A Comparative Study of Stock Scoring using Regression and Genetic-based Linear Models

Chien-Feng Huang
Department of Computer Science and Information Engineering
National University of Kaohsiung
Kaohsiung, Taiwan, R.O.C.
cfhuang15@nuk.edu.tw

Tsung-Nan Hsieh
Department of Computer Science and Information Engineering
National University of Kaohsiung
Kaohsiung, Taiwan, R.O.C.
m0995513@mail.nuk.edu.tw

Bao Rong Chang
Department of Computer Science and Information Engineering
National University of Kaohsiung
Kaohsiung, Taiwan, R.O.C.
brchang@nuk.edu.tw

Chih-Hsiang Chang
Department of Finance
National University of Kaohsiung
Kaohsiung, Taiwan, R.O.C.
cch@nuk.edu.tw

***Abstract*—Stock selection has long been a challenging and important task in investment and finance. Researchers and practitioners in this area often use regression models to tackle this problem due to their simplicity and effectiveness. Recent advances in machine learning (ML) are leading to significant opportunities to solve these problems more effectively. In this paper, we present a comparative study between the traditional regression-based and ML-based linear models for stock scoring, which is crucial to the success of stock selection. In ML-based models, Genetic Algorithms (GA), a class of well-known search algorithms in the area of ML, is used for optimization of model parameters and selection of input variables to the stock scoring model. We will show that our proposed genetic-based method significantly outperforms the traditional regression-based method as well as the benchmark. As a result, we expect this genetic-based methodology to advance the research in machine learning for finance and provide an attractive alternative to stock selection over the regression-based approach.**

Keywords- Stock selection; stock scoring; regression; genetic algorithms; optimization; model validation

I. INTRODUCTION

Investment has long been recognized as a challenging and important problem in finance. Researchers and practitioners in this area often rely heavily on regression models to study this subject due to their simplicity and effectiveness. Examples include the testing of the implications of a multi-asset equilibrium model [1]; the forecasting of stock market returns using the dividend yield, the earnings growth, and the price-earnings ratio growth [2]; the investigation of the relationship between portfolio diversification and risk reduction [3-4]; the classification of winning and losing stocks for portfolio construction [5], to name just a few.

Apart from these traditional regression-based approaches, recent advances in computational intelligence and machine learning are leading to attractive alternatives to solving these problems more effectively. Feasible quantitative models include methodologies stemming from soft computing [6] for prediction of financial time series, multi-objective optimization of expected investment return and risk reduction, and portfolio management – selection of investment instruments based on asset ranking using a variety of input variables and historical data, etc. [7]. All these research efforts were in an attempt to facilitate the task of decision-making for investment.

In the research area of stock selection and portfolio optimization, several machine learning methodologies have been developed, including artificial neural networks (ANNs), support vector machines (SVMs), evolutionary algorithms (EAs) as well as fuzzy inference models. Quah and Srinivasan [8] studied an ANN stock selection system to choose stocks that are top-ranked performers. They showed that their proposed model outperformed the benchmark model in terms of compounded actual returns overtime. Chapados and Bengio [9] also trained neural networks for estimation and prediction of asset behavior in order to facilitate decision-making in asset allocation. Although these models worked in some applications, they often suffer from the overfitting problem and may tend to fall into a local optimum.

For portfolio optimization, Kim and Han [10] proposed a genetic algorithm (GA) approach to feature discretization and the determination of connection weights for ANNs to predict the stock price index. They suggested that their approach was able to reduce the numbers of attributes and the prediction performance was enhanced. In addition, Caplan and Becker [11] employed genetic programming (GP) to develop a stock ranking model for the high technology manufacturing industry in the U.S. More recently, Becker et al. [12] explored various single-objective fitness functions for GP to construct stock selection models for particular investment specifics with respect to risk. In a nutshell, these GP-based models rank stocks from high to low according to a pre-defined objective function.

In the area of fuzzy applications in finance, earlier work includes, for instance, Chu *et al.*'s fuzzy multiple attribute decision analysis to select stocks for portfolio construction [13]. Analogously, Zargham and Sayeh [14] employed a fuzzy rule-based system to evaluate a set of stocks for the same purpose. Although these fuzzy approaches denote early efforts in employing computational intelligence for financial applications, they usually lack sufficient learning ability.

Despite the promising performance of the aforementioned approaches in finance, their success is highly contingent upon the input variables (features) to the model. Yang and Honavar [15] indicated that several classification issues are determined by the choice of features that describe given patterns presented to a classifier, such as the classification accuracy of the learned classifier, the computational cost needed for learning a classification function, and the number of training instances needed for learning. Therefore, feature selection may be used to identify useful, non-redundant subsets of features for a given machine learning task.

Furthermore, since the variables relevant to the machine learning models usually consist of not only the features but also the models parameters, it is expected that a successful modeling shall neglect neither of these two issues. Therefore, in our previous work [16], we devised a GA-based stock scoring model for the simultaneous task of feature selection and optimization of parameters. Based on the scores calculated, top-ranked stocks are chosen for portfolio construction. In [16], we showed that the portfolios constructed by our scheme substantially outperform the benchmark over the long period of time.

Despite the promising performance reported in [16], a serious comparison between the traditional regression-based and our ML-based methods for stock selection is still lacking. Our goal in this study is therefore to conduct such a comparison to show that ML-based method is indeed a valuable tool that provides an attractive alternative to stock selection over the regression-based method.

This paper is organized into five sections. Section 2 outlines the regression-based and our proposed methods. In Section 3, we describe the research data used in this study. In Section 4, we present the experimental design and empirical results are reported and discussed. Section 5 concludes this paper with future research directions.

II. METHODOLOGY

This section first describes the regression-based stock scoring scheme. Afterwards, our proposed scoring scheme along with model optimization (i.e., parameter optimization and feature selection) by the GA will be discussed.

A. Regression models

Regression is an approach to modeling the relationship between the output variable y and a set of input variables denoted as $\boldsymbol{X}$. In linear regression models, linear functions are used to model data, and unknown model parameters are then estimated from the data.

More specifically, consider a given set S with n training instances $\{(x_1(t), y_1(t)), (x_2(t), y_2(t)), ..., (x_n(t), y_n(t))\}$ at time t. Each training instance $x_i(t) \in R^p$, for $i = 1, ..., n$, serves as the inputs to generate a corresponding output $y_i(t) \in R$, where p is the input dimension. Using β and $\varepsilon(t)$ to denote the regression coefficients and the error terms, respectively, the linear regression model often takes the form

$$Y(t) = X(t)\,\beta + \varepsilon(t),$$

where

$$Y(t)=\begin{pmatrix} y_1(t) \\ y_2(t) \\ \vdots \\ y_n(t) \end{pmatrix}, X(\mathrm{t})=\begin{pmatrix} x_{11}(t) & \cdots & x_{1p}(t) \\ x_{21}(t) & \cdots & x_{2p}(t) \\ \vdots & \ddots & \vdots \\ x_{n1}(t) & \ldots & x_{np}(t) \end{pmatrix},\ \beta=\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_p \end{pmatrix},\ \ \varepsilon(t)=\begin{pmatrix} \varepsilon_1(t) \\ \varepsilon_2(t) \\ \vdots \\ \varepsilon_n(t) \end{pmatrix}.$$

In this study, we are concerned with the relative quality of stocks described by the fundamental variables, including firms' share price rationality, growth, profitability, liquidity, efficiency, and leverage attributes. These variables may be used as the inputs (i.e., x_i's) to the regression model and the returns of the stocks (i.e., y_i's) are the resultant output of the model. Therefore, once the regression model is constructed, it can be used to predict the stock returns in the future as long as the values of the fundamental variables are provided. In this study, we will use the predicted returns of stocks as surrogates to score stocks, so that the higher the predicted return of a stock, the higher its score. The goal of doing so is to imply the relative quality of stocks for future ranking as discussed shortly in Subsection *C*.

B. GA-based models

The objective of the GA-based stock scoring model proposed in our previous work [16] was to imply stocks of higher scores to possess higher potential in future price advancement. Based on these scores one can rank various stocks and top-ranked stocks are picked to construct the portfolio.

In [16], we developed a straightforward linear model using the fundamental variables to score stocks. More specifically, let $Z_{i,j}(t)$ denote the score of stock i assigned by variable j at time t, where $Z_{i,j}(t)$ depends on the value of variable j, $v_{i,j}(t)$, for stock i at time t. For instance, in the area of value investing, if the variable is the price-to-book ratio (P/B ratio), a smaller P/B ratio tends to imply the stock's higher potential of price increase in the future [16]. On the contrary, if the variable is return-on-assets (ROA), a higher ROA usually implies the stock's higher potential of price increase in the future.

Therefore, in [16] we proposed to sort the stocks according to their values of variable j and the individual score assigned to stock i at time t is:

$$Z_{i,j}(t) = \rho_{i,j}(t),$$

where $\rho_{i,j}(t) \in N$ is the ranking of stock i with respect to variable j at time t. Here we denote a stock sorting indicator I_j for variable j and consider two cases for the stock sorting scheme:

(1) $I_j = 0$: $\rho_{i,j}(t) \geq \rho_{k,j}(t)$ iff $v_{i,j}(t) \geq v_{k,j}(t)$ for $i \neq k$.

(2) $I_j = 1$: $\rho_{i,j}(t) \geq \rho_{k,j}(t)$ iff $v_{i,j}(t) \leq v_{k,j}(t)$ for $i \neq k$.

In addition, let W_j denote the weight of the j-th variable. Then the total score of stock i at time t, $y_i(t)$, can be defined as

$$y_i(t) = \sum_j W_j Z_{i,j}(t). \tag{1}$$

C. Stock ranking scheme

It is worthwhile to mention that the scores calculated for the stocks may not necessarily represent the precise values of various stocks. Rather, they can serve as surrogates for the actual quality to imply the relative rankings of the stocks. More specifically, given the scores ($y_i(t)$'s) for all stocks, the ranking of a stock can be defined as:

$$\alpha_i(t) = \rho(y_i(t)), \tag{2}$$

where $\rho(\cdot)$ is a ranking function so that $\alpha_i(t) \in N$ is the ranking of stock i at time t, and $\alpha_i(t) \geq \alpha_j(t)$ iff $y_i(t) \geq y_j(t)$.

The task of stock selection can be accomplished using these rankings whereby top-ranked m stocks (stocks corresponding to the top m α's) are selected as components of a portfolio. The performance of a portfolio can be evaluated by averaging the actual returns of the stocks in the portfolio, which is defined as:

$$\overline{R_t} = \frac{1}{m}\sum_{i=1}^{m} R_t(s_i(t)), \tag{3}$$

where $s_i(t)$ is the i-th ranked stock at time t; $R_t(\cdot)$ is the actual return for a stock at time t and $\overline{R_t}$ is the average return over all the m stocks in the portfolio at time t.

In this study we use the cumulative total (compounded) return, R_c, to evaluate the performance of a stock selection model, where R_c is defined as the product of the average annual return, $\overline{R_t}$, of the constitute stocks in a portfolio over n consecutive years as:

$$R_c = \prod_{t=1}^{n} \overline{R_t}. \tag{4}$$

D. Model optimization by genetic algorithms

The performance of the GA-based stock ranking model is determined by the set of input features $\boldsymbol{F}$, the set of stock sorting indicators $\boldsymbol{I}$, the weights of the fundamental variables $\boldsymbol{W}$. Therefore, the selection of optimal subsets of features $\boldsymbol{F}$, and the optimization of $\boldsymbol{I}$ and $\boldsymbol{W}$ will be critical to the effectiveness of the stock selection model. In [16], Genetic Algorithms were employed for simultaneous optimization with respect to these tasks. In the following we briefly describe the relevant GA-based optimization scheme for the stock selection model.

Among many paradigms of search algorithms GAs have been proven to have an advantage over traditional optimization methods in problems with many complex, discontinuous constraints in the search space. This methodology contributes for a global, population-based search in the search space, in contrast with the kind of local, greedy search conducted by most rule-induction and decision-tree algorithms. Lower computation cost is a general advantage of local, greedy search algorithms. However, the solution quality achieved by these algorithms can be greatly degraded if there exists a considerable degree of feature interactions, which is usually the case for real-world problems. Since GAs can be designed to perform a global search for various combinations of sets of features that improve given optimization criteria, this class of algorithms are expected to cope better with feature interaction problems.

Apart from feature selection, two sets of free parameters, $\boldsymbol{I}$ and $\boldsymbol{W}$, are to be provided for the GA-based stock ranking model. In [16] the GA was employed for simultaneous optimization of these tasks. In the overall encoding design, the composition of a chromosome was devised to consist of three portions — the candidate set of features $\boldsymbol{F}$, the stock sorting indicators $\boldsymbol{I}$ and the weighs $\boldsymbol{W}$. In Fig. 1, loci b_f^1 through b_f^n represent candidate features 1 through n, respectively. For these features, allele '1' or '0' corresponds to the feature being selected or not. Loci b_i^1 through b_i^n represent the sorting indicators, where 0 represents the variable being used for case (1) of the stock sorting scheme, and 1 represents case (2), respectively. On the right-hand side of Fig. 1 is the encoding of the set of parameters $\boldsymbol{W}$. Fig. 2 shows the detailed binary encoding for the weight of each individual variable, where the value of W_i (the weight for variable i) is encoded by loci b_{Wi}^1 through $b_{Wi}^{n_i}$.

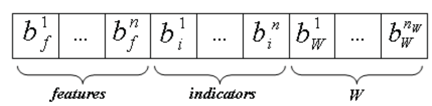

Figure 1. Chromosome encoding

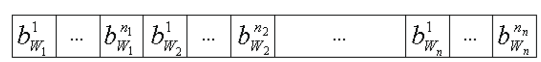

Figure 2. Detailed encoding of the weighting terms

In the coding scheme, the portion in the chromosome representing the genotypes of parameter W_i's is to be transformed into the phenotype by Eq. (5) for further fitness computation. The precision representing each parameter depends on the number of bits used to encode it in the chromosome, which can be determined as follows:

$$y = \min{}_y + \frac{d}{2^l - 1} \times (\max{}_y - \min{}_y), \tag{5}$$

where y is the corresponding phenotype for the particular parameter; $\min_y$ and $\max_y$ are the minimum and maximum of the parameter; d is the corresponding decimal value, and l is the length of the block used to encode the parameter in the chromosome.

With this encoding scheme, the fitness function of a chromosome can be defined as the annualized return of the portfolio:

$$fitness = \sqrt[n]{R_c}, \tag{6}$$

Table I. Attributes used in the stock selection model

Attribute	Ratios	Description	Ref.
Share price rationality	(1) PE ratio	Price-to-earnings ratio = share price / earnings per share	[17, 19-20, 22]
	(2) PB ratio	Price-to-book ratio = share price / book value per share	[17-21]
	(3) PS Ratio	Price-to-sales ratio = share price / sales per share	[17]
Profitability	(4) ROE	Return on equity (after tax) = net income after tax / shareholders' equity	[23-24]
	(5) ROA	Return on assets (after tax) = net income after tax / total assets	[23]
	(6) OPM	Operating profit margin = operating income / net sales	[25]
	(7) NPM	Net profit margin = net income after tax / net sales	[24]
Leverage	(8) DE ratio	Debt-to-equity ratio = total liabilities / shareholders' equity	[23]
Liquidity	(9) CF ratio	Cash flow ratio = cash flow from operating activities / current liabilities	[26]
	(10) CR	Current ratio = current assets / current liabilities	[23]
	(11) QR	Quick ratio = quick assets / current liabilities	[23]
Efficiency	(12) ITR	Inventory turnover rate = cost of goods sold / average inventory	[23]
	(13) RTR	Receivables turnover rate = net credit sales / average accounts receivable	[27]
Growth	(14) OIG	Operating income growth rate = (operating income at the current year – operating income at the previous year) / operating income at the previous year	[28]
	(15) NIG	Net income growth rate = (net income after tax at the current year – net income after tax at the previous year) / net income after tax at the previous year	[29]

where R_c is the cumulative total return computed by Eq. (4).

To sum up, our proposed GA model in [16] for stock selection is a multi-stage process, including feature selection and parameter optimization by the GA, stock scoring, stock ranking and selection, as well as performance evaluation. The flowchart of this algorithm is shown in Fig. 3.

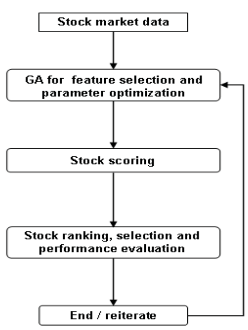

Figure 3. Flow chart of the GA model

III. DATA AND FUNDAMENTAL VARIABLES

We use the constituent stocks of the 200 largest market capitalizations listed in the Taiwan Stock Exchange as the investment universe. The yearly financial statement data and stock returns used for this research are retrieved from the TEJ (Taiwan Economic Journal Co. Ltd., http://www.tej.com.tw/) database for the period of time from 1995 to 2009. For the choice of fundamental variables, early studies indicated that several financial ratios play key roles in future stock returns. Most of them applied profitability (e.g., ROE, ROA, operating profit margin, and net profit margin), leverage (e.g., DB ratio), liquidity (e.g., current ratio and quick ratio), efficiency (e.g., inventory turnover rate and receivables turnover rate), and growth (e.g., operating income growth rate and net income growth rate) related ratios to examine the relationship between fundamentals and stock returns. Mukherji *et al.* [17], Jensen *et al.* [18], Danielson and Dowdell [19], Lewellen [20], Fama and French [21], and Hjalmarsson [22] also showed that the ratios relating to share price rationality, e.g., PE, PB, and PS ratios, are likely to influence future stock returns. According to the previous literature, Table 1 provides the aforementioned six attributes that are to be employed for this study, including fifteen financial ratios. For each year, investable stocks are described by these fifteen financial ratios and their historical returns are provided.

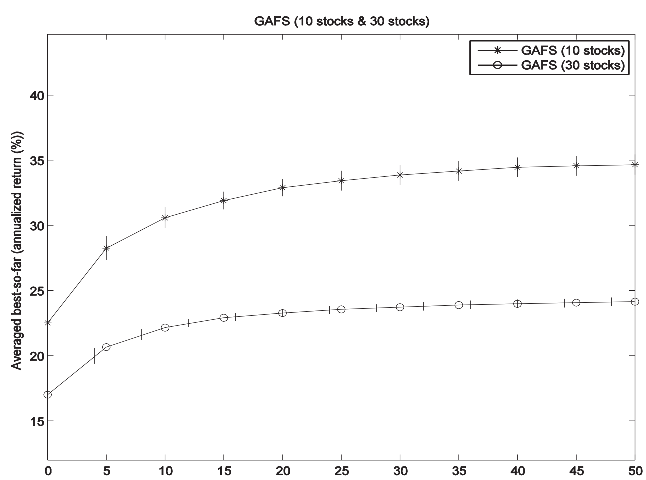

Figure 4. Best-so-far curves by the GAFS

Table II. Statistics of the benchmark, regression-based v.s. GA-based stock selection models for 30 stocks

Training period	Annualized benchmark return (%)	Annualized regression-model return (%)	Mean of annualized GA-model return (%)	Standard deviation of annualized GA-model return (%)	Testing period	Annualized benchmark return (%)	Annualized regression-model return (%)	Mean of annualized GA-model return (%)	Standard deviation of annualized GA-model return (%)
1995	-36.5079	-23.5091	-12.6511	0.7489	1996-2009	3.5111	6.6250	19.3132	1.9465
1995-1996	-10.1106	3.0185	9.1239	0.8107	1997-2009	1.8793	6.2228	15.8839	2.3647
1995-1997	12.8345	35.8647	57.1157	1.4235	1998-2009	-2.7400	2.7879	11.8631	1.5427
1995-1998	2.1765	19.3320	36.7931	1.1984	1999-2009	-0.5192	5.9096	13.4321	2.0628
1995-1999	2.1065	24.9227	39.4048	1.6962	2000-2009	-0.7508	6.4038	10.7420	1.9226
1995-2000	-0.4645	17.7135	30.4963	1.4909	2001-2009	0.6331	5.0830	11.8939	2.3368
1995-2001	-8.5626	12.5616	20.7217	1.0137	2002-2009	8.5385	12.4623	18.8043	2.4823
1995-2002	-5.6438	12.7313	23.0776	0.9618	2003-2009	7.3061	11.4078	19.3081	2.5509
1995-2003	-3.4626	15.8575	24.9503	0.7549	2004-2009	5.9366	13.6387	16.7977	3.1189
1995-2004	-2.9502	18.4946	25.3182	0.7039	2005-2009	6.7869	12.4526	14.6793	2.6588
1995-2005	-2.0844	16.5665	26.0959	0.6210	2006-2009	6.7311	10.4200	11.2700	2.8401
1995-2006	-0.6369	16.1631	25.4571	0.6748	2007-2009	3.5806	5.3973	10.6500	2.6348
1995-2007	2.2965	18.0834	28.3132	1.1360	2008-2009	-12.4666	-12.3408	-9.1414	2.1001
1995-2008	-0.1914	17.6252	24.1399	0.8497	2009	5.7270	3.9852	5.3718	3.2659

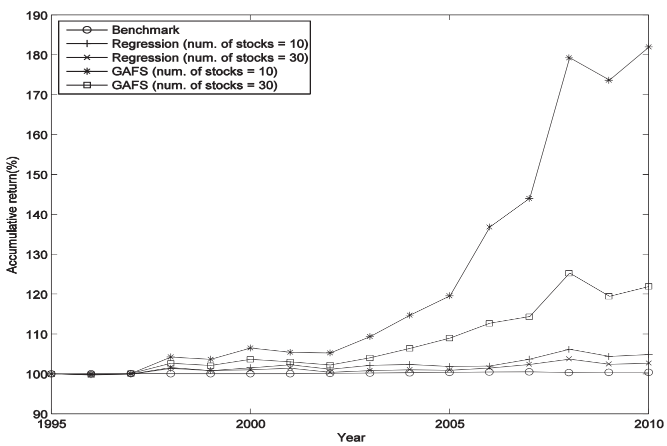

Figure 5. Cumulative returns of benchmark v.s. longing top-ranked stocks by the regression-based and GA-based models

IV. EMPIRICAL RESULTS

To illustrate our proposed GA models with feature selection (GAFS), stock data of all the years is used. Stocks are ranked based on the scores obtained. Top *m* stocks are selected and the average yearly returns of these selected stocks are calculated. Fig. 4 displays the averaged best-so-far values of the annualized returns over 50 runs and 50 generations for m=10 and 30. The averaged best-so-far curve is calculated by averaging the best-so-fars obtained by the GA at each generation for all 50 runs, where the vertical bars overlaying the curves represent the 95-percent confidence intervals about the means.

Fig. 5 displays an illustration of the cumulative benchmark return (the product of the average yearly returns of the 200 stocks in the investment universe) and the cumulative returns of longing a number of top-ranked stocks recommended by the regression-based and GA-based models. This figure clearly shows that the GA-based models can significantly outperform the regression-based models and the benchmark.

In order to examine the validity of the GA-based models we proposed, further statistical validation is conducted. In reality, the learned model has to be tested by unseen data. Here, we use the stock data of the first several years to train the GA model, and the remaining years are used for testing the learned model. Thus the average yearly return of the selected stocks for each of the testing year is calculated. The average yearly returns are then compounded to obtain the cumulative total return of the portfolio over the testing years.

Notice that this setup is different from the regular cross-validation procedure where the process of data being split into two independent sets is randomly repeated several times without taking into account the data's temporal order. However, in the stock selection study here, temporal order is critical as practically one would like to use all available data so far to train the model and to apply the models in the future to gain profits.

