FORECASTING EQUITY FUND PERFORMANCE VIA GA

SHUENN-REN CHENG¹, JUEI-CHAO CHEN²*, wen-hung wu³ and ya-tzu liu²

¹Department of Business Administration
Cheng Shiu University
Kaohsiung, Taiwan
tommy@csu.edu.tw

²Institute of Applied Statistics
Fu Jen Catholic University
Taipei, Taiwan
006684@mail.fju.edu.tw

*Corresponding author: yatzu0122@yahoo.com.tw

³Department of Business Administration
Kang-Ning Junior College of Medical Care and Management
Taipei, Taiwan
wu410226@knjc.edu.tw

Received August 2009; accepted November 2009

Abstract. The purpose of this article is to establish a mutual fund return rate forecasting model to forecast the return rate of mutual funds. This study is based on "Jensen performance model" which Jensen (1968) extend from CAPM. Correlation analysis was used to find factors that may influence mutual fund’s net return rate and added into the regression model as independent variables. The least square method was used to estimate parameters of the forecasting model. On the other hand, the regression model was used as the fitness function of Genetic Algorithm (GA) to estimate the regression parameters. The results discovered that the results from GA have better forecasting ability, especially since its Root Mean Square Error (RMAE) and Mean Absolute Error (MAE) were better than the least square method’s.

Keywords: Data mining, Genetic algorithm, Intelligent data analysis, Root mean square error

1. Introduction. In a liberalize and globalize investment market, aside from having more opportunity to obtain high return, investors at the same time will also face lots of investment risks. Therefore, how to obtain the balance in return and risk became one of the subjects investors are concerned with. Besides supplementing insufficient professional investment knowledge of every individual investor, collecting the capital of many investors and undergoing "mutual fund" of specialized management through fund managers with professional knowledge can also solve the limitation of inability to adequately segregate risk due to insufficient capital. The numbers of mutual funds increase everyday, investors now faced dilemmas in selecting mutual funds. In an international setting, because investors from different countries have their own personal concerns on fund performances, researches and forecasts of fund performance became important to many scholars. For example, Droms and Walker [1] used the international equity funds of US as subject in the study of performance persistence. In the Australian setting, Gallagher and Jarneicic [2] researched on the dynamics between performance of international equity funds and investor flows. Benson and Faff [3] evaluated on the performance of international equity funds and further researched on the relationship between the performance of international equity funds and money flows. Chen et al. [4] adopted logistic regression to forecast Taiwan’s international equity funds performances.
The purpose of this study in establishing a fund performance forecasting model is to use the model proposed by Jensen [5] as foundation and using correlation analysis to find the important factors that influence mutual fund performance. Base on the factors to build a regression model as the forecasting model then apply the least square method and genetic algorithm to find estimators of parameters, separately. Last, the RMAE and MAE were applied to the results to evaluate forecasting ability. The structure of this article is divided into 5 sections. Section 2 shows the research subjects and lecture review, Section 3 shows the variables, