accuracy of the model (that is, the information coefficient, or IC) by the historical volatility of accurate forecasts (represented by the standard deviation of ICs). Based on the concept of mean-variance efficiency, a blended model that performs asset allocations based on the stocks ranked and selected by two separate expected return factor models would achieve an average IC score that approximates the weighted average IC score of the constituent models with its standard deviation of ICs less than the weighted average of the standard deviations of the constituent models. This chapter attempts to build a blended stock selection model from the best linear model identified in Chapter 6 and the best ANN model identified in Chapter 7 or Chapter 8. The comparison of the summarised performance scores of the best linear model, the best ANN backpropagation model and the best ANN Kalman model is demonstrated in Table 9.1. Although the best ANN backpropagation model achieves the best in-sample scores for all performance measurement criteria except the QH IR score, its out-of-sample scores are the

lowest amongst the competing models. By contrast, the best ANN Kalman model achieves robust out-of-sample results across the performance measurement criteria. Overall, the Linear QH IR model achieves the best out-of-sample scores for all criteria and hence is the best expected return factor model constructed in this research. As a result, the Linear QH IR model and the best ANN Kalman model are used as the constituent models to build the blended model.

Table 9.1 Summarised Performances of the Expected Return Factor Models

The performance scores for the best linear, ANN backpropagation and the ANN Kalman models identified in Chapter 6, Chapter 7 and Chapter 8 over the examination period from 01 January 1997 to 31 December 2007 are demonstrated in Table 9.1. The best linear model is the model that employs style attributes that maximise the QH IR score over the training periods. The best ANN backpropagation model is the one that applies the cascade variable pre-selection technique and has no direct connection between the input and the output layers in its network. The best ANN Kalman model is the one that applies standard variable selection technique with direct connection in its network architecture.

	Linear QH IR	ANN Backprop_No Direct (CSV)	ANN Kalman_Direct (SVS)
Avg IC (In-Sample)	0.175	0.192	0.161
Avg IC (Out-of-Sample)	0.169	0.117	0.140
Grinold IR (In-Sample)	1.664	2.168	1.814
Grinold IR (Out-of-Sample)	1.737	1.369	1.634
Qian and Hua IR (In-Sample)	1.290	1.225	1.056
Qian and Hua IR (Out-of-Sample)	1.114	0.838	0.976
Slope *t*-Statistic (In-Sample)	7.999	8.349	7.405
Slope *t*-Statistic (Out-of-Sample)	8.604	6.773	7.638

Section 9.2 describes the methodology of constructing the blended model. Section 9.3 provides an analysis of the blended model relative to that of the constituent models. Section 9.4 consolidates the test results. Fractile analysis of Achour, Harvey, Hopkins and Lang (1999) is conducted to evaluate the ability of the blended model in classifying future winners and losers compared to the constituent models.

9.2 Descriptive Statistics and Methodology

The fractile analysis of Achour *et al* (1999) requires the sample shares to be ranked by their expected returns estimated by the respective expected return factor models at the beginning of each month over the out-of-sample period from 01 January 2002 to 31 December 2007. The sample shares are then assigned in equal numbers to a pre-defined number of fractile on the basis of its rank. Based on the size of the sample in this research, quintile analysis is deemed appropriate.

The quintile analysis is conducted on the best linear model (that is, linear QH IR model), the best artificial neural network (ANN) model and the blended model that ranks stocks based on their weighted average expected returns of the constituent models. The weights allocated to the constituent linear and ANN models are optimised using the constituent models' in-sample average ICs, in-sample standard deviation of ICs and the correlation coefficient of their in-sample ICs. The optimisation procedure restricts the weights assigned to the best linear and ANN models to be positive and sum to 1.0 with the objective of maximising the in-sample QH IR for the blended model. A 5-year moving window procedure is adapted to update the weights annually. The first in-sample period is from 01 January 1997 to 31 December 2001 to determine the weight allocations for 2002. The in-sample period subsequently moves 12 months forward to determine the optimal weights that maximises the QH IR of the blended model from 01 January 1998 to 31 December 2002 to determine the weight allocations for 2003. The training period is updated 5 times to cover the 6-year out-of-sample period from 01 January 2002 to 31 December 2007.

The out-of-sample stock-selection abilities for the top and bottom quintiles of the best linear model, the best ANN model and the blended model are evaluated based on the measures employed by Achour *et al* (1999) such as the mean out-of-sample return, standard deviation, mean excess return, mean excess standard deviation, percentage of periods outperforming the market, percentage of periods outperforming the market up, percentage of period outperforming the market down, maximum consecutive months of outperformance, maximum positive excess monthly return, maximum negative excess return, percentage of periods positive to negative return, percentage of periods of negative return, maximum

consecutive months of positive return, maximum consecutive months of negative return and cumulative growth of R1 since the beginning of the out-of-sample period in 2002. The descriptions and mathematical computations of these performance diagnostics are displayed Table 9.2. The market proxy used to evaluate the performances of the stock selection models is an equally-weighted portfolio of monthly sample shares that serves as a benchmark of the pool where the models select their shares from.

The log cumulative returns for the top and the bottom quintiles for the respective models are evaluated against each other. In addition, the log cumulative spread between the top and the bottom quintile for each model is also evaluated against each other to determine the possibility of developing successful long-short strategies from the respective models. A successful stock selection model is the one that consistently selects future winners for the top quintile and future losers for the bottom quintile regardless of the market trend. Thus, the log cumulative spread for such model will appreciate over time throughout different phases of the business cycle.

Table 9.2 Performance Diagnostics of the Fractile Analysis

The performance diagnostics displayed in Table 9.2 are adapted from Achour, Harvey, Hopkins and Lang (1999). The diagnostics are used to evaluate the performances of the top and bottom quintiles of stocks selected by the best linear model, the best nonlinear model and the blended model over the out-of-sample period from 01 January 1997 to 31 December 2007.

NO.	PERFORMANCE DIAGNOSTICS	DEFINITION
1.	Portfolio Mean Return	Arithmetic average of the monthly quintile return over the evaluation period.
2.	Std Dev Mean Return	Standard deviation of monthly quintile return over the evaluation period.
3.	Sig Mean Return *t*-Statistic	Student's *t*-statistic of the monthly quintile return over the evaluation period.
4.	Sharpe Ratio	The reward-to-risk ratio that is equal to the portfolio excess return per unit of total risk as measured by portfolio standard deviation of returns.
5.	Mean Excess Return	Arithmetic average of the monthly quintile return in excess of the monthly market proxy return over the evaluation period.
6.	Std Dev Mean Excess Return	Standard deviation of post-rank portfolio excess returns above the market portfolio over all observation periods.
7.	Sig Mean Excess Return t-Stat	Test of whether average excess return is significantly different from zero.
8.	% periods > Market	Percentage of total observations that average post-rank portfolio returns was greater than the market portfolio return over the holding period.
9.	% periods > Market Up	Percentage of total observations that average post-rank portfolio returns was greater than the market portfolio return when the market portfolio return was greater than zero.
10.	% periods > Market Down	Percentage of total observations that average post-rank portfolio returns was greater than the market portfolio return when the market portfolio return was less than zero.
11.	Max Consecutive Outperformance	Longest string of consecutive observations where average post-rank portfolio return was greater than the market portfolio return.
12.	Max +ve Excess Return	Highest single post-rank portfolio excess positive return above market portfolio over all observation periods.
13.	Max -ve Excess Return	Lowest single post-rank portfolio excess negative return above market portfolio over all observation periods.
14.	% Periods +ve to -ve Mean Return	Ratio of portfolio post-rank average returns greater than zero to post-rank returns less than zero over all observation periods.
15.	% Periods –ve returns	Percentage of observations that portfolio post-rank returns were less than zero over all observation periods, indicative of the historical probability of losing money.
16.	Max Consecutive +ve returns	Longest string of consecutive observations where average post-rank portfolio return was greater than zero.
17.	Max Consecutive –ve	Longest string of consecutive observations where average post-rank portfolio return was less than zero.
18.	Cumulative Return (Growth of R1)	Value of R1 if invested at the first observation date and compounded over intervening periods.

9.3 Results

Table 9.3 presents the weight allocations and in-sample scores for the best linear (Linear QH IR model) and the best ANN Kalman model used to construct the blended model (QH IR-optimised) over the six overlapping in-sample periods. The top panel of Table 9.3 displays the results of the average IC, standard deviation of ICs and QH IR scores, respectively for the Linear QH IR model and the ANN Kalman model. Based on these periodic scores, the optimised weight allocations for the QH IR-optimised blended model, is presented in the bottom panel. The histograms in Figure 9.1 graphically depict the periodic weight allocations estimated under the optimisation procedure for the blended model with the periodic optimised QH IR score represented by the yellow trend line. The results reveal that the blended model weights more towards the Linear QH IR model in all the periods, except for the period from 1998 to 2002, where the weighting leans more towards the ANN Kalman model (51.80% for the ANN Kalman model and 48.20% for Linear QH IR model). The overall average weight allocation over the overlapping periods remains intact around 50%, which is a split of 54.01% in the Linear QH IR model and 45.99% in the ANN Kalman model.

The fact that the standard deviation of ICs for the blended model is lower than that of each constituent model means that the forecasts of the constituent models are less than 100% correlated, which is evident in the middle panel of Table 9.3. Thus, the blended approach effectively diversifies the forecasting errors embedded in the constituent models, which enables the QH IR of the blended model to be optimised over the respective in-sample periods. The periodic QH IR score achieved by the blended models is approximately twice that of the scores of the constituent models. However the robustness of these scores in the corresponding out-of-sample periods is yet to be determined.

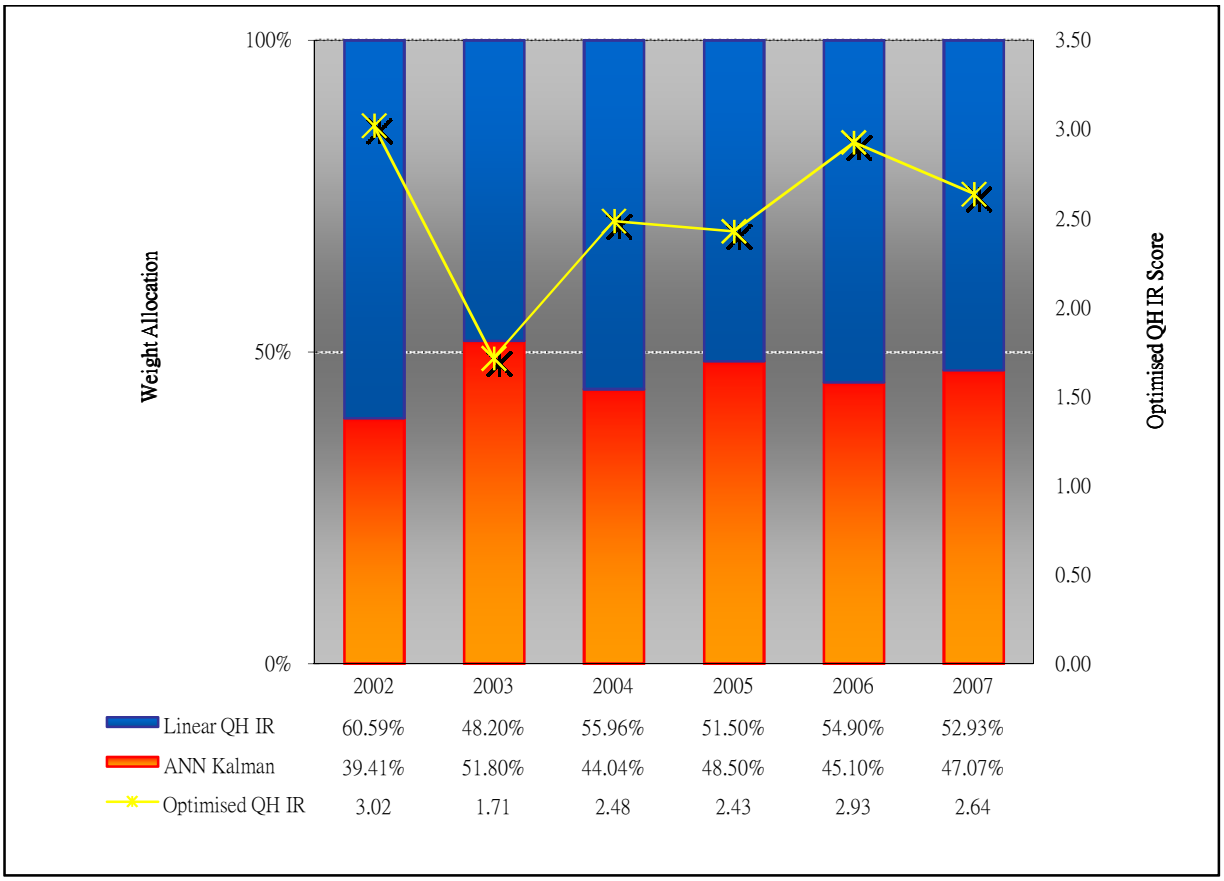
Table 9.3 Weight Allocations and In-Sample Scores for the Qian and Hua (QH) Information Ratio-Optimised Blended Model

Table 9.3 demonstrates the diagnostics for the best linear model (Linear QH IR model) and the best ANN model (ANN Kalman model) employed in the optimisation procedure to construct the QH IR-optimised stock selection model. The top panel provides the periodic in-sample average IC, standard deviation of ICs and QH IR scores, as well as the respective consolidated scores for the Linear QH IR model and the ANN Kalman model. The QH IR of the blended model (bottom panel) estimates the expected returns of the sample shares as the weighted average forecasts of the constituent models. The periodic and consolidated weights, the in-sample average IC and the in-sample standard deviation of ICs and the in-sample optimised QH IR scores for the blended model are demonstrated in the bottom panel of Table 9.3. The middle panel provides the periodic correlation coefficients of ICs between the forecasts of the Linear QH IR model and the ANN Kalman model.

	1997~2001	*1998~2002*	*1999~2003*	*2000~2004*	*2001~2005*	*2002~2006*	**Average**
Linear QH IR Model							
Avg. IC:	20.53%	15.48%	14.58%	17.65%	19.55%	17.18%	**17.50%**
Stdev IC:	13.40%	16.66%	11.69%	13.73%	13.19%	13.56%	**13.70%**
QH IR (Avg. IC/Stdev IC):	1.53	0.93	1.25	1.29	1.48	1.27	**1.29**
ANN Kalman Model							
Avg. IC:	11.27%	13.23%	17.59%	19.57%	20.30%	14.52%	**16.08%**
Stdev IC:	18.13%	15.14%	14.66%	14.76%	15.48%	14.47%	**15.44%**
QH IR (Avg. IC/Stdev IC):	0.62	0.87	1.20	1.33	1.31	1.00	**1.06**
Correlation Coefficient between the ICs of the Linear ANN model and the ANN Kalman Model:	0.74	0.46	0.51	0.42	0.54	0.63	**0.55**
QH IR Optimised (Blended) Model	*1997~2001*	*1998~2002*	*1999~2003*	*2000~2004*	*2001~2005*	*2002~2006*	**Average**
Estimated Weight (Linear QH IR):	60.59%	48.20%	55.96%	51.50%	54.90%	52.93%	**54.01%**
Estimated Weight (ANN Kalman):	39.41%	51.80%	44.04%	48.50%	45.10%	47.07%	**45.99%**
Blended Avg. IC:	16.88%	14.31%	15.91%	18.58%	19.89%	15.92%	**16.92%**
Blended Stdev IC:	5.60%	8.35%	6.40%	7.66%	6.80%	6.04%	**6.81%**
Optimised QH IR (Avg. IC/Stdev IC):	3.02	1.71	2.48	2.43	2.93	2.64	**2.53**

Figure 9.1 Weight Allocations and In-Sample Scores for the Qian and Hua (QH) Information Ratio-Optimised Blended Model

The optimal in-sample weight allocations and corresponding QH IR scores for the blended model are graphically depicted. The blue portion of the histogram, in each in-sample period represents the weight allocation to the linear QH IR model, while the orange portion of the histogram in each in-sample period represents the weight allocation to the ANN Kalman model. The yellow trend line (asterisks) represents the periodic optimised QH IR score for the blended stock selection model over each overlapping in-sample period.



PANEL (A) through PANEL (F) of Table 9.4 present the summarised periodic in- and out-of-sample scores with regard to the average IC, the standard deviation of ICs and the QH IR for the Linear QH IR model, the ANN Kalman model and the QH IR-optimised blended model respectively. The consolidated results over the six sub-periods are displayed in PANEL (G) of Table 9.4. Comparing the performances of the blended model to that of the Linear QH IR model and the ANN Kalman model reveals that the blended model continues to outperform the constituent models in the out-of-sample period since the blended model achieves higher average IC score with medium-to-below-average standard deviation. The consolidated result in PANEL (G) shows that the blended model has achieved the best average IC score with the lowest standard deviation of ICs for the consolidated out-of-sample period from 2002 to 2007, which leads to the dominant outperformance in terms of its QH IR score.

Figure 9.2 graphically depicts the consolidated in- and out-of-sample performance scores of the blended model relative to the Linear QH IR model and the ANN Kalman model. In line with the objective of the blended model, the model has a lower in-sample standard deviation of ICs compared to that of the constituent models. The out-of-sample standard deviation of ICs for the blended model is robust in that it is moderately lower than that of the constituent models. With higher out-of-sample IC and lower out-of-sample standard deviation of ICs, the blended model has achieved the highest out-of-sample forecasting score represented by the average IC per unit of the volatility of its forecasting error.

Table 9.4 Periodic Performance Scores of the Qian and Hua (QH) Information Ratio-Optimised Blended Model

The periodic performance scores of the Linear QH IR model, the ANN Kalman model and the blended model are cross-examined on the basis of their respective average ICs, standard deviation of ICs and QH IR scores during the evaluation periods. The performance scores of the models for the 6 moving-window in-sample periods and their corresponding out-of-sample (forecasting) periods from 2002 to 2007 are displayed in PANEL (A) through PANEL (F) respectively. PANEL (G) displays the consolidated in-sample and out-of-sample scores for the respective models.

PANEL (A)	**In-Sample: 1 Jan 1997 to 31 Dec 2001**	**Out-of Sample: 1 Jan 2002 to 31 Dec 2002**	
	Linear QH IR	ANN Kalman	Blended: Linear_Kalman
Average IC (In-Sample)	0.2053	0.1128	0.1688
Average IC (Out-Sample)	0.1167	0.1268	0.1781
Stdev. IC (In-Sample)	0.1340	0.1815	0.0560
Stdev. IC (Out-Sample)	0.1843	0.1372	0.1657
QH IR (In-Sample)	1.5325	0.6215	3.0150
QH IR (Out-Sample)	0.6333	0.9245	1.0752

PANEL (B)	**In-Sample: 1 Jan 1998 to 31 Dec 2002**	**Out-of Sample: 1 Jan 2003 to 31 Dec 2003**	
	Linear QH IR	ANN Kalman	Blended: Linear_Kalman
Average IC (In-Sample)	0.1548	0.1323	0.1431
Average IC (Out-Sample)	0.2183	0.1994	0.2303
Stdev. IC (In-Sample)	0.1665	0.1514	0.0835
Stdev. IC (Out-Sample)	0.1250	0.1079	0.1389
QH IR (In-Sample)	0.9295	0.8737	1.7134
QH IR (Out-Sample)	1.7464	1.8488	1.6584

Table 9.4 Periodic Performance Scores of the Qian and Hua (QH) Information Ratio-Optimised Blended Model – Continued

PANEL (C)	In-Sample: 1 Jan 1999 to 31 Dec 2003	Out-of Sample: 1 Jan 2004 to 31 Dec 2004	
	Linear QH IR	ANN Kalman	Blended: Linear_Kalman
Average IC (In-Sample)	0.1458	0.1759	0.1591
Average IC (Out-Sample)	0.2375	0.2099	0.2602
Stdev. IC (In-Sample)	0.1168	0.1466	0.0640
Stdev. IC (Out-Sample)	0.2101	0.1847	0.2092
QH IR (In-Sample)	1.2480	1.1998	2.4847
QH IR (Out-Sample)	1.1303	1.1365	1.2439

PANEL (D)	In-Sample: 1 Jan 2000 to 31 Dec 2004	Out-of Sample: 1 Jan 2005 to 31 Dec 2005	
	Linear QH IR	ANN Kalman	Blended: Linear_Kalman
Average IC (In-Sample)	0.1765	0.1956	0.1858
Average IC (Out-Sample)	0.1228	0.0911	0.1218
Stdev. IC (In-Sample)	0.1373	0.1476	0.0766
Stdev. IC (Out-Sample)	0.1227	0.1731	0.1616
QH IR (In-Sample)	1.2857	1.3255	2.4256
QH IR (Out-Sample)	1.0005	0.5264	0.7536

PANEL (E)	In-Sample: 1 Jan 2001 to 31 Dec 2005	Out-of Sample: 1 Jan 2006 to 31 Dec 2006	
	Linear QH IR	ANN Kalman	Blended: Linear_Kalman
Average IC (In-Sample)	0.1955	0.2030	0.1989
Average IC (Out-Sample)	0.1500	0.0820	0.1493
Stdev. IC (In-Sample)	0.1319	0.1548	0.0680
Stdev. IC (Out-Sample)	0.0885	0.0832	0.0797
QH IR (In-Sample)	1.4826	1.3116	2.9264
QH IR (Out-Sample)	1.6951	0.9851	1.8733

Table 9.4 Periodic Performance Scores of the Qian and Hua (QH) Information Ratio-Optimised Blended Model – Continued

PANEL (F)	In-Sample: 1 Jan 2002 to 31 Dec 2006	Out-of Sample: 1 Jan 2007 to 31 Dec 2007	
	Linear QH IR	ANN Kalman	Blended: Linear_Kalman
Average IC (In-Sample)	0.1718	0.1451	0.1592
Average IC (Out-Sample)	0.1761	0.1179	0.1715
Stdev. IC (In-Sample)	0.1356	0.1446	0.0604
Stdev. IC (Out-Sample)	0.1038	0.0893	0.0930
QH IR (In-Sample)	1.2665	1.0035	2.6387
QH IR (Out-Sample)	1.6964	1.3200	1.8437

PANEL (G)	Consolidated In-Sample and Out-of-Sample Performance Statistics		
	Linear QH IR	**ANN Kalman**	**Blended: Linear_Kalman**
Average IC (In-Sample)	**0.1750**	**0.1608**	**0.1692**
Average IC (Out-Sample)	**0.1689**	**0.1397**	**0.1852**
Stdev. IC (In-Sample)	**0.1356**	**0.1523**	**0.0681**
Stdev. IC (Out-Sample)	**0.1517**	**0.1431**	**0.1413**
QH IR (In-Sample)	**1.2908**	**1.0560**	**2.5340**
QH IR (Out-Sample)	**1.1144**	**0.9762**	**1.4080**

Figure 9.2 Consolidated Performance Scores of the Qian and Hua (QH) Information Ratio-Optimised Blended Model Relative to the Constituent Models

Figure 9.2 plots the moving-window (rolling) Linear QH IR model, the ANN Kalman model and the blended (Linear_Kalman) model in the diagram based on their respective consolidated average IC scores and standard deviation of ICs for the in-sample and out-of-sample periods. The slope of the line connecting the origin to the position of the model in the diagram is essentially the model's QH IR score. The filled labels represent in-sample scores while the corresponding out-of-sample scores are represented by empty labels.

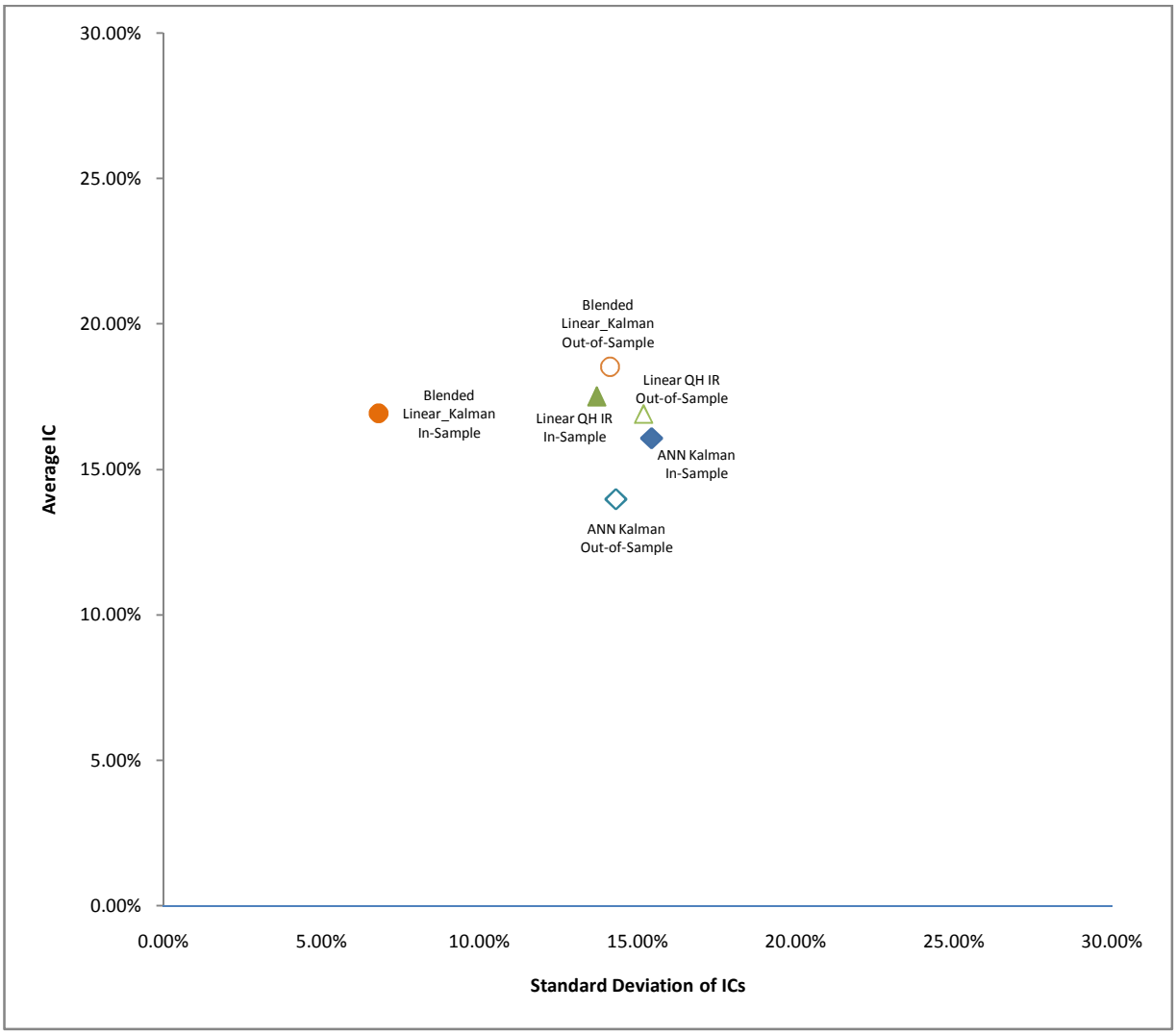


Figure 9.3 provides the log cumulative performances of the top and bottom quintiles of the Linear QH IR model, the ANN Kalman model and the blended model over the out-of-sample period from 01 January 2002 to 31 December 2007. It is apparent that the top quintile (Q1) of the ANN Kalman model outperforms the top quintile of the blended model, which in turn outperforms the top quintile of the Linear QH IR model over the first half of the out-of-sample period from 01 January 2002 to 31 December 2004. The performances of the top quintiles of the respective models subsequently reverse in the second half of the out-of-sample period from 01 January 2005 to 31 December 2007. Although no distinctive performances of the bottom quintiles of the respective models are identified over the first half of the out-of-sample period, the Linear QH IR model appears to identify the worst performers for the bottom quintile (Q5) compared to the ANN Kalman model and the blended model over the second half of the out-of-sample period. Overall, the top quintiles of all three models earn higher cumulative returns relative to the market proxy, which in turn outperforms the bottom quintiles of the respective models.

Figure 9.4 depicts the out-of-sample log cumulative quintile spread as the difference between the top (Q1) and bottom quintile (Q5) returns of the three models over the examination period from 01 January 2002 to 31 December 2007. This is representative of a scenario that invests in the top quintile of the respective strategies and simultaneously holds an equal amount of short position in the bottom quintile. The strategy is rebalanced every month to maintain a neutral position in the top and the bottom quintiles of the strategy. A clear kink in the cumulative spread of the respective models since the beginning of 2005 is documented. This suggests that it becomes less likely for the long-short strategies to accumulate value in the second half of the out-of-sample period.

Figure 9.3 Out-of-Sample Log Cumulative Quintile Returns of the Stock Selection Models

Figure 9.3 depicts the log cumulative returns of the top and bottom quintiles for the Linear QH IR model, the ANN Kalman model and the blended (Linear_Kalman) model over the out-of-sample period from 01 January 2002 to 31 December 2007. The solid red, blue and yellow trend lines represent the log cumulative returns for top quintiles selected by the Linear QH IR model, the ANN Kalman model and the blended model respectively. The dashed red, blue and yellow trend lines represent the log cumulative returns for bottom quintiles selected by the Linear QH IR model, the ANN Kalman model and the blended model respectively. The solid white trend line represents the equally-weighted sample shares listed on the JSE (EWALSI) from which the respective models select their shares.

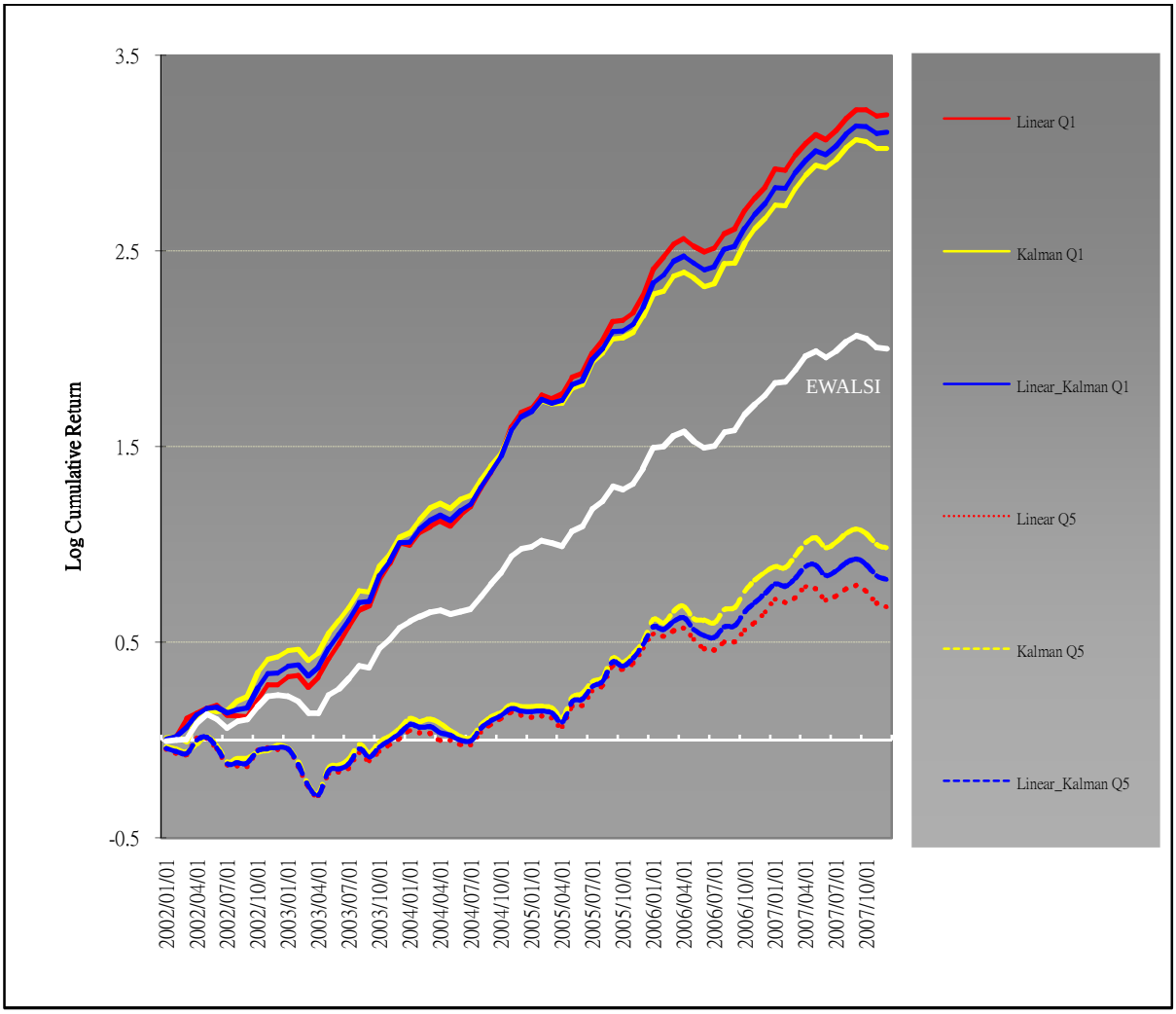


Figure 9.4 Out-of-Sample Log Cumulative Quintile Spreads of the Stock Selection Models

Figure 9.4 depicts the out-of-sample log cumulative quintile spreads as the difference between the top and bottom quintile returns for the Linear QH IR model, the Kalman ANN model and the blended (Linear_Kalman) model over the examination period from 01 January 2002 to 31 December 2007. The red, yellow and blue trend lines represent strategies that invest in the top quintiles and simultaneously hold an equal amount of short positions in the bottom quintiles of the respective models. The position is rebalanced every month to maintain a neutral position in the long and short positions respectively.

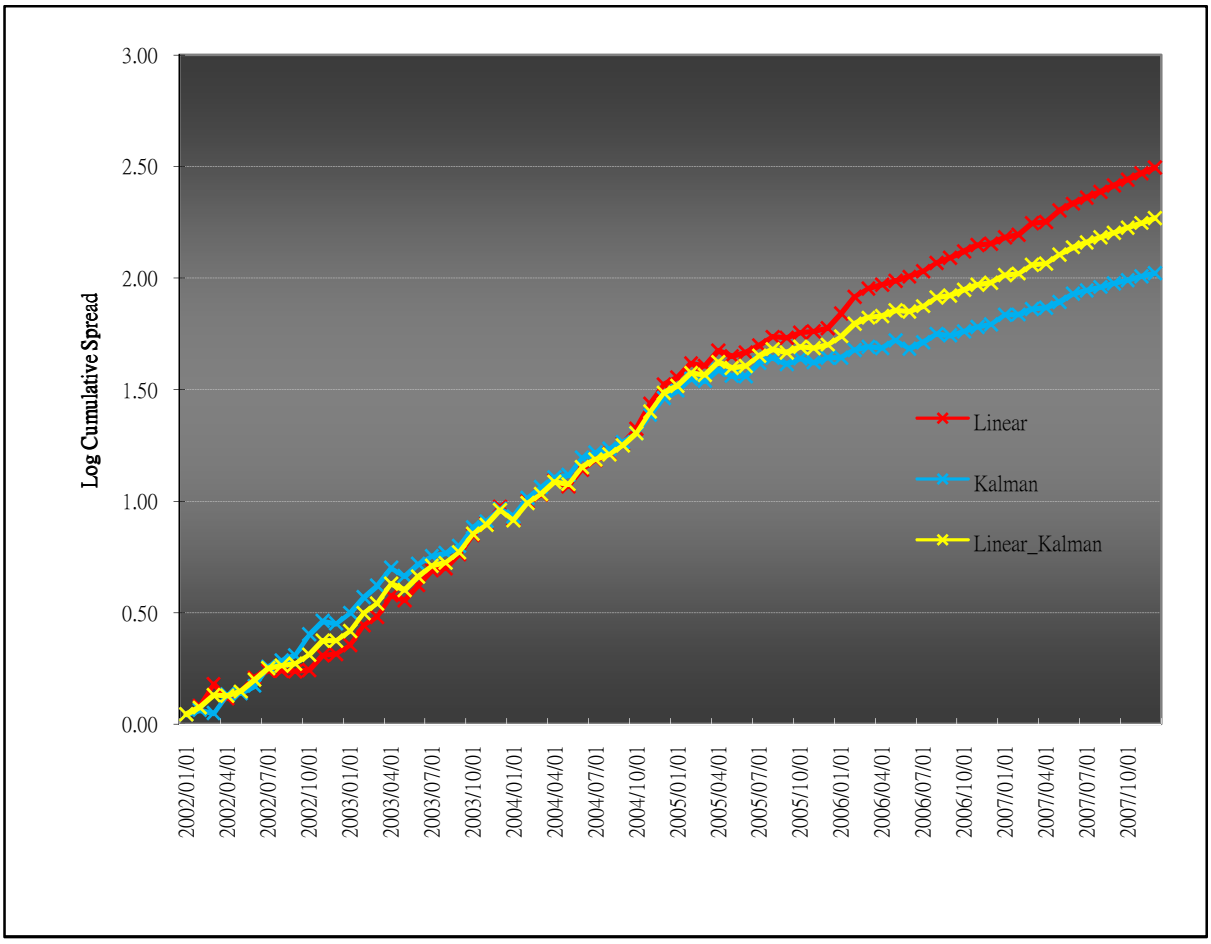


Table 9.5 demonstrates the out-of-sample performance diagnostics for the top and bottom quintiles of the Linear QH IR model, the ANN Kalman and the blended model. Examining the top quintile (Quintile 1) performances of the respective models in the out-of-sample period from 01 January 2002 to 31 December 2007, it is apparent that the portfolio mean return for the blended model is in-between the portfolio mean returns of its constituent models, while its standard deviation is less than the standard deviation of its constituent models. This leads to a greater *t*-statistic for the mean portfolio return of the blended model (8.529) relative to the *t*-statistics for the mean portfolio returns of the Linear QH IR model (8.235) and the ANN Kalman model (8.146). The resulting Sharpe Ratio of the top quintile for the blended model (0.889) outperforms the Sharpe ratios of the top quintiles for the Linear QH IR model (0.863) and for the ANN Kalman model (0.846). When the performance of the top quintile for each of the models is compared to the equally-weighted market proxy comprised of sample shares, the mean excess return of the blended model (0.31) remain in-between that of the Linear QH IR model (0.32) and the ANN Kalman model (0.30) with its standard deviation of the mean excess return (0.038) lower than that of the Linear QH IR model (0.041) and the ANN Kalman model (0.041). As a result, the *t*-statistic of the mean excess return for the blended model (6.588) is higher than that of the Linear QH IR model (6.412) and the ANN Kalman model (5.944).

Overall, the top quintiles of all three models demonstrate their abilities to earn significant mean portfolio returns and outperform the market proxy at the 5% significance level. These results are confirmed by their abilities of beating the market over the out-of-sample period, during upturn of the market and during the market downturn (refer to diagnostics 8, 9 and 10 respectively). The highest probabilities of outperformance for the three scenarios are achieved by the blended model (83.3%, 78.0% and 92.0% respectively) over the examination period. The maximum consecutive outperformance (diagnostic 11) over the out-of-sample period is also achieved by the blended model (12 times). While the ANN Kalman model has the highest positive excess return of 14.0% (diagnostic 12), the lowest negative return (diagnostic 13) is obtained by the Linear QH IR model (-7.1%). On the other hand, the blended model only makes loss in 12.1% of the months over the out-of-sample period (diagnostic 15) compared to 18.2% for the Linear QH IR model and 15.2% for the ANN Kalman model. As a result, the ratio of the percentage periods positive to negative mean return (diagnostic 14) is the highest for the blended model (7.25 times) compared to the Linear QH IR model (4.50 times) and the

ANN Kalman model (5.60 times). The maximum consecutive positive performance (diagnostic 16) is the same for all three models and the maximum consecutive negative performance (diagnostic 17) is also similar for all three models. Overall, the Linear QH IR model accumulates the highest return (21.507) over the out-of-sample period, followed by the blended model (19.922) and the ANN Kalman model (18.642).

With regard to the performances of the bottom quintiles (Quintile 5) for the respective models, the blended model produces monthly mean return (0.014) in-between that of the Linear QH IR model (0.012) and the ANN Kalman model (0.016). The standard deviations of the mean portfolio returns for all three models are in line with each other. As a result, the Sharpe ratio (diagnostic 4) of the blended model (0.140) is also in-between that of the Linear QH IR model (0.099) and the Kalman model (0.183). The *t*-statistics of the mean portfolio returns (diagnostic 3) for the bottom quintiles of the respective models remain significant at the 5% level, but are much lower than the *t*-statistics of the mean portfolio returns for the top quintiles. When the performances of the bottom quintiles are evaluated against the market proxy comprised of the sample shares, the mean excess returns (diagnostic 5) are insignificantly negative for the Linear QH IR model and the blended model. However, the mean excess return for the ANN Kalman model is approximately zero. Although the performance of the bottom quintile for the Linear QH IR model is relatively lower compared to the other two models, the overall results suggest that all three models have failed to successfully identify future losers amongst the sample shares. In addition, the probabilities for the bottom quintiles to outperform the market proxy in the out-of-sample period (diagnostic 8) are below but close to 50% for all three models. These probabilities are even higher and above 60% during upturn of the market (diagnostic 10). These observations serve as evidence supporting the above suggestion that all three models are ineffective in identifying market losers. Overall, the bottom quintile of the Linear QH IR model accumulates the least return (2.048) followed by the bottom quintiles of the blended model (2.320) and the ANN Kalman model (2.673) over the out-of-sample period.

Table 9.5 Out-of-Sample Model Performance Diagnostics for the Stock Selection Models

Table 9.5 displays the out-of-sample performance diagnostics for the top and bottom quintiles of the Linear QH IR model, the ANN Kalman model and the blended (Linear_Kalman) model respectively. The left panel displays the out-of sample stock selection results of the bottom quintile (fifth quintile) of the respective models, while the right panel represents the out-of-sample stock selection results of the top quintile (first quintile) of the respective models.

Out-of-Sample Period :	**Bottom Quintile (Quintile 5)**			**Top Quintile (Quintile 1)**		
1 January 2002 to 31 December 2007	**Linear QH IR**	**ANN Kalman**	**Blended Linear_Kalman**	**Linear QH IR**	**ANN Kalman**	**Blended Linear_Kalman**
1. Portfolio Mean Return	0.012	0.016	0.014	0.049	0.046	0.047
2. Std Dev Mean Return	0.048	0.049	0.048	0.048	0.046	0.045
3. Sig Mean Return t-Statistics	2.029	2.697	2.377	8.235	8.146	8.529
4. Sharpe Ratio	0.099	0.183	0.140	0.863	0.846	0.889
5. Mean Excess Return	-0.004	0.000	-0.003	0.032	0.030	0.031
6. Std Dev Mean Excess Return	0.033	0.026	0.029	0.041	0.041	0.038
7. Sig Mean Excess Return t-Statistics	-1.079	-0.089	-0.710	6.412	5.944	6.588
8. % periods > Market	0.439	0.470	0.439	0.788	0.803	0.833
9. % periods > Market Up	0.317	0.366	0.317	0.756	0.732	0.780
10. % periods > Market Down	0.640	0.640	0.640	0.840	0.920	0.920
11. Max Consecutive Outperformance	5	5	5	11	9	12
12. Max +ve Excess Return	0.090	0.054	0.071	0.113	0.140	0.106
13. Max -ve Excess Return	-0.090	-0.051	-0.072	-0.071	-0.088	-0.078
14. % Periods +ve to -ve Mean Return	1.200	1.538	1.444	4.500	5.600	7.250
15. % Periods –ve	0.455	0.394	0.409	0.182	0.152	0.121
16. Max Consecutive +ve	6	6	6	13	13	13
17. Max Consecutive –ve	5	4	4	3	2	2
18. Cumulative Return (Growth of R1)	2.048	2.673	2.320	21.507	18.642	19.922

9.4 Conclusion

The results of the optimisation procedure reveal that the optimal in-sample Qian and Hua (QH) information ratio is achieved by allocating greater weight to the best linear factor model (Linear QH IR model) relative to the weight allocated to the best ANN model (ANN Kalman model with direct connection from the input layer neurons to the output neuron and standard variable selection technique). The periodic QH IR scores achieved by the blended model are approximately twice the scores of the constituent models over the six overlapping in-sample periods, and are robust over the out-of-sample period in that the blended model continues to obtain higher average IC score with median-to-low level of standard deviation of ICs. Overall, the blended model has achieved the best average IC score with the lowest standard deviation of ICs for the consolidated out-of-sample period from 01 January 2002 to 31 December 2007. As a result, the blended model achieves the best QH IR score in the out-of-sample period. This indicates that the blended model has the highest out-of-sample forecasting ability per unit of the volatility of forecasting error.

In terms of the periodic fractile analysis over the out-of-sample period, the bottom quintiles (Quintile 5) of the three models are successfully identified and compatible of each other prior to 2005 in that they fail to accumulate positive returns and remain underperforming the market proxy over the period. However, their log cumulative returns start to take off and turn positive after 2005. During this period, the bottom quintile of the ANN Kalman model outperforms that of the bottom quintile of the blended model, which in turn outperforms the bottom quintile of the Linear QH IR model. The failure for the three models to identify the true loser shares since 2005 severely drags down their abilities to accumulate the top-bottom quintile spread. The consolidated out-of-sample diagnostics also reveal that the three models successfully identify true winning shares for the top quintiles that earn statistically significant returns above the market return. However, strong evidence is found that all three models are ineffective in identifying future losers. Overall, the blended model effectively diversifies the volatility of the forecasting errors of the constituent Linear QH IR model and the ANN Kalman model, and hence is the best stock-selection model identified in this research.

Appendix A
Computations of Firm-Specific Style Attributes

The formulae used to compute the respective firm-specific attributes are based on the information provided by the DataStream International database. The full descriptions of the attribute codes is displayed in Table 4.1. Note that cash earnings refers to the net cash flow from operating, financing and investment activities (from the cash flow statement), whereas cash refers to cash holdings as part of the current assets. The return index takes into account both dividend yield and capital gains from the appreciation or depreciation of the share price.

(I) FUNDAMENTAL VALUES RELATIVE TO SHARE PRICE

1. BVTP
   Book Value of Equity / Share Price

2. CFTP
   Cash Earnings per Share / Share Price

3. DY
   Ordinary Shareholders Dividends per Share / Share Price

4. EY
   EPS / Share Price

5. SALESTP
   Sales per Share / Share Price

(II) SOLVENCY AND LIQUIDITY

6. CFTCURRLIABS
   Net Cash Flow / Current Liabilities

7. CFTDEBT
   Net Cash Flow / Total Liabilities

8. CURRENTRATIO
   Current Assets / Current Liabilities

9. DEBTTMVE
   Total Liabilities / Market Value of Equity

10. DEBTTBVE
    Total Liabilities / Book Value of Equity

11. ICBT
    Profit Before Interest and Tax / Accrued Interest

(III) FUNDAMENTAL GROWTH

12. G12MCPS
    (Current Cash Holdings per Share / Prior 12-Month Cash Holdings per Share) - 1

13. G12MDPS
    (Current DPS / Prior 12-Month DPS) - 1

14. G12MEPS
    (Current EPS / Prior 12-Month EPS) - 1

15. G12MGPMARGIN
    (Current Gross Profit Margin / Prior 12-Month Gross Profit Margin) - 1

16. G12MNPMARGIN
    (Current Net Profit Margin / Prior 12-Month Net Profit Margin) - 1

17. G12MSALES
    (Current Sales / Prior 12-Month Sales) - 1

18. G24MEPS
    (Current EPS / Prior 24-Month Earnings per Share) - 1

19. GROWTH
    Return on Equity * (1 – Dividend Payout Ratio)

(IV) OPERATING PERFORMANCE

20. GPMARGIN
    Gross Profit / Sales

21. NPMARGIN
    Net Profit after Tax / Sales

22. PAYOUT
    Current Dividend per Share /Current Earnings per Share

23. ROA
    Net Profit Before Tax / Total Assets

24. ROE
    Net Profit After Tax / Ordinary Shareholders' Equity

25. TATURNOVR
    Sales / Total Assets

(V) SIZE AND RETURN MOMENTUM

26. LAGLPRICE
    Ln (Prior 1-Month Share Price)

27. LPRICE
    Ln (Current Share Price)

28. LSIZE
    Ln (Market Capitalisation)

29. MOM1
    (Current Return Index /Prior 1-Month Return Index) - 1

30. MOM12
    (Current Return Index / Prior 12-Month Return Index) - 1

31. MOM12-1
    (Prior 1-Month Return Index / Prior 12-Month Return Index) - 1

32. MOM24
    (Current Return Index / Prior 24-Month Return Index) - 1

33. MOM3
    (Current Return Index / Prior 3-Month Return Index) - 1

34. MOM6
    (Current Return Index / Prior 6-Month Return Index) - 1

(VI) CONSENSUS ANALYST FORECAST

35. EARNREV
    (Consensus Next EPS Forecast / Consensus Previous EPS Forecast) - 1

36. EG1
    (Consensus Next EPS Forecast / Current EPS) - 1

37. FOREPS1
    Consensus EPS Forecast 1-year Forward

38. FOREPS2
    Consensus EPS Forecast 2-year Forward

Appendix B.1 Average Information Coefficient Model (IC)

Appendix B.1 demonstrates the order in which each of the style attributes are included in the average IC model in each of the sub-periods. The performance scores for the models with and without a regression constant are displayed in the top and bottom panels of the table respectively. The style attributes with significant mean factor payoffs are in italics and highlighted in bold. The style attributes assigned with a single asterisk are those being eliminated in the model due to the fact that they do not contribute to the improvements of the respective performance scores. On the other hand, the style attributes assigned with double asterisks are those attributes that are highly correlated with a correlation coefficient of greater than 0.70 or less than -0.70 to one or more of the attributes in the prior iterations.

In-Sample: 1 Jan 1997 to 31 Dec 2001
Out-of-Sample: 2002

With Regression Constant:

		In	Out
1	***BVTP***	0.0891	0.1602
2	***FOREY2***	0.1512	0.0169
3	***CFTP***	0.1657	0.0096
4	CURRENTRATIO	0.1936	0.1510
5	DEBTTMVE	0.2088	0.1279
6	EY	0.2249	0.1540
7	***LSIZE***	0.2291	0.1217
8	***MOM1***	0.2299	0.1109
9	****SALESTP***	0.2282	0.1135
10			

Without Regression Constant:

		In	Out
1	***CFTP***	0.0867	0.0244
2	FOREY2	0.1541	0.0051
3	CURRENTRATIO	0.1910	0.1507
4	***EY***	0.2048	0.1489
5	DEBTTMVE	0.2260	0.1453
6	****BVTP***	0.2257	0.1491
7			
8			
9			
10			

In-Sample: 1 Jan 1998 to 31 Dec 2002
Out-of-Sample: 2003

With Regression Constant:

		In	Out
1	***BVTP***	0.1202	0.1439
2	MOM12	0.1417	0.1437
3	DEBTTBVE	0.1464	0.1720
4	*MOM1	0.1457	0.1644
5			
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.1227	0.1439
2	MOM12	0.1416	0.1440
3	***DEBTTBVE***	0.1458	0.1712
4	*MOM1	0.1458	0.1648
5			
6			
7			
8			
9			
10			

In-Sample: 1 Jan 1999 to 31 Dec 2003
Out-of-Sample: 2004

With Regression Constant:

		In	Out
1	***BVTP***	0.1502	0.0909
2	***ROE***	0.1621	0.1921
3	***CFTP***	0.1748	0.2183
4	DEBTTMVE	0.1810	0.2159
5	*****DEBTTBVE***	0.1820	0.2132
6	****LPRICE***	0.1774	0.2354
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.1502	0.0909
2	ROE	0.1610	0.1965
3	***CFTP***	0.1745	0.2207
4	DEBTTMVE	0.1805	0.2199
5	*****DEBTTBVE***	0.1809	0.2110
6	****LPRICE***	0.1766	0.2315
7			
8			
9			
10			

In-Sample: 1 Jan 2000 to 31 Dec 2004
Out-of-Sample: 2005

With Regression Constant:

		In	Out
1	***BVTP***	0.1549	0.0848
2	ROE	0.1830	0.0967
3	***CFTP***	0.1943	0.1039
4	G12MGPMARGIN	0.2093	0.1584
5	DEBTTMVE	0.2102	0.1566
6	MOM1	0.2123	0.1650
7	*CFTCURRLIABS	0.2116	0.1481
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.1549	0.0848
2	ROE	0.1831	0.0926
3	***CFTP***	0.1930	0.0978
4	DEBTTMVE	0.1983	0.0905
5	***LPRICE***	0.1994	0.0983
6	*MOM12	0.1973	0.0963
7			
8			
9			
10			

In-Sample: 1 Jan 2001 to 31 Dec 2005
Out-of-Sample: 2006

With Regression Constant:

		In	Out
1	***BVTP***	0.1390	0.0478
2	ROE	0.1712	0.0485
3	***CFTP***	0.1859	0.0714
4	G12MGPMARGIN	0.2159	0.0514
5	*ICBT	0.2151	0.0442
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.1390	0.0478
2	ROE	0.1712	0.0485
3	***CFTP***	0.1852	0.0264
4	G12MGPMARGIN	0.1935	-0.1008
5	ICBT	0.1947	-0.0923
6	DEBTTMVE	0.1966	-0.0887
7	*MOM1	0.1931	-0.0223
8			
9			
10			

In-Sample: 1 Jan 2002 to 31 Dec 2006
Out-of-Sample: 2007

With Regression Constant:

		In	Out
1	***LSIZE***	0.1145	0.0487
2	***G12MNPMARGIN***	0.1288	0.0137
3	***CFTP***	0.1523	0.0717
4	GROWTH	0.1580	0.0562
5	MOM12-1	0.1668	0.1141
6	***MOM3***	0.1777	0.1783
7	***BVTP***	0.1839	0.1705
8	TATURNOVR	0.1852	0.1729
9	*****LPRICE***	0.1861	0.1710
10	****MOM12***	0.1858	0.1868

Without Regression Constant:

		In	Out
1	***LSIZE***	0.1145	0.0487
2	***G12MNPMARGIN***	0.1291	0.0206
3	***CFTP***	0.1495	0.0674
4	GROWTH	0.1590	0.0539
5	***MOM12-1***	0.1678	0.1004
6	MOM3	0.1770	0.1712
7	*****LPRICE***	0.1772	0.1699
8	*G12MSALES	0.1760	0.1794
9			
10			

Appendix B.2 Grinold Information Ratio Model (Grinold IR)

Appendix B.2 demonstrates the order in which each of the style attributes are included in the Grinold model in each of the sub-periods. The performance scores for the models with and without a regression constant are displayed in the top and bottom panels of the table respectively. The style attributes with significant mean factor payoffs are in italics and highlighted in bold. The style attributes assigned with a single asterisk are those being eliminated in the model due to the fact that they do not contribute to the improvements of the respective performance scores. On the other hand, the style attributes assigned with double asterisks are those attributes that are highly correlated with a correlation coefficient of greater than 0.70 or less than -0.70 to one or more of the attributes in the prior iterations.

In-Sample: 1 Jan 1997 to 31 Dec 2001
Out-of-Sample: 2002

With Regression Constant:

		In	Out
1	***BVTP***	0.9459	1.7633
2	***FOREY2***	1.3190	0.1621
3	***CFTP***	1.4356	0.0958
4	CURRENTRATIO	1.4712	1.1950
5	DEBTTMVE	1.5789	0.9901
6	EY	1.6714	1.1695
7	***LSIZE***	1.7039	0.9170
8	****MOM1***	1.6933	0.8153
9			
10			

Without Regression Constant:

		In	Out
1	***CFTP***	0.9057	0.2705
2	FOREY2	1.3329	0.0550
3	CURRENTRATIO	1.4547	1.1911
4	DEBTTMVE	1.5360	0.8999
5	***EY***	1.6796	1.1021
6	****BVTP***	1.6770	1.1305
7			
8			
9			
10			

In-Sample: 1 Jan 1998 to 31 Dec 2002
Out-of-Sample: 2003

With Regression Constant:

		In	Out
1	***BVTP***	1.2934	1.6138
2	MOM12	1.4549	1.5762
3	DEBTTBVE	1.4761	1.8629
4	****CFTP***	1.4438	2.1105
5			
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	1.3185	1.6138
2	MOM12	1.4535	1.5787
3	***DEBTTBVE***	1.4695	1.8544
4	*MOM1	1.4365	1.7605
5			
6			
7			
8			
9			
10			

In-Sample: 1 Jan 1999 to 31 Dec 2003
Out-of-Sample: 2004

With Regression Constant:

		In	Out
1	***BVTP***	1.6194	1.0243
2	***DEBTTBVE***	1.7190	1.3830
3	***CFTP***	1.8486	1.8008
4	ROA	1.8655	2.0340
5	*DEBTTMVE	1.8467	1.9682
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	1.6194	1.0243
2	***DEBTTBVE***	1.7177	1.3766
3	***CFTP***	1.8384	1.7459
4	*ROE	1.8361	2.3935
5			
6			
7			
8			
9			
10			

In-Sample: 1 Jan 2000 to 31 Dec 2004
Out-of-Sample: 2005

With Regression Constant:

		In	Out
1	***BVTP***	1.6848	0.9671
2	ROE	1.9478	1.0870
3	***CFTP***	2.0697	1.1683
4	DEBTTMVE	2.1031	1.1295
5	***LPRICE***	2.1144	1.1832
6	*MOM12	2.1054	1.0980
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	1.6848	0.9671
2	ROE	1.9492	1.0408
3	***CFTP***	2.0561	1.1003
4	DEBTTMVE	2.1024	1.0160
5	***LPRICE***	2.1144	1.1053
6	****LSIZE***	2.0883	1.0151
7			
8			
9			
10			

In-Sample: 1 Jan 2001 to 31 Dec 2005
Out-of-Sample: 2006

With Regression Constant:

		In	Out
1	***BVTP***	1.5344	0.5579
2	ROE	1.8477	0.5568
3	***CFTP***	2.0066	0.8194
4	***LPRICE***	2.0753	0.7057
5	***MOM12***	2.0929	1.2361
6	ROA	2.1710	1.0891
7	MOM3	2.1993	1.4999
8	****LSIZE***	2.1829	1.4156
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	1.5344	0.5579
2	ROE	1.8482	0.5555
3	***CFTP***	1.9984	0.3017
4	***LPRICE***	2.0724	0.2930
5	MOM12	2.0906	0.8962
6	*****LAGLPRICE***	2.1182	1.3071
7	****DEBTTBVE***	2.0857	1.2367
8			
9			
10			

In-Sample: 1 Jan 2002 to 31 Dec 2006
Out-of-Sample: 2007

With Regression Constant:

		In	Out
1	***LSIZE***	1.3190	0.6042
2	MOM12-1	1.3737	1.6451
3	***EY***	1.5570	1.4643
4	***MOM12***	1.7153	2.0971
5	***BVTP***	1.7687	2.3058
6	ROE	1.8460	2.3617
7	***CFTP***	2.0081	2.4345
8	***MOM3***	2.0353	2.2615
9	*****LPRICE***	2.0443	2.2481
10	*****LAGLPRICE***	2.0780	1.7977
11	*TATURNOVR	2.0595	1.6783

Without Regression Constant:

		In	Out
1	***LSIZE***	1.3190	0.6042
2	***BVTP***	1.3710	0.4958
3	ROE	1.5694	0.2342
4	***MOM12-1***	1.7025	1.2820
5	***CFTP***	1.7308	1.4753
6	***EY***	1.7948	1.6670
7	MOM12	1.8594	2.4686
8	****LPRICE***	1.8543	2.4463
9			
10			

Appendix B.3 Qian and Hua Information Ratio Model (Q&H IR)

Appendix B.3 demonstrates the order in which each of the style attributes are included in the QH IR model in each of the sub-periods. The performance scores for the models with and without a regression constant are displayed in the top and bottom panels of the table respectively. The style attributes with significant mean factor payoffs are in italics and highlighted in bold. The style attributes assigned with a single asterisk are those being eliminated in the model due to the fact that they do not contribute to the improvements of the respective performance scores. On the other hand, the style attributes assigned with double asterisks are those attributes that are highly correlated with a correlation coefficient of greater than 0.70 or less than -0.70 to one or more of the attributes in the prior iterations.