Here we compare the cumulative benchmark return (the product of the average yearly returns of the 200 stocks in the investment universe) and the cumulative average return of longing a number of top-ranked stocks using the regression-based and GA-based models. Table 2 shows the model validation. An inspection on the means of the annualized GA-model returns shows that in 13 out of 14 testing cases both the regression-based and GA-based models outperformed the benchmark. Furthermore, the GA-based model outperformed the regression-based model in all the 14 testing cases. As a result, these validations provide statistical evidence that our proposed GA-based stock selection methodology provides a superior solution to the traditional regression-based methods and is thus promising for investment in research and practice.

V. CONCLUSIONS

In this paper we presented a comparative study between the traditional regression-based and GA-based models for stock scoring and selection. Based on the designed scoring mechanism for a set of stocks, top-ranked stocks can be selected as components in a portfolio. In the mean time, the GA was employed for optimization of model parameters and feature selection. We have evaluated the models statistically and validated the effectiveness of this method by comparing the investment return of the models with that of the benchmark. The empirical results showed that the investment return provided by our proposed GA-based methodology can significantly outperform the benchmark and the regression-based methods. Therefore, we expect this ML-based model to advance the research in computational finance and provide a promising solution to stock selection in practice.

In the future, a plausible research direction is to develop more advanced scoring models to investigate how performance of portfolios can be further improved. In addition, in our current model, we consider the first several years as the training set and the next several years as the test set. This may not be sufficient to generate a feasible model; e.g., it might be difficult to infer plausible scores for stocks in the period from 1996 till 2009 if the model uses only the stocks of 1995 for training. Therefore, in the future work, we intend to study a model that would be capable of generating more time-dependent patterns to account for the impact of the more recent past of the stocks.

ACKNOWLEDGMENT

This work is fully supported by the National Science Council, Taiwan, Republic of China, under grant number NSC 100-2221-E-390 -033.

REFERENCES

[1] S.C. Andrade, C. Chang, and M.S. Seasholes, “Trading imbalances, predictable reversals, and cross-stock price pressure,” *Journal of Financial Economics*, Vol. 88, No. 2, , 2008, pp. 406–423.

[2] M.A. Ferreira, and P. Santa-Clara, “Forecasting stock market returns: The sum of the parts is more than the whole,” *Journal of Financial Economics*, Vol. 100, No. 3, , 2011, pp. 514-537.

[3] J. L. Evans and S. H. Archer, “Diversification and the Reduction of Dispersion: An Empirical Analysis,” *Journal of Finance*, Vol. 23, No. 5, 1968, pp. 761-767.

[4] E. J. Elton and M. J. Gruber, “Risk Reduction and Portfolio Size: An Analytical Solution,” *Journal of Business*, Vol. 50, No. 4, 1977, pp. 415-437.

[5] J. Piotroski, “Value investing: the use of historical financial statement information to separate winners from losers,” *Journal of Accounting Research*, Vol. 38, 2001, pp.1–41.

[6] A. Mochón, D. Quintana, Y. Sáez, “Soft computing techniques applied to finance”, *Applied Intelligence*, Vol.29, No. 2, 2008, pp.111-115.

[7] D. Zhang, L. Zhou, “Discovering golden nuggets: data mining in financial application”, *IEEE Transactions on Systems, Man, and Cybernetics—Part C: Applications and Review*, Vol.34, No. 4, 2004, pp. 513–522.

[8] T. S. Quah, B. Srinivasan, “Improving returns on stock investment through neural network selection”, *Expert Systems with Applications*, Vol.17, 1999, pp. 295–301.

[9] N. Chapados, Y. Bengio, “Cost functions and model combination for VaR-based asset allocation using neural networks”, *IEEE Transactions on Neural Networks*, Vol. 12, 2001, pp. 890–906.

[10] K. J. Kim, I. Han, “Genetic algorithms approach to feature discretization in artificial neural networks for the prediction of stock price index”, *Expert Systems with Applications*, Vol.19, No. 2, 2000, pp. 125–132.

[11] M. Caplan, Y. Becker, “Lessons learned using genetic programming in a stock picking context”, in: U.-M. O'Reilly, T. Yu, R. Riolo, B. Worzel (Eds.), *Genetic Programming Theory and Practice II*, Springer, Ann Arbor, Michigan, 2004, pp. 87–102, Chapter 6.