In-Sample: 1 Jan 1997 to 31 Dec 2001
Out-of-Sample: 2002

With Regression Constant:

		In	Out
1	FOREY2	0.6072	0.4106
2	***CFTP***	0.9036	0.0406
3	CURRENTRATIO	1.0412	0.9206
4	***BVTP***	1.1982	0.8787
5	EY	1.3863	0.9365
6	TATURNOVR	1.4730	0.7593
7	***SALESTP***	1.5324	0.7705
8	***LAGLPRICE***	1.5325	0.6333
9	****MOM1***	1.5299	0.5435
10			

Without Regression Constant:

		In	Out
1	FOREY2	0.6072	0.4106
2	***CFTP***	0.8996	0.0272
3	CURRENTRATIO	1.0584	0.9205
4	***BVTP***	1.2851	0.9769
5	***EY***	1.5158	0.9064
6	TATURNOVR	1.6178	0.8920
7	****SALESTP***	1.6104	0.7879
8			
9			
10			

In-Sample: 1 Jan 1998 to 31 Dec 2002
Out-of-Sample: 2003

With Regression Constant:

		In	Out
1	***BVTP***	0.5755	1.0825
2	***MOM12-1***	0.8555	1.3319
3	DEBTTMVE	0.8853	1.6450
4	EY	0.8877	1.4720
5	***CFTP***	0.9295	1.7464
6	****FOREY2***	0.9267	1.2789
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.5918	1.0825
2	***MOM12-1***	0.8591	1.3616
3	DEBTTMVE	0.8884	1.6861
4	****DEBTTBVE***	0.8714	1.6680
5			
6			
7			
8			
9			
10			

In-Sample: 1 Jan 1999 to 31 Dec 2003
Out-of-Sample: 2004

With Regression Constant:

		In	Out
1	***BVTP***	0.8016	0.8538
2	MOM12	1.0095	1.1546
3	***DEBTTBVE***	1.0445	1.3505
4	***CFTP***	1.0781	1.3184
5	***EY***	1.1239	1.1776
6	CFTDEBT	1.2480	1.1303
7	*TATURNOVR	1.2075	1.1234
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.8016	0.8538
2	***MOM12-1***	1.0107	1.1314
3	***DEBTTBVE***	1.0383	1.3175
4	***CFTP***	1.0511	1.2633
5	ROA	1.1015	1.1918
6	TATURNOVR	1.1222	1.2170
7	***EY***	1.1338	1.1834
8	*DEBTTMVE	1.1329	1.2620
9			
10			

In-Sample: 1 Jan 2000 to 31 Dec 2004
Out-of-Sample: 2005

With Regression Constant:

		In	Out
1	***BVTP***	0.8757	1.1118
2	***MOM12-1***	1.1888	0.8084
3	***DEBTTBVE***	1.2857	1.0005
4	**DEBTTMVE	1.3005	0.8752
5	****LAGLPRICE***	1.2841	0.8340
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.8757	1.1118
2	***MOM12-1***	1.1950	0.7943
3	***DEBTTBVE***	1.2760	0.8977
4	**DEBTTMVE	1.3142	0.7620
5	****LAGLPRICE***	1.2722	0.7546
6			
7			
8			
9			
10			

In-Sample: 1 Jan 2001 to 31 Dec 2005
Out-of-Sample: 2006

With Regression Constant:

		In	Out
1	BVTP	0.8623	0.4015
2	***MOM12-1***	1.1445	0.7906
3	GROWTH	1.3161	0.9395
4	***CFTP***	1.4066	1.3674
5	MOM1	1.4366	1.7113
6	ICBT	1.4623	1.7460
7	G12MCPS	1.4826	1.6951
8	*MOM3	1.4806	1.5742
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	0.8623	0.4015
2	***MOM12-1***	1.1489	0.7897
3	***DEBTTBVE***	1.2405	0.7483
4	***CFTP***	1.3310	0.5570
5	****LPRICE***	1.3281	0.5284
6			
7			
8			
9			
10			

In-Sample: 1 Jan 2002 to 31 Dec 2006
Out-of-Sample: 2007

With Regression Constant:

		In	Out
1	***LPRICE***	0.7132	0.3307
2	***G12MNPMARGIN***	0.7562	0.1919
3	***CFTP***	0.9288	0.7295
4	***BVTP***	1.0170	0.6268
5	***DEBTTBVE***	1.0934	0.7288
6	MOM1	1.1630	1.2053
7	***MOM24***	1.1924	1.4822
8	***MOM12***	1.2251	1.6466
9	***MOM3***	1.2665	1.6964
10	*TATURNOVR	1.2449	1.6987

Without Regression Constant:

		In	Out
1	***LPRICE***	0.7132	0.3307
2	***G12MNPMARGIN***	0.7577	0.1918
3	***CFTP***	0.8951	0.6414
4	MOM1	0.9461	1.2398
5	MOM12	1.0183	1.8205
6	GROWTH	1.0610	1.2379
7	G12MCPS	1.0888	1.1655
8	MOM3	1.1267	1.1966
9	****MOM24***	1.1246	1.3797
10			

Appendix B.4 Slope *t*- Statistic Model

Appendix B.4 demonstrates the order in which each of the style attributes are included in the slope *t* -static model in each of the sub-periods. The performance scores for the models with and without a regression constant are displayed in the top and bottom panels of the table respectively. The style attributes with significant mean factor payoffs are in italics and highlighted in bold. The style attributes assigned with a single asterisk are those being eliminated in the model due to the fact that they do not contribute to the improvements of the respective performance scores. On the other hand, the style attributes assigned with double asterisks are those attributes that are highly correlated with a correlation coefficient of greater than 0.70 or less than -0.70 to one or more of the attributes in the prior iterations.

In-Sample: 1 Jan 1997 to 31 Dec 2001
Out-of-Sample: 2002

With Regression Constant:

		In	Out
1	EY	3.5741	1.1069
2	***MOM12-1***	5.2957	0.8613
3	***CFTP***	6.5400	3.9841
4	***MOM1***	6.6670	2.8962
5	****LAGLPRICE***	6.5610	2.8404
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***EY***	3.7445	1.0953
2	***MOM12-1***	5.1763	0.7830
3	***CFTP***	6.5954	4.2024
4	MOM12	6.6985	2.6430
5	****LAGLPRICE***	6.6477	2.6161
6			
7			
8			
9			
10			

In-Sample: 1 Jan 1998 to 31 Dec 2002
Out-of-Sample: 2003

With Regression Constant:

		In	Out
1	FOREY2	3.6970	-1.3705
2	CURRENTRATIO	3.7277	-0.8149
3	***CFTP***	5.7766	2.9686
4	***BVTP***	6.0671	3.4076
5	*TATURNOVR	6.0121	2.1183
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	FOREY2	3.5443	1.1225
2	CURRENTRATIO	3.6524	0.7010
3	***CFTP***	5.3377	3.6247
4	***BVTP***	6.6558	3.0398
5	TATURNOVR	6.8253	2.9850
6	****DEBTTBVE***	6.1813	3.2092
7			
8			
9			
10			

In-Sample: 1 Jan 1999 to 31 Dec 2003
Out-of-Sample: 2004

With Regression Constant:

		In	Out
1	***BVTP***	4.9455	3.0453
2	***MOM12-1***	6.1820	3.6518
3	***SALESTP***	6.6801	3.8898
4	***CFTP***	7.1391	4.0281
5	***EY***	7.9872	4.0283
6	TATURNOVR	8.2272	4.0775
7	****DEBTTBVE***	7.9761	4.1484
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	4.9323	3.0350
2	***MOM12-1***	6.1350	3.5872
3	***SALESTP***	6.8125	3.6932
4	***CFTP***	7.1630	3.7195
5	ROA	7.8376	3.9207
6	****EY***	7.8141	3.7391
7			
8			
9			
10			

In-Sample: 1 Jan 2000 to 31 Dec 2004
Out-of-Sample: 2005

With Regression Constant:

		In	Out
1	***BVTP***	5.5341	3.3786
2	***G12MDPS***	8.0234	2.1066
3	***SALESTP***	8.0513	2.1512
4	*DEBTTMVE	7.9714	1.4982
5			
6			
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	5.5443	3.3586
2	***MOM12-1***	7.2821	2.7489
3	***DEBTTBVE***	7.6182	3.0482
4	**DEBTTMVE	8.2467	2.6593
5	MOM3	8.7118	2.7335
6	***LSIZE***	8.8759	1.7888
7	*****MOM6***	8.9970	1.8688
8	****LAGLPRICE***	8.5164	1.8323
9			
10			

In-Sample: 1 Jan 2001 to 31 Dec 2005
Out-of-Sample: 2006

With Regression Constant:

		In	Out
1	***BVTP***	6.1666	1.0277
2	***MOM12***	7.6327	2.4129
3	GROWTH	8.2291	2.4180
4	***CFTP***	8.8478	4.4060
5	MOM1	9.4003	4.8776
6	***LPRICE***	9.5930	4.8210
7	MOM3	9.8246	4.5163
8	*ICBT	9.8191	4.4958
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	6.1710	1.0200
2	G12MCPS	7.6250	1.7262
3	***CFTP***	8.8107	1.7210
4	TATURNOVR	9.2626	1.8754
5	***DEBTTBVE***	9.5005	1.6944
6	****SALESTP***	9.2830	1.5919
7			
8			
9			
10			

In-Sample: 1 Jan 2002 to 31 Dec 2006
Out-of-Sample: 2007

With Regression Constant:

		In	Out
1	***BVTP***	4.5921	-0.9815
2	G12MEPS	6.9009	0.6846
3	G12MCPS	7.9254	0.0520
4	***CFTP***	8.9107	2.2939
5	***DEBTTBVE***	8.9473	2.7897
6	*MOM1	8.8900	3.3012
7			
8			
9			
10			

Without Regression Constant:

		In	Out
1	***BVTP***	4.6194	1.0819
2	***G12MEPS***	6.7652	0.5500
3	G12MCPS	7.7433	0.3280
4	****MOM12-1***	7.6043	1.1141
5			
6			
7			
8			
9			
10			

Appendix C Performances of the Backpropagation Models (*t*-Statistic > 1.0)

Variable Selection Pool: Attributes with greater than 1.0 for the absolute values of their mean factor payoffs in each sub-period are considered for inclusion in the models

Table C.1 Periodic Variable Selections of the Backpropagation Models

Table C.1 documents the periodic variable selection process of the backpropagation models developed for each of the six overlapping in-sample periods over the period from 1 January 1997 to 31 December 2007. The backpropagation models are divided into models with or without using the cascade pre-selection procedure in addition to the existing standard variable selection procedure. When a variable is included by the algorithm as a selected input in the neural network, it is assigned with a letter "I". By contrast, variables failed to pass the standard variable selection procedure is assigned with a letter "V". For the models that pre-select variables using the cascade variable selection technique, variables eliminated in the pre-selection step are assigned with a letter "C" without going through the standard variable selection procedure. Variables with less than 1.0 for the absolute values of their mean factor payoffs in the respective sub-periods are left with empty cells.

In-Sample Period:	1/1/97 ~ 31/12/01		1/1/98 ~ 31/12/02		1/1/99 ~31/12/03		1/1/00 ~ 31/12/04		1/1/01 ~ 31/12/05		1/1/02 ~ 31/12/06	
	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection
BVTP	I	I	I	I	I	I	I	I	I	I	I	I
CFTCURRLIABS												
CFTDEBT							V	V				
CFTP	I	I	I	I	I	I	I	I	I	I	I	I
CURRENTRATIO							I	V				
DEBTTBVE			V	C	V	C	V	C	V	C	V	C
DEBTTMVE	I	I			V	V	V	C				
DY	I	V	I	I	V	V	V	I	I	I	V	V
EARNREV	V	V	V	V	V	V	V	V	V	V	V	V
EG1					V	I	V	I	V	V	I	C
EY			I	I	V	V	I	V	V	V	V	V
FOREY1	I	V	I	C	V	C	V	V	V	I	I	I
FOREY2	V	V	V	V	V	C	V	C	V	V	I	V
G12MCPS			V	C			I	C	I	V		
G12MDPS	I	I	I	I	I	V	I	V	I	I	V	C
G12MEPS	V	C					V	C	I	V	I	V
G12MGPMARGIN												
G12MNPMARGIN	V	C	V	C			V	I	I	V	I	I
G12MSALES							V	C	I	C	V	V
G24MEPS							V	C	V	C	V	V
GPMARGIN												
GROWTH			V	C	V	C			V	C	V	V
ICBT							V	C	V	C	V	V
LAGLPRICE	V	C	V	V	V	I	I	V	V	I	V	V
LPRICE	V	C	V	I	V	C	I	V	V	V	I	V
LSIZE	I	C	V	V	V	V	V	I	I	V	I	I
MOM1	I	I	I	V	I	V						
MOM12	V	V	V	I	I	I	I	V	I	V	V	I
MOM12-1	I	I	I	I	I	I	I	I	I	V	I	I
MOM24	V	I	I	I	V	I	I	I	I	I	I	I
MOM3												
MOM6	V	V			V	I	V	I	I	V	V	V
NPMARGIN					V	V			V	V		
PAYOUT			V	V	I	I					V	V
ROA					V	C	V	C	I	C	I	V
ROE	V	V	V	V	V	C						
SALESTP	I	C	V	C	V	C	V	C	V	C	V	C
TATURNOVR					I	C	V	V	I	V	V	V
Total:	**10**	**7**	**9**	**9**	**8**	**9**	**11**	**9**	**15**	**7**	**12**	**8**

Figure C.1 In-Sample Scores

CHART (A) to CHART (D) illustrate the histograms of the in-sample performance scores for the ANN models (backpropagation training rule) under each of the four performance measurement criteria. The performance scores of the ANN models in each chart are represented by the red, yellow, blue and green bars respectively to enhance direct comparisons of the periodic in-sample performance of the non-linear expected return multifactor models.

CHART (A) In-Sample Average IC Scores

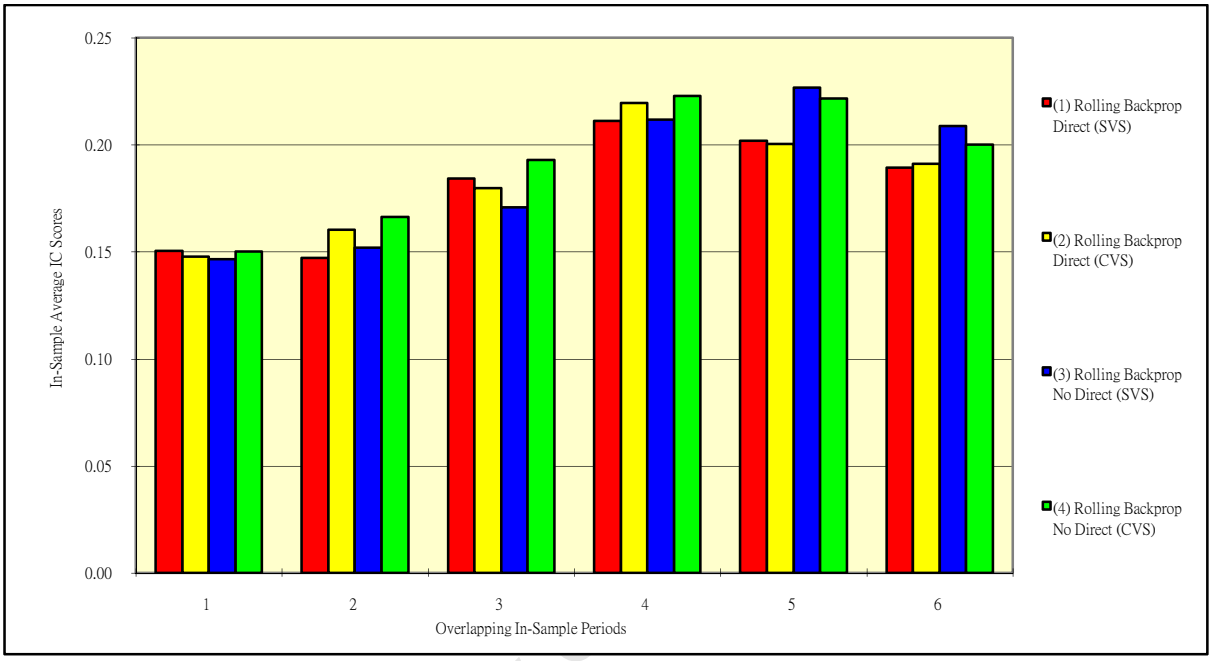


CHART (B) In-Sample Grinold IR Scores

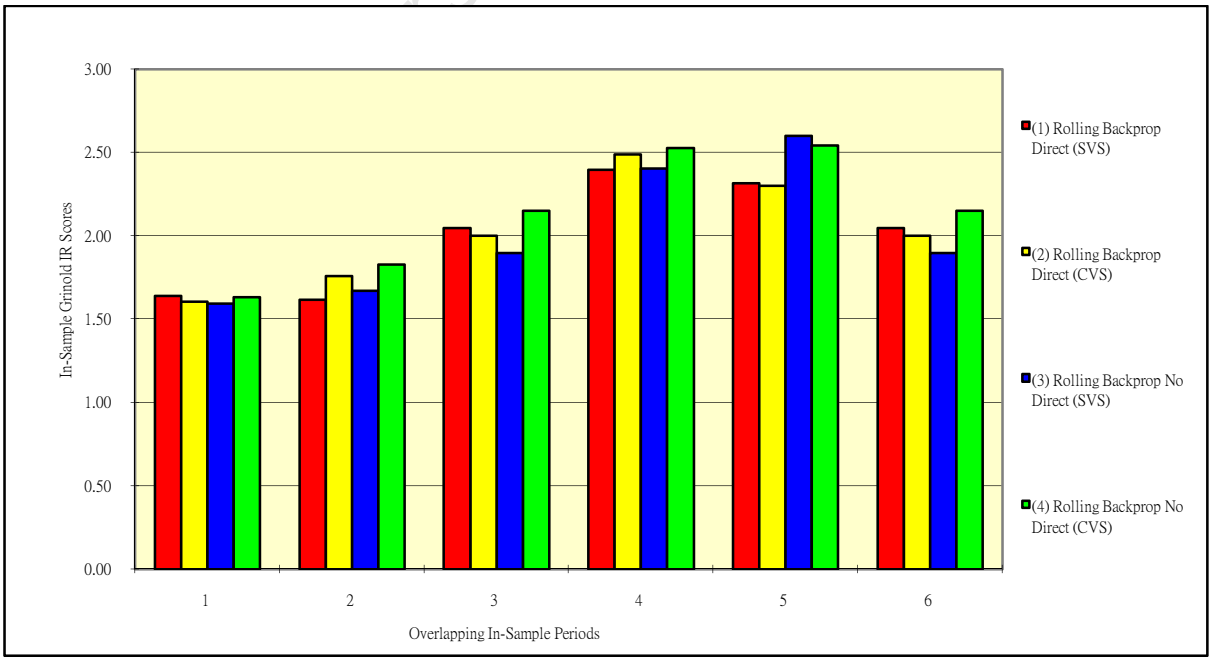


Figure C.1 In-Sample Scores - Continued

CHART (C) In-Sample Qian and Hua IR Scores

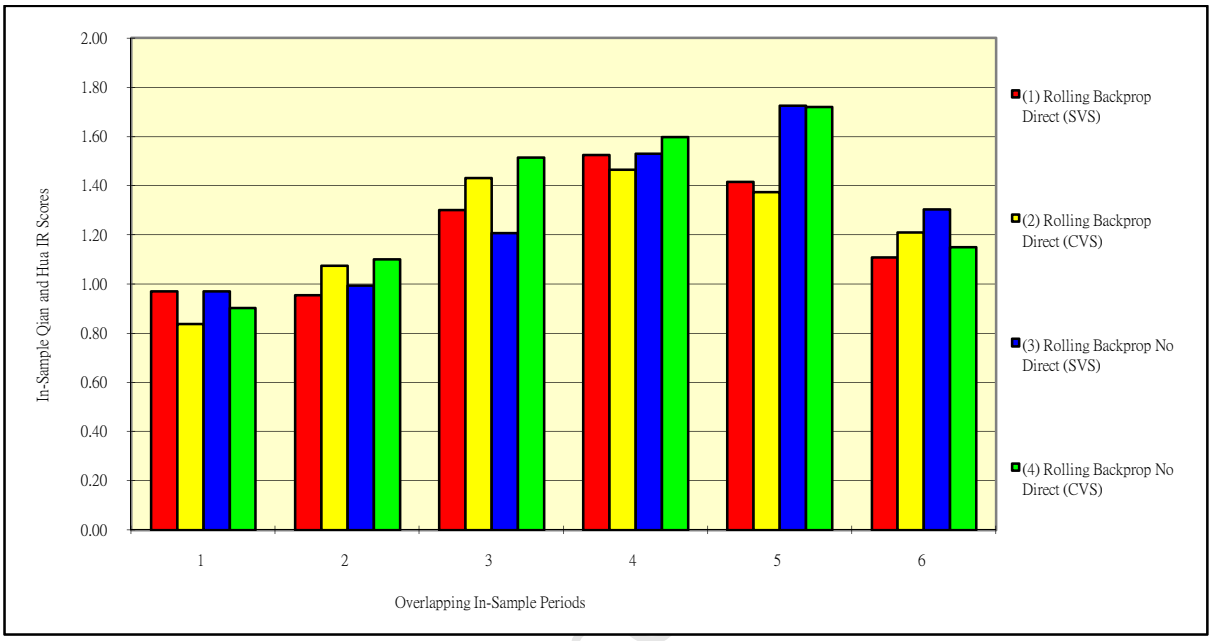


CHART (D) In-Sample Slope *t*-Statistic Scores

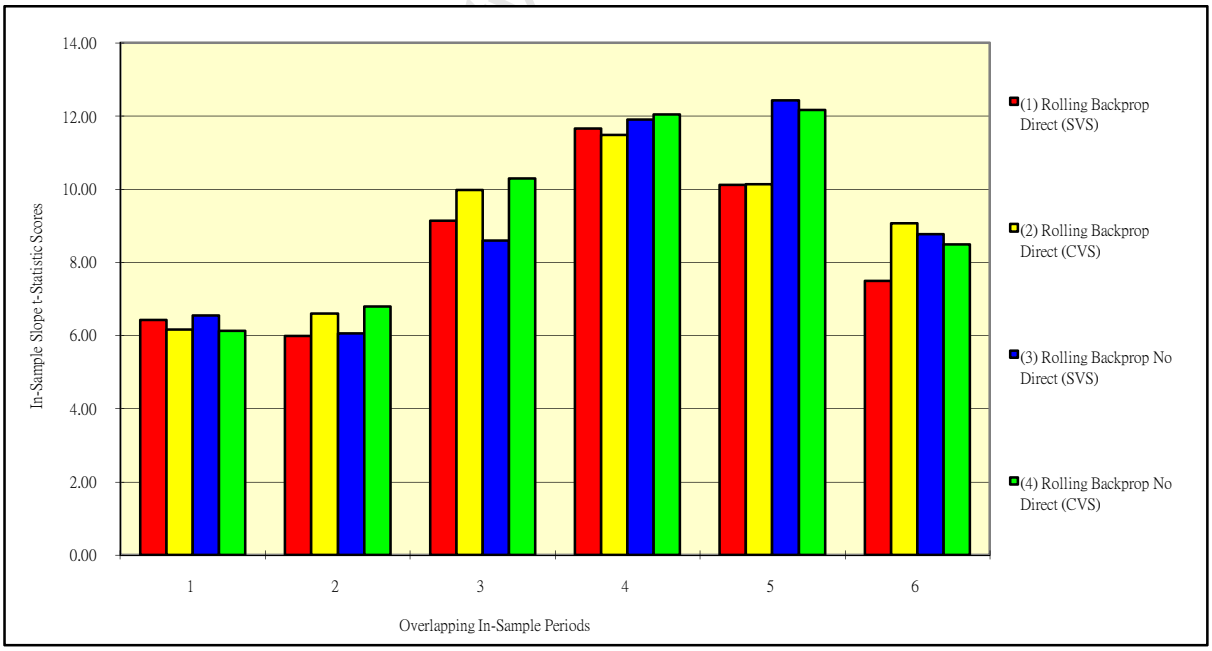


Figure C.2 Out-of-Sample Scores

CHART (A) to CHART (D) illustrate the periodic out-of-sample performance of the ANN (backpropagation) models under each of the performance measurement criteria. The performance scores of the models constructed under the moving-window procedure that updates the model inputs every 12 months are represented by the histograms. On the other hand, the performance scores of the models trained within the static period from 01January 1997 to 31 December 2001 are represented by the trend lines. The ANN models in each chart are highlighted in red, yellow, blue and green respectively to enhance the direct comparisons of the respective performance scores of the pre-specified models.

CHART (A) Out-of-Sample Average IC Scores

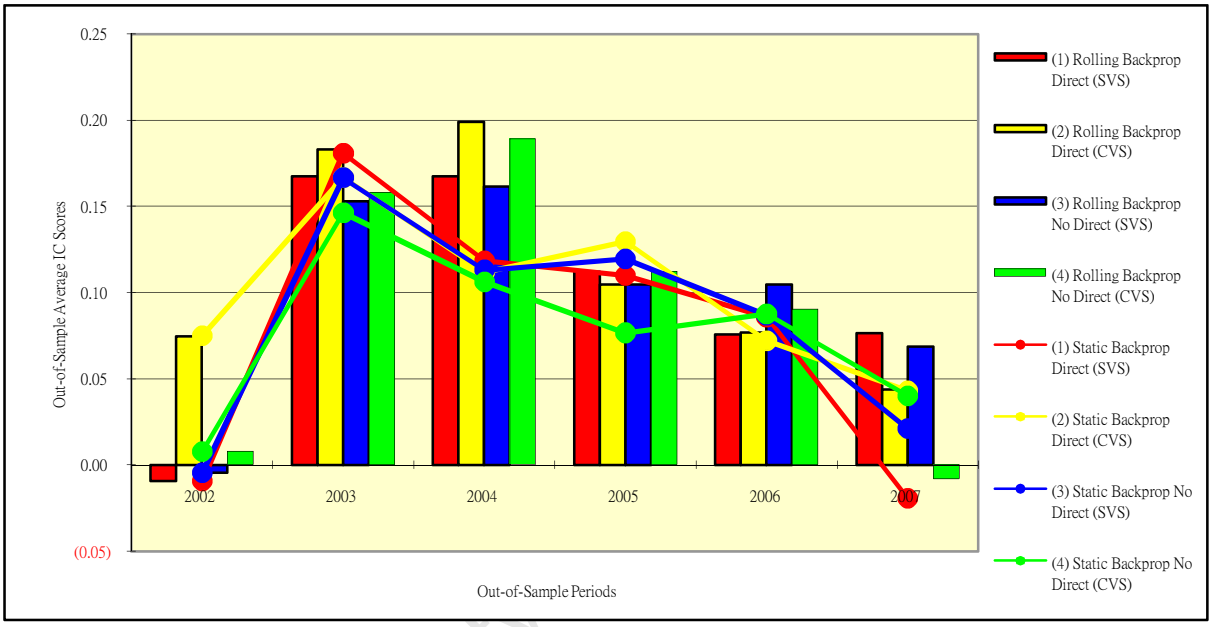


CHART (B) Out-of-Sample Grinold IR Scores

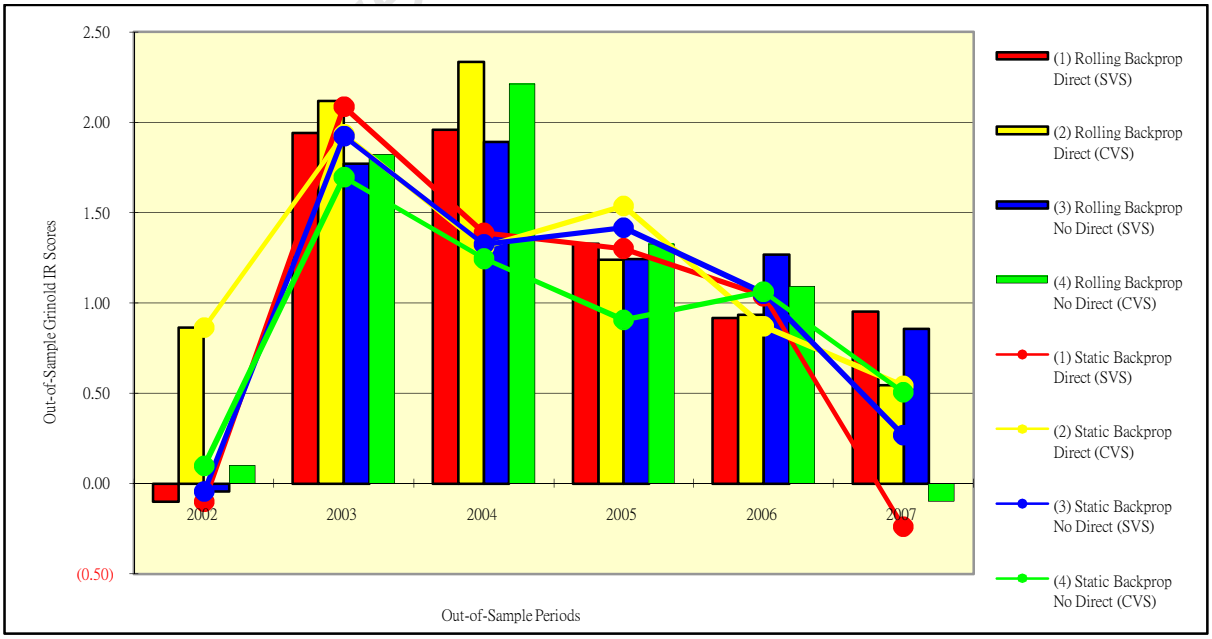


Figure C.2 Out-of-Sample Scores - Continued

CHART (C) Out-of-Sample Qian and Hua IR Scores

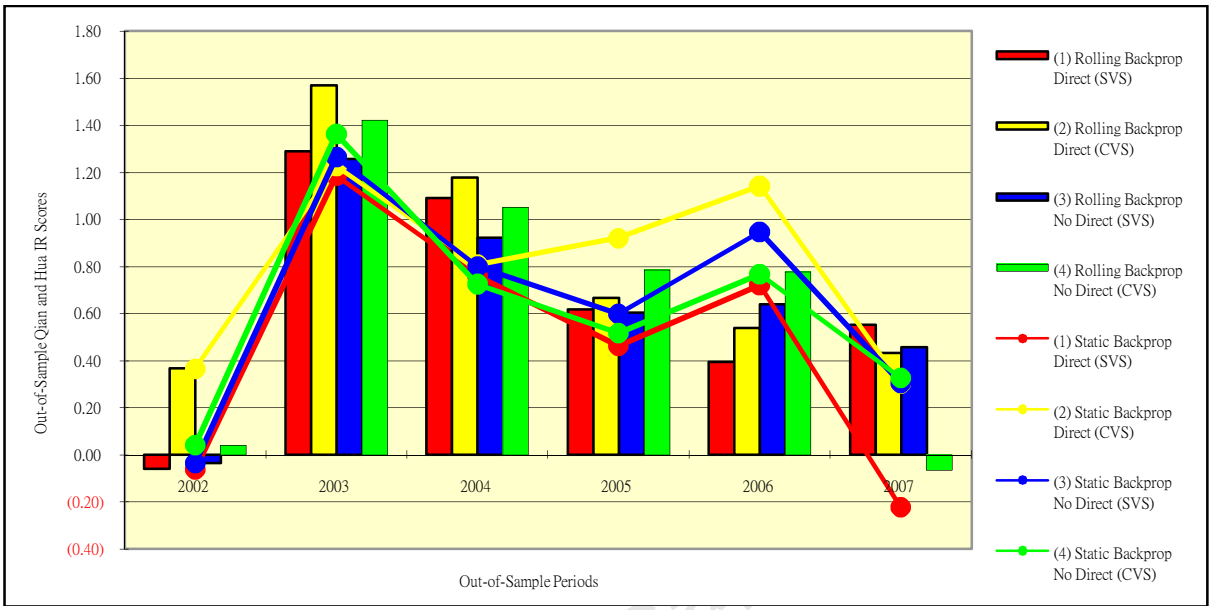


CHART (D) Out-of-Sample Slope *t*-Statistic Scores

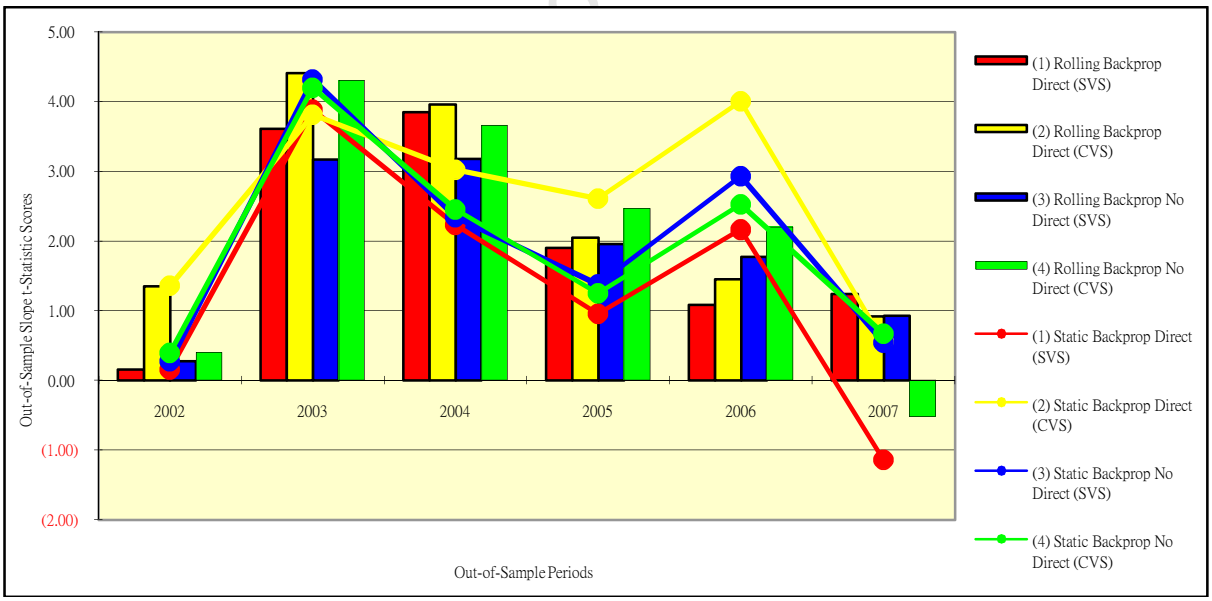


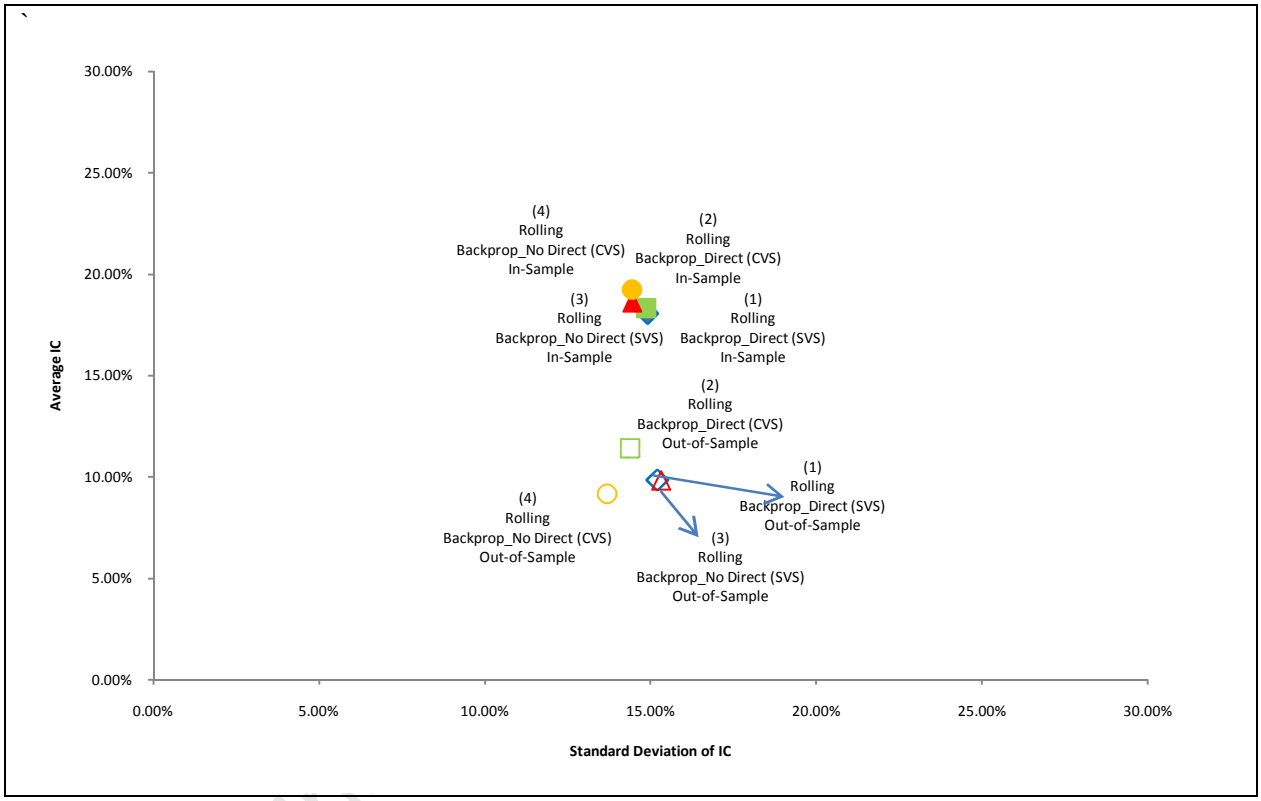
Table C.2 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Backpropagation Models

This table contains the average periodic in-sample and the corresponding out-of-sample performance scores of the rolling backpropagation models developed in this chapter. The models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and output. The backpropagation models applying the cascade pre-selection technique are assigned with CVS. On the other hand, the models applying standard variable selection technique are assigned with SVS. The best out-of-sample score is highlighted in bold.

	(1) Rolling Backprop Direct (SVS)	(2) Rolling Backprop Direct (CVS)	(3) Rolling Backprop No Direct (SVS)	(4) Rolling Backprop No Direct (CVS)
Avg IC (In-Sample)	0.1806	0.1831	0.1861	0.1923
Avg IC (Out-Sample)	0.0984	**0.1138**	0.0982	0.0916
Grinold's IR (In-Sample)	2.0107	2.0265	2.0106	2.1390
Grinold's IR (Out-Sample)	1.1661	**1.3386**	1.1639	1.0754
Qian and Hua's IR (In-Sample)	1.2114	1.2313	1.2879	1.3305
Qian and Hua's IR (Out-Sample)	0.6467	**0.7912**	0.6408	0.6690
Slope t-Statistic (In-Sample)	8.4692	8.9068	9.0518	9.3171
Slope t-Statistic (Out-Sample)	1.9767	**2.3599**	1.8834	2.0825

Figure C.3 Consolidated In-Sample and Out-of-Sample Performance Scores for the Backpropagation Models

Figure C.3 plots the moving-window backpropagation models in the diagram based on their respective consolidated average IC scores and standard deviation of IC for the in-sample and out-of-sample periods. The slope of the line connecting the origin to the position of the model in the diagram is essentially the model's Qian and Hua IR score. The filled labels represent in-sample scores while the corresponding out-of-sample scores are represented by empty labels. The four backpropagation models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and output. The backpropagation models with cascade variable selection are assigned with CVS. On the other hand, the models applying standard variable selection technique are assigned with SVS.



Appendix D Performances of the Kalman Models (*t*-Statistic > 1.0)

Variable Selection Pool: Attributes with greater than 1.0 for the absolute values of their mean factor payoffs in each sub-period are considered for inclusion in the models

Table D.1 Periodic Variable Selections of the Kalman Models

Table D.1 documents the periodic variable selection process of the Kalman models developed for each of the six overlapping in-sample periods over the period from 1 January 1997 to 31 December 2007. The Kalman models are divided into models with or without using the cascade pre-selection procedure in addition to the existing standard variable selection procedure. When a variable is included by the algorithm as a selected input in the neural network, it is assigned with a letter "I". By contrast, variables failed to pass the standard variable selection procedure is assigned with a letter "V". For the models that pre-select variables using the cascade variable selection technique, variables eliminated in the pre-selection step are assigned with a letter "C" without going through the standard variable selection procedure. Variables with less than 1.0 for the absolute values of their mean factor payoffs in the respective sub-periods are left with empty cells.

In-Sample Period:	1/1/97 ~ 31/12/01		1/1/98 ~ 31/12/02		1/1/99 ~31/12/03		1/1/00 ~ 31/12/04		1/1/01 ~ 31/12/05		1/1/02 ~ 31/12/06	
	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection	Standard Variable Selection	Cascade Pre-Selection
BVTP	V	I	I	I	I	V	I	I	I	I	V	V
CFTCURRLIABS												
CFTDEBT							V	V				
CFTP	I	I	I	I	I	I	I	V	V	I	I	I
CURRENTRATIO							V	V				
DEBTTBVE			V	V	V	I	I	V	V	V	V	C
DEBTTMVE	V	V			V	V	V	C				
DY	V	I	I	I	V	V	V	I	V	C	V	C
EARNREV	V	V	V	V	V	V	V	V	V	V	V	V
EG1					V	V	V	V	I	V	V	C
EY			V	V	V	V	V	V	I	C	V	V
FOREY1	V	V	I	C	V	I	V	C	V	I	V	V
FOREY2	V	V	V	V	V	I	V	C	V	V	I	C
G12MCPS			V	V			V	C	V	C		
G12MDPS	I	I	I	V	V	V	V	C	V	C	V	V
G12MEPS	I	V					V	V	I	I	V	V
G12MGPMARGIN												
G12MNPMARGIN	V	I	V	C			V	V	V	V	I	I
G12MSALES							V	V	V	C	V	V
G24MEPS							I	V	V	C	V	C
GPMARGIN												
GROWTH			V	V	V	V			V	C	V	V
ICBT							V	V	V	V	V	C
LAGLPRICE	V	V	V	V	V	V	V	C	I	V	V	C
LPRICE	V	C	V	V	V	V	V	C	V	V	V	V
LSIZE	I	V	V	C	V	V	V	V	V	V	I	I
MOM1	I	I	V	I	V	C						
MOM12	V	V	V	I	V	I	V	C	V	V	V	V
MOM12-1	V	V	V	V	V	I	I	V	V	V	V	V
MOM24	I	I	V	V	I	V	V	I	V	V	V	C
MOM3												
MOM6	V	V			V	I	V	I	I	V	V	V
NPMARGIN					V	V			I	I		
PAYOUT			I	V	I	V					V	V
ROA					I	V	V	V	V	C	V	V
ROE	V	V	V	C	V	I						
SALESTP	V	V	V	V	V	V	V	V	V	V	V	C
TATURNOVR					V	V	V	C	V	V	V	I
Total:	**6**	**7**	**6**	**5**	**5**	**8**	**5**	**4**	**7**	**5**	**4**	**4**

Figure D.1 In-Sample Scores

CHART (A) to CHART (D) illustrate the histograms of the in-sample performance scores for the ANN models (Kalman training rule) under each of the four performance measurement criteria. The performance scores of the ANN models in each chart are represented by the red, yellow, blue and green bars respectively to enhance direct comparisons of the periodic in-sample performance of the non-linear expected return multifactor models.