[12] Y. Becker, P. Fei, A. Lester, “Stock selection – An innovative application of genetic programming methodology”, in: R. Riolo, T. Soule, B. Worzel (Eds.), *Genetic Programming Theory and Practice IV, Genetic and Evolutionary Computation*, vol. 5, Springer, Ann Arbor, Michigan, 2006, pp. 315– 334, Chapter 12.

[13] T. C. Chu, C. T. Tsao, Y. R. Shiue, “Application of fuzzy multiple attribute decision making on company analysis for stock selection”, in: *Proceedings of Soft Computing on Intelligent Systems and Information Processing*, 1996, pp. 509–514.

[14] M. R. Zargham, M. R. Sayeh, “A web-based information system for stock selection and evaluation”, in: *Proceedings of the First International Workshop on Advance Issues of E-Commerce and Web-Based Information Systems*, 1999, pp. 81-83.

[15] J. Yang, V. Honavar, “Feature subset selection using a genetic algorithm”, *IEEE Intelligent Systems*, Vol. 13, No. 2, 1998, pp. 44-49.

[16] Huang, C.-F., Chang, C.-H., Chang, B. R. and Cheng, D.-W, “A Study of a Hybrid Evolutionary Fuzzy Model for Stock Selection,” *Proceeding of the 2011 IEEE International Conference on Fuzzy Systems*, Taipei, Taiwan, June 27-30, 2011, pp. 210-217.

[17] S. Mukherji, M. S. Dhatt, Y. H. Kim, “A fundamental analysis of Korean stock returns,” *Financial Analysts Journal*, Vol. 53, No. 3, 1997, pp. 75-80.

[18] G. R. Jensen, R. R. Johnson, J. M. Mercer, “New evidence on size and price-to-book effects in stock returns,” *Financial Analysts Journal*, Vol. 53, No. 6, 1997, pp. 34-42.

[19] M. G. Danielson, T. D. Dowdell, “The return-stages valuation model and the expectations within a firm’s P/B and P/E ratios,” *Financial Management*, Vol. 30, No. 2, 2001, pp. 93-124.

[20] J. Lewellen, “Predicting returns with financial ratio,” *Journal of Financial Economics*, Vol. 74, No. 2, 2004, pp. 209-235.

[21] E. F. Fama, K. R. French, “Average returns, B/M, and share issue,” *Journal of Finance*, Vol. 63, No. 6, 2008, pp. 2971-2995.

[22] E. Hjalmarsson, “Predicting global stock returns,” *Journal of Financial and Quantitative Analysis*, Vol. 45, No. 1, 2010, pp. 49-80.

[23] M. Omran, “Linear versus non-linear relationships between financial ratios and stock returns: empirical evidence from Egyptian firms,” *Review of Accounting and Finance*, Vol. 3, No. 2, 2004, pp. 84-102.

[24] R. Bauer, N. Guenster, R. Otten, “Empirical evidence on corporate governance in Europe: the effect on stock returns, firm value and performance,” *Journal of Asset Management*, Vol. 5, No. 2, 2004, pp. 91-104.

[25] M. T. Soliman, “The use of DuPont analysis by market Participants,” *The Accounting Review*, Vol. 83, No. 3, 2008, pp. 823-853.

[26] W. E. Ferson and C. R. Harvey, “Fundamental determinants of national equity market returns: A perspective on conditional asset pricing,” *Journal of Banking & Finance*, Vol. 21, 1998, pp. 1625–1665.

[27] T. A. Carnes, “Unexpected Changes in quarterly financial-statement line items and their relationship to stock prices,” *Academy of Accounting and Financial Studies Journal*, Vol. 10, No. 3, 2006, pp. 99-116.

[28] D. Ikenberry, J. Lakonishok, “Corporate governance through the proxy contest: evidence and implications,” *Journal of Business*, Vol. 66, No. 3, 1993, pp. 405-435.

[29] G. Sadka, R. Sadka, “Predictability and the earnings-returns relation,” *Journal of Financial Economics*, Vol. 94, No. 1, 2009, pp. 87-93.