CHART (A) In-Sample Average IC Scores

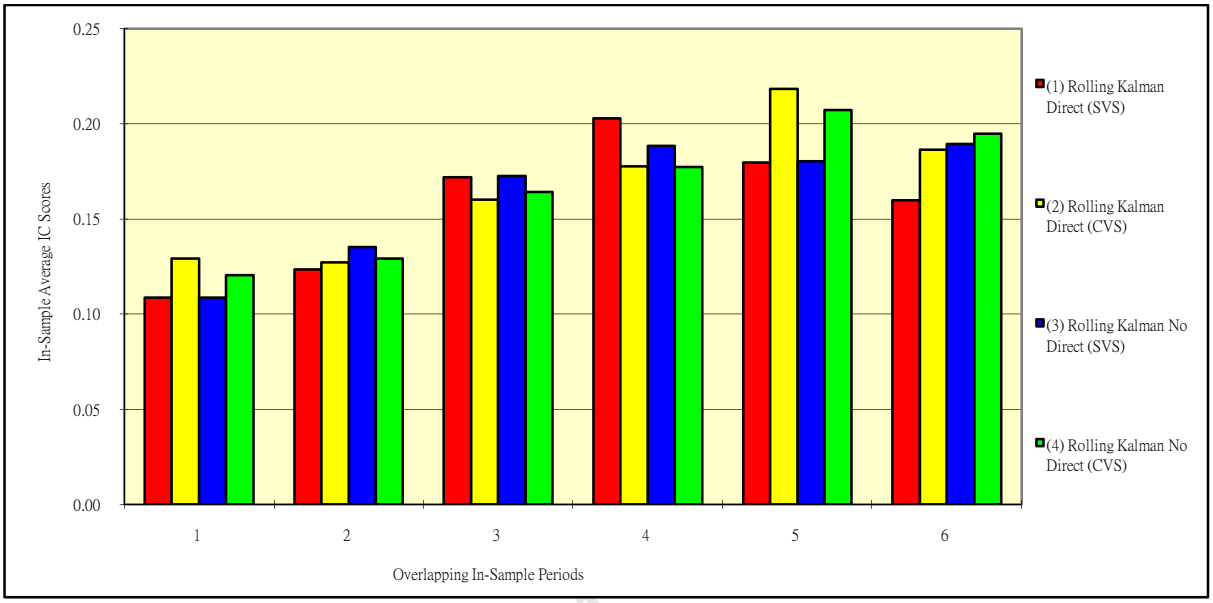


CHART (B) In-Sample Grinold IR Scores

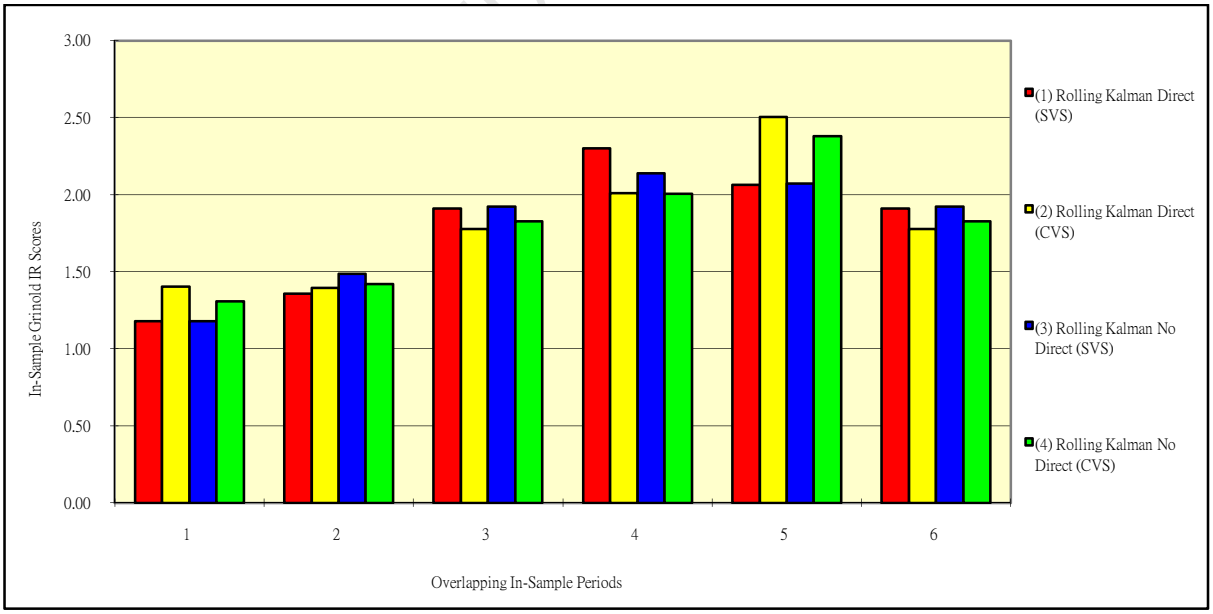


Figure D.1 In-Sample Scores - Continued

CHART (C) In-Sample Qian and Hua IR Scores

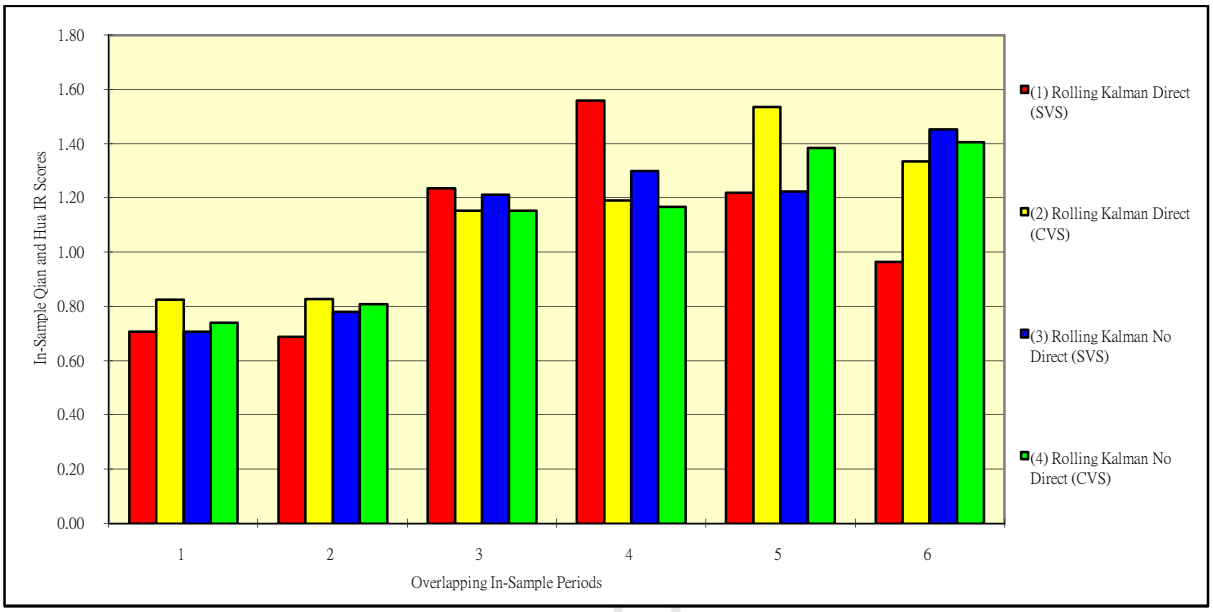


CHART (D) In-Sample Slope *t*-Statistic Scores

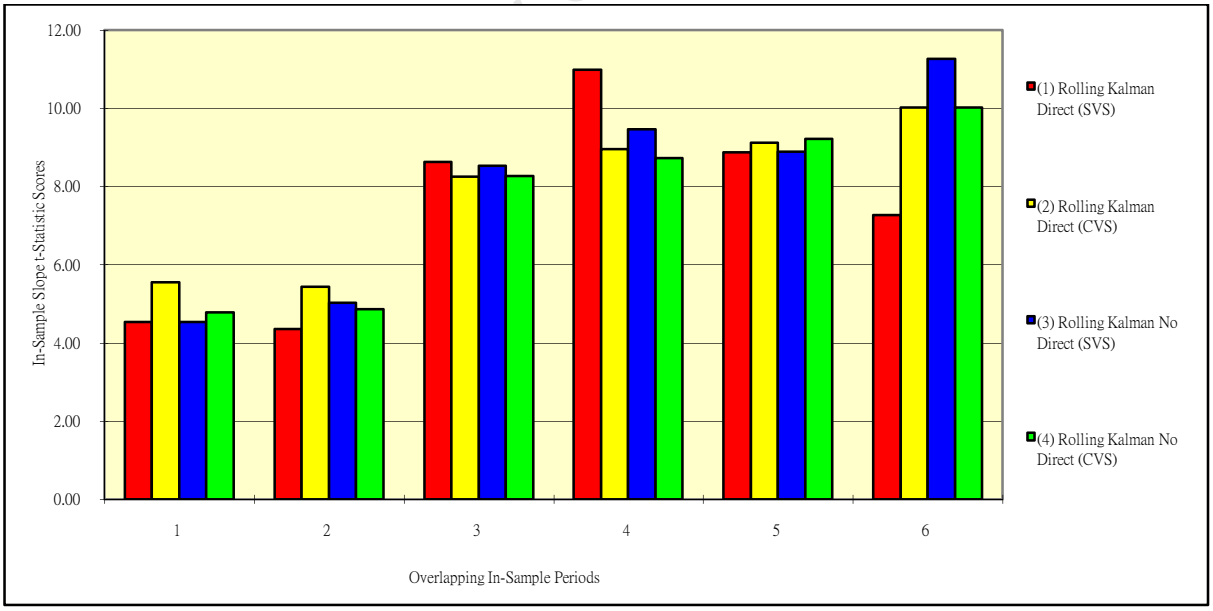


Figure D.2 Out-of-Sample Scores

CHART (A) to CHART (D) illustrate the periodic out-of-sample performance of the ANN (Kalman) models under each of the performance measurement criteria. The performance scores of the models constructed under the moving-window procedure that updates the model inputs every 12 months are represented by the histograms. On the other hand, the performance scores of the models trained within the static period from 01January 1997 to 31 December 2001 are represented by the trend lines. The ANN models in each chart are highlighted in red, yellow, blue and green respectively to enhance the direct comparisons of the respective performance scores of the pre-specified models.

CHART (A) Out-of-Sample Average IC Scores

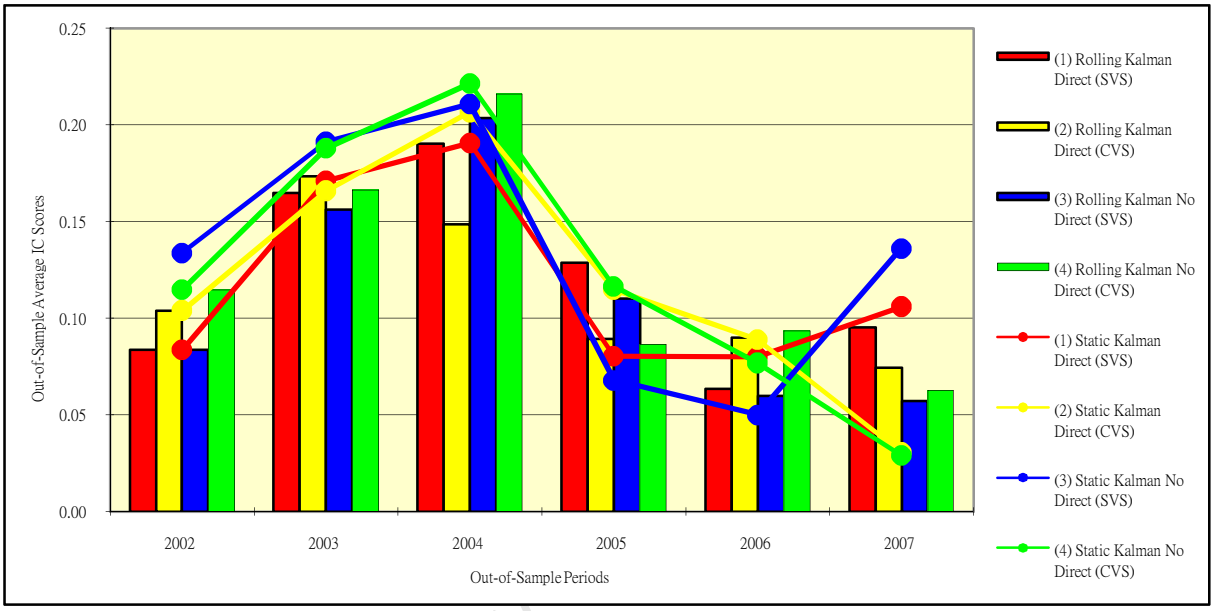


CHART (B) Out-of-Sample Grinold IR Scores

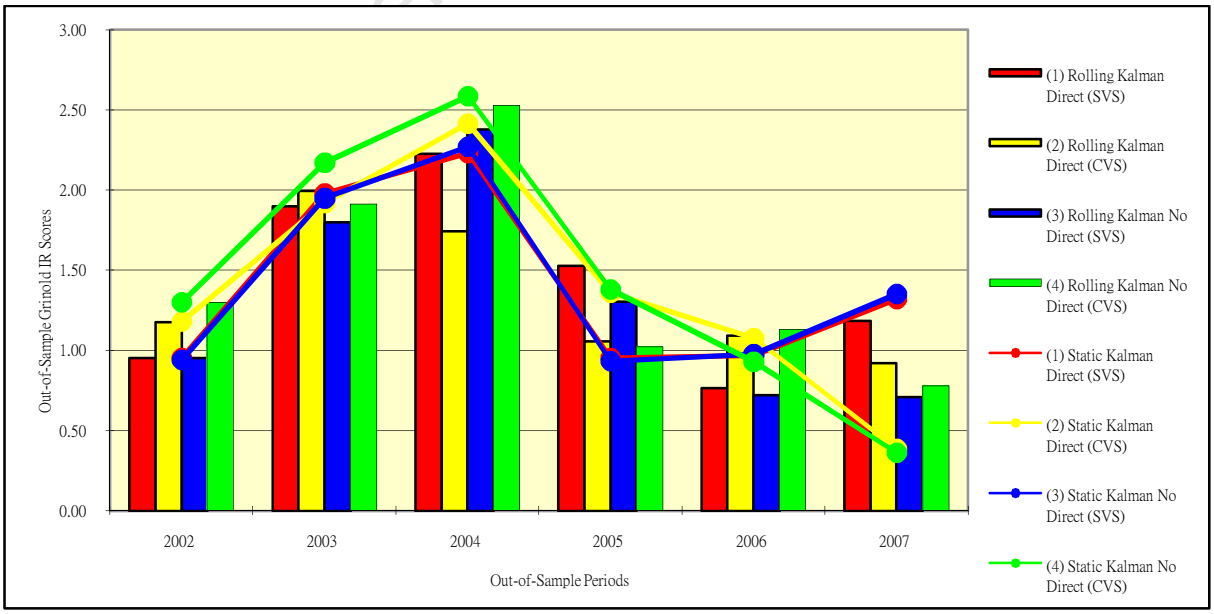


Figure D.2 Out-of-Sample Scores - Continued

CHART (C) Out-of-Sample Qian and Hua IR Scores

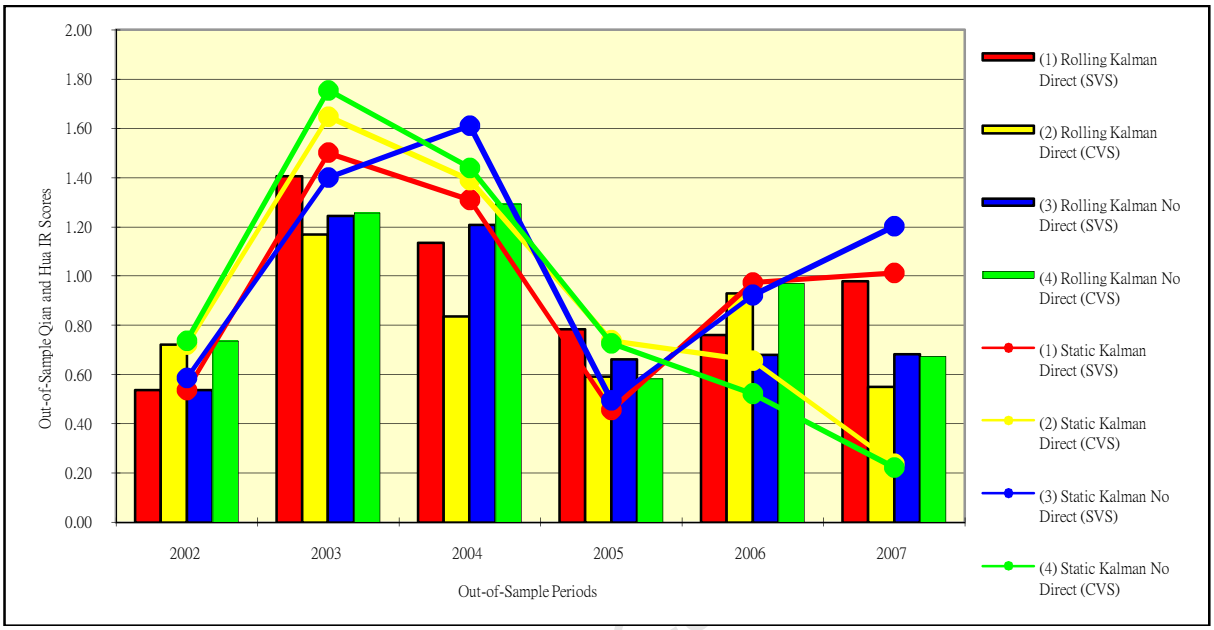


CHART (D) Out-of-Sample Slope *t*-Statistic Scores

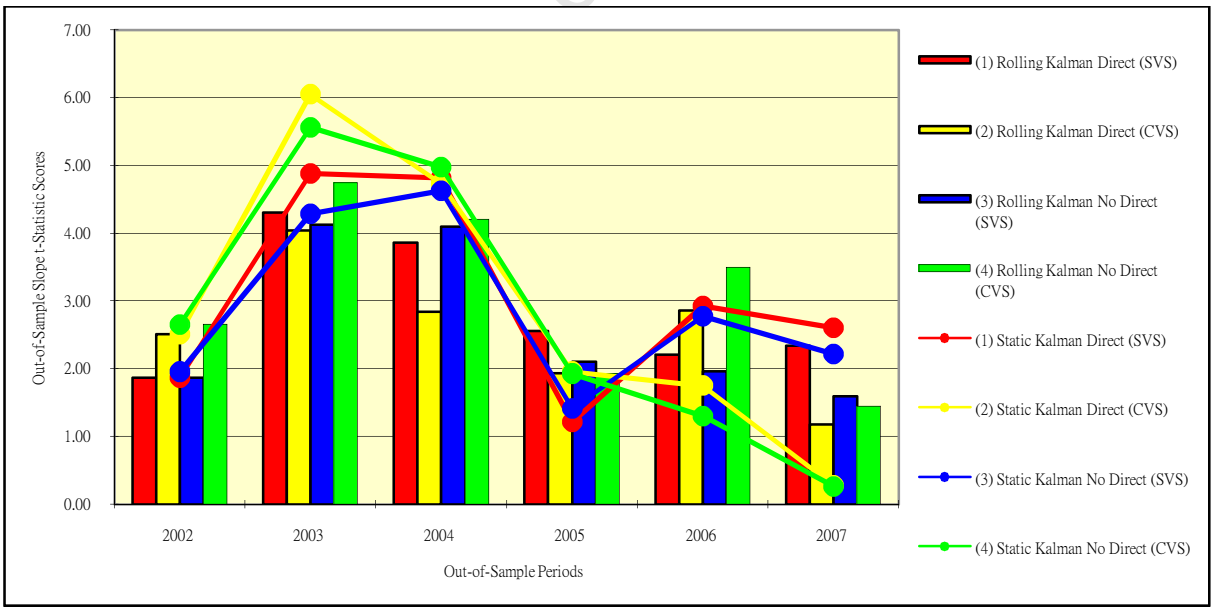


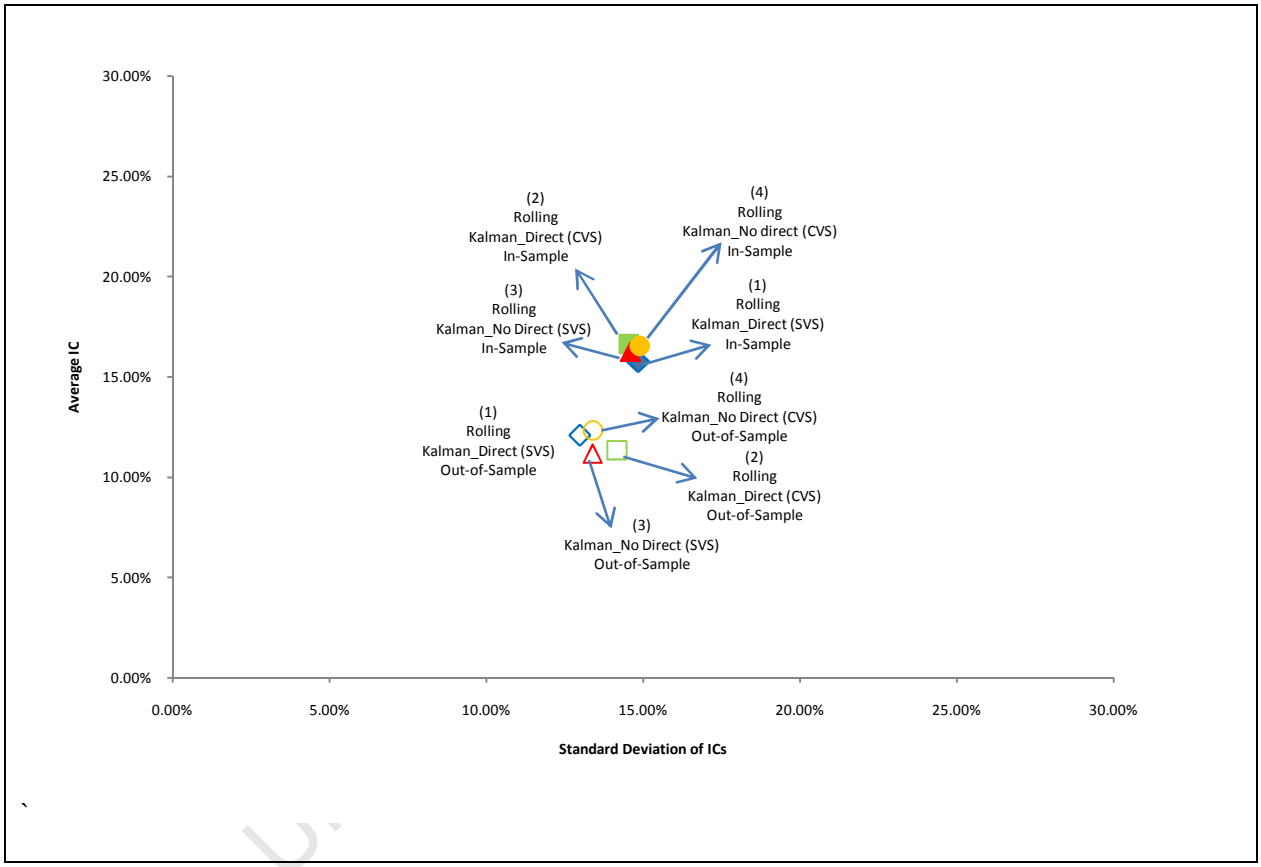
Table D.2 Consolidated In-Sample and Out-of-Sample Performance Scores for the Rolling Kalman Models

This table contains the average periodic in-sample and the corresponding out-of-sample performance scores of the rolling Kalman models developed in this chapter. The models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and output. The Kalman models applying the cascade pre-selection technique are assigned with CVS. On the other hand, the models applying standard variable selection technique are assigned with SVS. The best out-of-sample score is highlighted in bold.

	(1) Rolling Kalman Direct (SVS)	(2) Rolling Kalman Direct (CVS)	(3) Rolling Kalman No Direct (SVS)	(4) Rolling Kalman No Direct (CVS)
Avg IC (In-Sample)	0.1578	0.1665	0.1625	0.1655
Avg IC (Out-Sample)	0.1211	0.1134	0.1118	**0.1232**
Grinold's IR (In-Sample)	1.7883	1.8117	1.7872	1.7953
Grinold's IR (Out-Sample)	1.4252	1.3317	1.3111	**1.4453**
Qian and Hua's IR (In-Sample)	1.0628	1.1447	1.1130	1.1108
Qian and Hua's IR (Out-Sample)	**0.9336**	0.7998	0.8360	0.9188
Slope t-Statistic (In-Sample)	7.4368	7.8892	7.9477	7.6457
Slope t-Statistic (Out-Sample)	2.8583	2.5632	2.6286	**3.0780**

Figure D.3 Consolidated In-Sample and Out-of-Sample Performance Scores for the Kalman Models

Figure D.3 plots the moving-window Kalman models in the diagram based on their respective consolidated average IC scores and standard deviation of IC for the in-sample and out-of-sample periods. The slope of the line connecting the origin to the position of the model in the diagram is essentially the model's Qian and Hua IR score. The filled labels represent in-sample scores while the corresponding out-of-sample scores are represented by empty labels. The four Kalman models are distinguished by their variable selection method as to whether they apply the cascade pre-selection technique, and whether there is a direct connection between their inputs and output. The Kalman models with cascade variable selection are assigned with CVS. On the other hand, the models applying standard variable selection technique are assigned with SVS.



Bibliography

Abhyankar A, Copeland L and W Wong (1997), "Uncovering Nonlinear Structure in Real-Time Stock-Market Indexes: The S&P 500, the DAX, the Nikkei 225, and the FTSE-100", Journal of Business and Economic Statistics, American Statistical Association, vol. 15, no 1, 1-14

Achour D, Harvey C R, Hopkins G and C Lang (1998), "Stock Selection in Emerging Markets: Portfolio Strategies for Malaysia, Mexico, and South Africa", Emerging Markets Quarterly, vol. 2, 38-91

Achour D, Harvey C R, Hopkins G and C Lang (1999), "Firm Characteristics and Investment Strategies in Africa: The Case of South Africa", The African Finance Journal, vol. 1, no 1, 1-67

Attew D (2002), "Artificial Neural Networks", In Neural Networks and the Financial Markets, Shadbolt J and J G Taylor (eds.), Springer, 87-93

Avramov D (2004), "Stock Return Predictability and Asset Pricing Models", The Review of Financial Studies, vol. 17, 699-738

Avramov D and T Chordia (2006), "Asset Pricing Models and Financial Market Anomalies", The Review of Financial Studies, vol. 19, no 3, 1001-1040

Avramov D and T Chordia (2006), "Predicting Stock Returns", Journal of Financial Economics, vol. 82, 387-415

Baestaens D, Van Den Bergh W and D Wood (1994), Neural Network Solutions for Trading in Financial Markets, Pitman Publishing, London

Ball R (1978), "Anomalies in Relationships Between Securities' Yields and Yield Surrogates", Journal of Financial Economics, vol. 6, no 2, 103-126

Ball R, Kothari S and J Shanken (1995), "Problems in Measuring Portfolio Performance: An Application of Contrarian Investment Strategies", Journal of Financial Economics, vol. 38, 79-107

Banz R W (1981), "The Relationship between Return and Market Value of Common Stocks", Journal of Financial Economics, vol. 9, 3-18

Barberis N and A Shleifer (2003), "Style Investing", Journal of Financial Economics, 161-199

Barberis N, Shleifer A and R Vishny (1998), "A Model of Investor Sentiment", Journal of Financial Economics, vol. 49, 307-343

Basu S (1977), "The Investment Performance of Common Stocks in Relation to Their Price-to-Earnings Ratio: A Test of the Efficient Markets Hypothesis", Journal of Finance, vol. 12, no 1, 129-156

Basu S (1983), "The Relationship Between Earnings' Yield, Market Value and The Returns for NYSE Common Stocks: Further Evidence", Journal of Financial Economics, vol. 32, no 3, 663-682

Bauman C, Conover C M and R E Miller (1998), "Growth versus Value and Large-Cap versus Small-Cap Stocks in International Markets", Financial Analysts Journal, vol. 54, no 2, 75-89

Bekaert G and C R Harvey (1995), "Time-varying World Market Integration", Journal of Finance, vol. 50, no 2, 401-444

Bekaert G and C R Harvey (1997), "Foreign Speculators and Emerging Equity Markets, Working Paper 1441, Graduate School of Business, Stanford University

Beltratti A, Margarita S and P Terna (1996), Neural Networks For Economic and Financial Modelling, International Thomson

Berghage T and G Orlander (2007), Predicting the Unpredictable: How Neural Networks Enhance Predictability and Performance in Portfolio Management, Acacia Publishing

Berk J B (2000), "Sorting Out Sorts", Journal of Finance, vol. 55, 407-427

Black F (1972), "Capital Market Equilibrium with Restricted Borrowing", Journal of Business, vol. 45, no 3, 444-455

Bhandari L C (1988), "Debt-Equity Ratio and Expected Common Stock Returns", Journal of Finance, vol. 43, no 2, 507-528

Black F, Jensen M C and M Scholes (1972), "The Capital Asset Pricing Model: Some Empirical Tests", Studies in the Theory of Capital Markets, Praeger Publisher, New York

Blume M and R Stambaugh (1983), "Biases in Computing Returns: An Application to the Size Effect", Journal of Financial Economics, vol. 28, 19-33

Bodie Z, Kane A and A J Marcus (2008), Investments, 7th Edition, McGraw Hill

Bolland P and J Connor (1997), "A Constrained Neural Network Kalman Filter for Price Estimation in High Frequency Financial Data", International Journal of Neural Systems, vol. 8, no 4, 399-415

Bradfield D J, Barr G D and J F Affleck-Graves (1988), "Asset Pricing in Small Markets-The South African Case", South African Journal of Business Management, vol. 19, no 1, 11-21

Brennan M T, Chordia and A Subrahmanyam (1998) "Alternative Factor Specification, Security Characteristics and the Cross-Section of Stock Returns", Journal of Financial Economics, vol. 49, 345-373

Brown P, Kleidon A W and T A Marsh (1983), "New Evidence on the Nature of Size-Related Anomalies in Stock Prices", Journal of Financial Economics, vol. 12, 33-56

Brown S and W Goetzmann (1995), "Performance Persistence", Journal of Finance, vol. 50, no 2, 679-698

Brown S, Goetzmann W and S Ross (1995), "Survival", Journal of Finance, vol. 50, no 3, 853-873

Buckle D (2004), "How to Calculate Breadth: An Evolution of the Fundamental Law of Active Portfolio Management", Journal of Asset Management, vol. 4, no 6, 393-405

Campbell J Y (2000), "Asset Pricing at the Millennium", Journal of Finance, vol. 55, 1515-1567

Cao Q, Leggio K and M Schiederjans (2005), "A Comparison between Fama and French's Model and Artificial Neural Networks in Predicting the Chinese Stock Market", Computers and Operations Research, vol. 32, 2499-2512

Cao Q, Parry M and B Leggio (2009), "The Three-Factor Model and Artificial Neural Networks: Predicting Stock Price Movement in China", Annals of Operations Research, Springer

Chan K C (1988), "On the Contrarian Investment Strategy", Journal of Business, vol. 61, no 2, 147-163

Chan K C, Hameed A and W Tong (2000), "Profitability of Momentum Strategies in the International Equity Markets", Journal of Financial and Quantitative Analysis, vol. 35, no 2, 153-172

Chan L K C and J Lakonishok (2004), "Value and Growth Investing: Review and Update", Financial Analysts Journal, vol. 60, no 1, 71-86

Chandra B and P Varghese (2007), "Applications of Cascade-Correlation Neural Networks for Cipher System Identification", World Academy of Science, Engineering and Technology, vol. 26, 311-314

Chen N, Roll R and S A Ross (1986), "Economic Forces and Stock Market", Journal of Business, vol. 59, no 3, 389-403

Chopra N, Lakonishok J and J R Ritter (1992), "Measuring Abnormal Performance – Do Stocks Overreact?", Journal of Financial Economics, vol. 31, 235-268

Claessens S, Dasgupta S and J Glen (1998), "The Cross-Section of Stock Returns: Evidence from the Emerging Markets", Emerging Markets Quarterly, vol. 2, 4-13

Clarke R, de Silva H and S Thorley (2002), "Portfolio Constraints and the Fundamental Law of Active Management", Financial Analysts Journal, vol. 58, 48-66

Clarke R, de Silva H and S Thorley (2006), "The Fundamental Law of Active Portfolio Management", Journal of Investment Management, vol. 4, no 3, 54-72

Coats P K and L F Fant (1993), "Recognizing Financial Distress Patterns Using a Neural Network Tool", Financial Management, 142-155

Conover W J (1999), Practical Nonparametric Statistics, Wiley

Conrad J, Cooper M and G Kaul (2003), "Value versus Glamour", Journal of Finance, vol. 58, no 5, 1969-1995

Cybenko G (1989), "Approximation by Superpositions of a Sigmoidal Function", Mathematics of Control, Signals and Systems, vol. 2, 303-314

Daniel K and S Titman (1997), "Evidence on the Characteristics of Cross Sectional Variation in Share Returns", Journal of Finance, vol. 52, no 1, 1-33

Daniel K, Titman S and K Wei (2001), "Explaining the Cross-Section of Stock Returns in Japan: Factors or Characteristics", Journal of Finance, vol. 56, 743-766

Daniel K, Hirshleifer D and A Subrahmanyam (1998), "Overconfidence, Arbitrage, and Equilibrium Asset Pricing", Journal of Finance, vol. 56, no 3, 921-965

Daniel K, Hirshleifer D and A Subrahmanyam (1998), "Investor Psychology and Security Market Under- and Overreactions", Journal of Finance, vol. 53, no 6, 1839-1885

Daniel K, Sheridan T and K C Wei (2001), "Explaining the Cross-Section of Stock Returns in Japan: Factors or Characteristics?", Journal of Finance, vol. 56, no 2, 743-766

Davis J L (1994), "The Cross-Section of Realised Stock Returns: The Pre-COMPUSTAT Evidence", Journal of Finance, vol. 49, no 5, 1579-1593

Davis J L, Fama E F and KR French (2000), "Characteristics, Covariances, and Average Returns: 1929 to 1997", Journal of Finance, vol. 55, no 1, 389-406

De Bondt W F M and R H Thaler (1985), "Does the Stock Market Overreact?", Journal of Finance, vol. 40, no 3, 793-805

De Bondt W F M and R H Thaler (1987), "Further Evidence on Investor Overreaction and Stock Market Seasonality", Journal of Finance, vol. 42, no 3, 557-581

De Villiers J, Lowlings A, Pettit A and J Affleck-Graves (1986), "An Investigation into the Small-Firm Effect on the JSE", South African Journal of Business Management, vol. 17, no 4, 191-195

DeFusco R A, McLeavey D W, Pinto J E and D E Runkle (2004), Quantitative Methods for Investment Analysis, 2nd Edition, CFA Institute

Diamantopoulou M J (2006), "Tree-Bole Estimation on Standing Pine Trees Using Cascade-Correlation Artificial Neural Network Models", Agricultural Engineering International: The GIGR Ejournal. Manuscript IT 06 002, vol. 8, 1-14

Dutta S and S Shekhar (1988), "Bond Rating: A Non-Conservative Application of Neural Networks", Proceedings of the IEEE International Conference on Neural Networks 2, 443-450

Eakins S and R Stansell (2003), "Can Value-Based Selection Criteria Yield Superior Risk-Adjusted Returns: An Application of Neural Networks", International Review of Financial Analysis, vol. 12, 83-97

Eakins S, Stansell R and J Buck (1998), "Analyzing the Nature of Institutional Demand for Common Stocks", Quarterly Journal of Business and Economics, vol. 37

Enke D and S Thawornwong (2005), "The Use of Data Mining and Neural Networks for Forecasting Stock Market Returns", Expert Systems with Applications, vol. 29, 927-940

Fahlman S E and C Liebere (1990/1991), "The Cascade-Correlation Learning Algorithm", Technical Report, CMU-CS-90-100, Carnegie Mellon University, 1-13

Fama E F (1965), "The Behavior of Stock Market Prices", Journal of Business, vol. 38, 34-105

Fama E F (1970), "Efficient Capital Markets: A Review of Theory and Empirical Work", Journal of Finance, vol. 25, no 2, 383-417

Fama E F (1991), "Efficient Capital Markets: II", Journal of Finance, vol. 46, no 5, 1575-1617

Fama E F (1998), "Market Efficiency, Long-term Returns, and Behavioural Finance", Journal of Financial Economics, vol. 49, 283-306

Fama E F and K R French (1988), "Dividend Yields and Expected Stock Returns", Journal of Financial Economics, vol. 22, 3-25

Fama E F and K R French (1992), "The Cross-Section of Expected Stock Returns", Journal of Finance, vol. 47, 427-465

Fama E F and K R French (1993), "Common Risk Factors in the Returns on Stocks and Bonds", Journal of Financial Economics, vol. 33, no 1, 3-56

Fama E F and K R French (1995), "Size and Book-to-Market Factors in Earnings and Returns", Journal of Finance, vol. 50, no 1, 131-156

Fama E F and K R French (1996), "Multifactor Explanations of Asset Pricing Anomalies", Journal of Finance, vol. 51, no 4, 55-84

Fama E F and K R French (1998), "Value versus Growth: The International Evidence", Journal of Finance, vol. 53, no 6, 1975-1999

Fama E F and K R French (2004), "The Capital Asset Pricing Model: Theory and Evidence", Journal of Economic Perspectives, vol. 18, no 3, 25-46

Fama E F and K R French (2006), "Profitability, Investment and Average Returns", Journal of Financial Economics, vol. 82, 491-518

Fama E F and K R French (2008), "Dissecting Anomalies", Journal of Finance, vol. 63, no 4, 1653-1678

Fama E F and J D MacBeth (1973), "Risk, Return, and Equilibrium", Journal of Political Economy, vol. 81, no 3, 607-636

Fama E F, Fisher L, Jensen M and R Roll (1969), "The Adjustment of Stock Prices to New Information", International Review, vol. 10

Ferguson M and R Shockley (2003), "Equilibrium 'Anomalies'", Journal of Finance, vol. 58, no 6, 2549-2580

Ferson W and C Harvey (1999), "Conditioning Variables and the Cross Section of Stock Returns", Journal of Finance, vol. 54, 1325-1360

Forner C and J Marhuenda (2003), "Contrarian and Momentum Strategies in the Spanish Stock Market", European Financial Management, vol. 9, 667-88

Frankish (2004), Multi-Factor Models and the Cross-Section of Equity Returns, Unpublished Master's Thesis, University of Cape Town

Fraser E and M J Page (2000), "Value and Momentum Strategies: Evidence from the JSE", Investment Analysts Journal, vol. 51, 25-35

Fuller R J (1981), Capital Asset Pricing Theories: Evolution and New Frontiers, Financial Analysts Research Foundation

Goetzmann W and P Jorion (1993), "Testing the Predictive Power of Dividend Yields", Journal of Finance, vol. 68, no 4, 483-508

Goetzmann W and P Jorion (1995), "A Longer Look at Dividend Yields", Journal of Business, vol. 68, no 4, 483-508

Goldberg D E (1989), "Genetic Algorithms in Search, Optimisation and Machine Learning", Addison-Wesley

Gonzalez S (2000), "Neural Networks for Macroeconomic Forecasting: A Complementary Approach to Linear Regression Models", Finance Canada Working Paper

Goodwin T (1998), "The Information Ratio", Financial Analysts Journal, vol. 54, no 4, 34-43

Graham B and D Dodd (1934), Security Analysis, McGraw Hill, New York

Grauer R (1999), "On the Cross-Sectional Relation between Expected Returns, Betas, and Size", Journal of Finance, vol. 54, no 2, 773-789

Grewal M S and A P Andrews (2001), "Kalman Filtering Using Matlab", John Wiley & Sons

Griffin J M (2002), "Are the Fama and French Factors Global or Country Specific?", The Review of Financial Studies, vol. 15, no 3, 783-803

Grinold R C (1989), "The Fundamental Law of Active Management", The Journal of Portfolio Management, vol. 15, 30-37

Grinold R C (1994), "Alpha is Volatility Times IC Times Score, or Real Alphas Don't Get Eaten", The Journal of Portfolio Management, vol. 20, 9-16

Grinold R C (1999), "Mean-Variance and Scenario-Based Approaches to Portfolio Selection", The Journal of Portfolio Management, vol. 25, no 2, 10-22

Grinold R C and R N Kahn (2000), Active Portfolio Management, McGraw-Hill, New York

Gupta F, Prajogi R and E Stubbs (1999), "The Information Ratio and Performance", The Journal of Portfolio Management, vol. 26, no 1, 33-39

Hanna J D and M J Ready (2005), "Profitable Predictability in the Cross Section of Stock Returns", Journal of Financial Economics, vol. 78, 463-505

Harrington D R (1987), Modern Portfolio Theory, The Capital Asset Pricing Model and Arbitrage Pricing Theory: A User's Guide, 2nd Edition, Prentice-Hall

Harvey C (1995), "Predictable Risk and Returns in Emerging Markets", Review of Financial Studies, vol. 8, no 3, 79-88

Haugen R A (1995), The New Finance: The Case Against Efficient Markets, Prentice Hall, USA

Haugen R A and N L Baker (1996), "Commonality in the Determinants of Expected Stock Returns", Journal of Financial Economics, vol. 41, 401-439

Haugen R A and N L Baker (2009), "Case Closed", The Handbook of Portfolio Construction: Contemporary Applications of Markowitz, J B Guerard Jr. (ed.), Forthcoming

Haykin S (2001): Kalman Filtering and Neural Networks, Wiley,

Hirshleifer (2001), "Investor Psychology and Asset Pricing", Journal of Finance, vol. 56, no 4, 1533-1597

Hodnett K and H Hsieh (2009), "Tests of the Overreaction Hypothesis on the Johannesburg Stock Exchange (JSE) in the Post Millennium ERA", South African Finance Association (SAFA) Conference

Holland J H (1975), "Adaptation in Natural and Artificial Systems", University of Michigan Press, Ann Arbor

Hong H and J C. Stein (1999), "A Unified Theory of Underreaction, Momentum Trading and Overreaction in Asset Markets", Journal of Finance, vol. 54, 2143-2184

Hornik K, Stinchcombe M and H White (1989), "Multilayer Feedforward Networks are Universal Approximators", Neural Networks, vol. 2, 359-366

Hsieh D (1995), "Nonlinear Dynamics in Financial Markets", Financial Analysts Journal, 55-62

Hung S, Liang T and V Liu (1996), "Integrating Arbitrage Pricing Theory and Artificial Neural Networks to Support Portfolio Management", Decision Support Systems, vol. 18, 301-316

Ibbotson R G (1975), "Price Performance of Common Stocks New Issues", Journal of Financial Economics, vol. 2, 232-272

Jarvis RM and R Goodacre (2005), "Genetic Algorithm Optimisation for Pre-Processing and Variable Selection of Spectroscopic Data", Bioinformatics, vol. 21, no 7, 860-868

Jegadeesh N (1990), "Evidence of Predictable Behaviour of Security Returns", Journal of Finance, vol. 45, no 3, 881-898

Jegadeesh N and S Titman (1993), "Returns to Buying Winners and Selling Losers: Implications for Stock Market Efficiency", Journal of Finance, vol. 48, no 1, 65-91

Jegadeesh N and S Titman (2001), "Profitability of Momentum Strategies: An Evaluation of Alternative Explanations", Journal of Finance, vol. 56, no 2, 699–720

Jensen M (1972), "Capital Markets: Theory and Evidence", The Bell Journal of Economics and Management Science, vol. 3, no 2, 357-398

Jensen M and R Ruback (1983), "The Market for Corporate Control: The Scientific Evidence", Journal of Financial Economics, vol. 11, 5-50

Kaastra I and M Boyd (1996), "Designing a Neural Network for Forecasting Financial and Economic Time Series", Neurocomputing, vol. 10, 215-236

Kahneman D and A Tversky (1979), "An Analysis of Decision under Risk", Econometrica, vol. 47, no 2, 263-291

Kalman R (1960), "A New Approach to Linear Filtering and Prediction Problems", Transactions of the ASME- Journal of Basic Engineering, 35-45

Kanas A and A Yannopoulos (2001), "Comparing Linear and Nonlinear Forecasts for Stock Returns", International Review of Economics and Finance, vol. 10, 383-398

Kendall M and A B Hill (1953), "The Analysis of Economic Time Series – Part I: Prices", Journal of the Royal Statistical Society, vol. 116, no 1, 11-34

Keim D B (1983), "Size-Related Anomalies and Stock Return Seasonality: Further Empirical Evidence", Journal of Financial Economics, vol. 12, 13-32

Kingdom J and K Feldman (1995), "Genetic Algorithms for Bankruptcy Prediction", Search Space Report, No. 09-95, Search Space Limited, London

Kothari S, Shanken J and G Sloan (1995), "Another Look at the Cross-Section of Expected Returns", Journal of Finance, vol. 50, no 1, 185-224

Kothari S, Lewellen J and J Warner (2006), "Stock Returns, Aggregate Earnings Surprises, and Behavioural Finance", Journal of Financial Economics, vol. 79, 537-568

Kovalishyn V, Tetko I, Luik A, Kholodovych V, Villa A and D Livingstone (1998), "Variable Selection in the Cascade-Correlation Learning Architecture", Journal of Chemical Information and Computer Science, vol. 38, 651-659

Kuan C and H White (1994), "Artificial Neural Networks: An Econometric Perspective", Economic Review, vol. 13, 1-91

La Porta R (1996), "Expectations and the Cross-Section of Returns", Journal of Finance, vol. 51, 1715-1742

La Porta R, Lakonishok J, Shleifer A and R W Vishny (1997), "Good News for Value Stocks", Journal of Finance, vol. 52, no 2, 859-874

Lacher R, Coats P, Sharma S and F Fant (1995), "A Neural Network For Classifying the Financial Health of a Firm", European Journal of Operational Research, vol. 85, 53-65

Lakonishok J, Shleifer A and R W Vishny (1994), "Contrarian Investment, Extrapolation and Risk", Journal of Finance, vol. 49, no 5, 1541-1578

Lary D J and H Y Mussa (2004), "Using An Extended Kalman Filter Learning Algorithm for Feed-Forward Neural Networks to Describe Tracer Correlations" Atmospheric Chemistry and Physics, vol. 4, no 3, 3653-3667

Lewellen J (2004), "Predicting Returns with Financial Ratios", Journal of Financial Economics, vol. 74, 209-235

Lintner J (1965), "The Valuation of Risky Assets and the Selection of Risky Investments in Stock Portfolios and Capital Budgets", Review of Economics and Statistics, vol. 47, no 1, 13-37

Litzenberger R and K Ramaswamy (1979), "The Effects of Personal Taxes and Dividends on Capital Asset Prices: Theory and Empirical Evidence", Journal of Financial Economics, vol. 7, 163-195

Lo A and A C MacKinlay (1990), "Data-Snooping Biases in Tests of Financial Asset Pricing Models", Review of Financial Studies, vol. 3, no 3, 431-468

Lo A and A C MacKinlay (1990), “When are Contrarian Profits Due to Stock Market Overreaction?”, Review of Financial Studies, vol. 3, no 2, 175-205

Mahfoud S and G Mani (1995), “Genetic Algorithms for Predicting Individual Stock Performance”, Proceedings of the 3rd International Conference on Artificial Intelligence Applications on Wall Street, 171-181

Malkiel B G (1990), A Random Walk Down Wall Street, Norton and Company, New York

Mao K Z and S A Billings (1999), “Variable Selection in Non-Linear Systems Modelling”, Mechanical Systems and Signal Processing, vol. 13, no 2, 351-366

Markowitz H M (1952), "Portfolio Selection", Journal of Finance, vol. 7, no 1, 77–91

Markowitz H M (1959), Portfolio Selection: Efficient Diversification of Investments, John Wiley & Sons, New York

Markowitz H M (2005), "Market Efficiency: A Theoretical Distinction and So What?", Financial Analysts Journal, vol. 61, no 5, 17-30

Marose R A (1990), "A Financial Neural-Network Application", Al Expert, vol. 5, no 5, 50-53

McNelis P (2005), Neural Networks in Finance: Gaining Predictive Edge in the Market, Elsevier

Modigliani F and Pogue G (1988), “Risk, Return and CAPM: Concepts and Evidence”, The Financial Analyst’s Handbook, 2nd Edition, Levine S (ed.), Dow Jones Irwin, Homewood

Moore (2008), An Investigation of Firm Specific and Macroeconomic Variables and their Influence on Emerging Market stock Returns, Unpublished Master’s Thesis, University of Cape Town

Mossin J (1966), "Equilibrium in a Capital Asset Market", Econometrica, vol. 34, no 4, 768-783

Motiwalla L F and M Wahab (2000), "Predictable Variation and Profitable Trading of US Equities: A Trading Simulation Using Neural Networks" Computers and Operations Research, vol. 27, 1111-1129

Muller C (1999), "Investor Overreaction on the Johannesburg Stock Exchange", Investment Analysts Journal, no 49, 5-17

Mutooni R and C Muller (2007), "Equity Style Timing", Investment Analysts Journal, no 65, 15-24

Nechyba M and Y Xu (1997), "Cascade Neural Networks with Node-Decoupled Extended Kalman Filter", Proceedings on IEEE Intelligence Symposium on Computational Intelligence in Robotics and Automation, vol. 1, 214-219

Nofsinger J R (2005), The Psychology of Investing, 3rd Edition, Prentice Hall

Orderud F (2005), "Comparison of Kalman Filter Estimation Approaches for State Space Models with Nonlinear Measurements", In Proceedings of Scandinavian Conference on Simulation and Modeling

Page M J (1996), "Further Evidence of Firm Size and Earnings Anomalies on the Johannesburg Stock Exchange", De Ratione, vol. 10, no 1, 27-44

Page M J and F Palmer (1993), "The Relationship between Excess Returns, Firm Size and Earnings on the Johannesburg Stock Exchange", South African Journal of Business Management, vol. 22, no 3, 63-73

Page M J and C V Way (1992/1993), "Stock Market Overreaction: The South African Evidence", Investment Analysts Journal, vol. 36, 35-49

Pesaran M and A Timmermann (1995), "Predictability of Stock Returns: Robustness and Economic Significance", Journal of Finance, vol. 50, no 4, 1201-1228

Qi M (1999), "Nonlinear Predictability of Stock Returns Using Financial and Economic Variables", Journal of Business and Economic Statistics, vol. 17, no 4, 419-429

Qi M and G Maddala (1995), "Option Pricing Using ANN: The Case of S&P 500 Index Call Options", Proceedings of the 3rd International Conference on Neural Networks in the Capital Markets, London

Qi M and G Maddala (1998), "Economic Factors and the Stock Market: A New Perspective", Journal of Forecasting, vol. 18, 151-166

Qian E and R Hua (2003), "The Information Ratio of Active Management", Putnam Investments

Qian E and R Hua (2004), "Active Risk and The Information Ratio", Journal of Investment Management, vol. 2, no 3, 1-15

Quah T and B Srinivasan (1999), "Improving Returns on Stock Investments through Neural Network Selection", Expert Systems with Applications, vol. 17, 295-301

Reilly F K and Brown K C (2003), Investment Analysis and Portfolio Management, 7th Edition, Thomson Learning

Reinganum M R (1981), "Misspecification of Capital Asset Pricing – Empirical Anomalies Based on Earnings' Yields and Market Values", Journal of Financial Economics, vol. 9, 19-46

Reinganum M R (1983), "The Anomalous Stock Market Behaviour of Small Firms in January – Empirical Tests for Tax-Loss Selling Effects", Journal of Financial Economics, vol. 12, 89-104

Rojas R (1996), Neural Networks – A Systematic Introduction, Springer, Berlin

Roll R (1977), "A Critique of the Asset Pricing Theory's Tests Part I: On Past and Potential Testability of the Theory", Journal of Financial Economics", vol. 14, no 2, 129-176

Roll R (1978), "Ambiguity When Performance is Measured by the Security Market Line", Journal of Finance, vol. 33, no 4, 1051-1069

Roll R (1981), "A Possible Explanation of the Small Firm Effect", Journal of Finance, vol. 36, no 4, 879-888

Roll R and S A Ross (1980), "An Empirical Investigation of the Arbitrage Pricing Theory", Journal of Finance, vol. 35, no 5, 121-130

Roll R and S A Ross (1984), "The Arbitrage Pricing Theory Approach to Strategic Portfolio Planning", Financial Analysts Journal, vol. 40, no 1, 14-26

Rosenberg B, Reid K and R Lanstein (1985), "Persuasive Evidence of Market Inefficiency", Journal of Portfolio Management, vol. 11, 9-17

Ross S A (1976), "The Arbitrage Theory of Capital Asset Pricing", Journal of Economic Theory, vol. 13, no 2, 341-360

Rousseau R and P Van Rensburg (2004), "Time and the Payoff to Value Investing", Journal of Asset Management, vol. 4, no 5, 318-325

Rouwenhorst K G (1999), "Local Return Factors and Turnover in Emerging Stock Markets", Journal of Finance, no 54, 1439-1464

Ruck W, Rogers S, Kabrisky M, Maybeck P and M Oxley (1992), "Comparative Analysis of Backpropagation and the Extended Kalman Filter for Training Multilayer Perceptrons", IEEE Transactions on Pattern Analysis and Machine Intelligence, vol. 14, no 6, 686-691

Rumelhart D E and J L McClelland (1986), “Parallel Distributed Processing”, Cambridge Mass.: Massachusetts Institute of Technology Press, vol.1

Rutan E (1993), “Experiments with Optimal Stock Screens”, Proceedings of the 3rd International Conference on Artificial Intelligence Applications on Wall Street, 269-273

Samuelson P (1965), "Proof that Properly Anticipated Prices Fluctuate Randomly", Industrial Management Review, vol. 6, no 2, 41–49

Schiereck D, De Bondt W F M and M Weber (1999), “Contrarian and Momentum Strategies in Germany”, Financial Analysts Journal, vol. 55, no 6, 104-116

Schwert, G W (1983), “Size and Stock Returns, and Other Empirical Regularities”, Journal of Financial Economics, vol. 12, 3-12

Sen T, Oliver R and N Sen (1995), “Predicting Corporate Mergers”, In Neural Networks in Capital Markets, Refenes, (eds), John Wiley and Sons, 325-340

Serra A P (2000), “Country and Industry Factors in Returns: Evidence from Emerging Markets’ Stocks”, Emerging Markets Review, vol. 1, 127-151

Serra A P (2002), “The Cross-Sectional Determinants of Returns: Evidence from Emerging Markets’ Stocks”, Working Paper

Serra A P (2003), “The Cross-Sectional Determinants of Returns: Evidence from Emerging Markets’ Stocks, Journal of Emerging Market Finance, vol. 2, no 2, 123-162

Sharpe W F (1964), “Capital Asset Prices: A Theory of Market Equilibrium under Conditions of Risk”, Journal of Finance, vol. 19, no 3, 425-442

Sharpe W F (1966), “Mutual Fund Performance”, Journal of Business, vol. 39, no 1, 119-138.

Sharpe W F (1992), "Asset Allocation: Management Style and Performance Measurement", Journal of Portfolio Management, vol. 18, 7-19

Shefrin H and M Statman (1985), "The Disposition to Sell Winners Too Early and Ride Losers too Long: Theory and Evidence", Journal of Finance, vol. 40, no 3, 777-782

Sheskin I (1997), Handbook of Parametric and Nonparametric Statistical Procedures, CRC-Press

Singhal S and L Wu (1989), "Training Multilayer Perceptrons with the Extended Kalman Algorithm", Advances in Neural Information Processing Systems1, Touretzky D (ed.), Morgan Kaufmann Publishers, 133-140

Singleton J C and A J Surkan (1990), "Neural Networks for Bond Rating Improved by Multiple Hidden Layers", Proceedings of the IEEE International Conference on Neural Networks, San Diego, CA, 2, 163-168

Singleton J C and A J Surkan (1990), "Bond Rating with Neural Networks", In Neural Networks in the Capital Markets, Refenes, (ed.), John Wiley & Sons, 301-307

Sofge D A (2002), "Using Genetic Algorithm Based Variable Selection to Improve Neural Network Models for Real-World Systems", Proceedings of the 2002 International Conference on Machine Learning and Applications

Sorensen E, Qian E, Schoen R, and R Hua (2004), Multiple Alpha Sources and Active Management, Journal of Portfolio Management, vol. 30, no 2, 39-46.

Statman M (1999), "Behavioural Finance: Past Battles and Future Engagements", Financial Analysts Journal, vol. 55, no 6, 18-27

Stattman D (1980), "Book Values and Expected Stock Returns", Unpublished MBA dissertation, University of Chicago, Illinois

Stehle R, Bunke O and V Sommerfeld (1997), "Semiparametric Modelling of the Cross-Section of Expected Returns in the German Stock Market", Humboldt University Berlin, Germany

Sum J, Leung C, Young G and W Kan (1999), "On the Kalman Filtering Method in Neural Network Training and Pruning", IEEE Transactions on Neural Networks, vol. 10, no 1, 161-166

Swanson N R and H White (1995a), "A Model-Selection Approach to Assessing the Information in the Term Structure Using Linear Models and Artificial Neural Networks", Journal of Business & Economic Statistics, vol. 13, 265-275

Swanson N R and H White (1995b), "A Model-Selection Approach to Real-Time Macroeconomic Forecasting Using Linear Models and Artificial Neural Networks", Working Paper, Department of Economics, Penn State University

Takane Y (1998), "Nonlinear Multivariate Analysis by Neural Network Models", In Data Science, Classification, and Related methods, Hayashi C, Ohsumi N, Yajima K, Tanaka Y, Bock H and Y Baba (eds.), Tokyo: Springer, 527-538.

Taylor J G (2002), "Neural Networks", In Neural Networks and the Financial Markets, Shadbolt J and J G Taylor, (eds.) Springer, 87-93

Trippi R and J K Lee (1996), Artificial Intelligence in Finance and Investing, Irwin

Tobin J (1958), "Liquidity Preference as Behaviour Towards Risk", The Review of Economic Studies, vol. 25, 65-86

Utans J and J Moody (1991), "Selecting Neural Network Architectures via the Prediction Risk: Application to Corporate Bond Rating Prediction", Proceedings of the 1st International Conference on Artificial Intelligence Applications on Wall Street, IEEE Computer Society Press, CA

Van Rensburg P (1997), "Employing the Pre-Specified Variable Approach to APT Factor Identification on the Segmented Johannesburg Stock Exchange", South African Journal of Accounting Research, vol. 11, no 1, 57-75

Van Rensburg P (2001), "A Decomposition of Style-Based Risk on the JSE", Investment Analysts Journal, vol. 54, 45-60

Van Rensburg P and E Janari (2008), "Firm-Specific Attributes and the Cross-Section of Australian Stock Exchange Returns", Journal of Asset Management, vol. 9, no 3, 193-214

Van Rensburg P and M Robertson (2003a), "Style Characteristics and the Cross-Section of JSE Returns", Investment Analysts Journal, no 57, 1-10

Van Rensburg P and M Robertson (2003b), "Size, Price-to-Earnings and Beta on the JSE", Investment Analysts Journal, vol. 58, 1-11

Van Rensburg P and K B E Slaney (1997), "Market Segmentation on the Johannesburg Stock Exchange", Journal for Studies in Economics and Econometrics, vol. 23, no 3, 1-23

Vuolteenaho T (2002), "What Drives Firm-level Stock Returns?", Journal of Finance, vol. 57, no 1, 233-264

Wallet B C, Marchette D J, Solka J L and E J Wegman (1996), "A Genetic Algorithm for Best Subset Selection in Linear Regression", Proceedings of the 28th Symposium on the Interface

Warner B and M Misra (1996), "Understanding Neural Networks as Statistical Tools", The American Statistician, vol. 50, no 4, 284-293

Welch G and G Bishop (2006), An Introduction to the Kalman Filter, Department of Computer Science, University of North Carolina

Werbos P J (1974/1994), The Roots of Backpropagation, NY: John Wiley& Sons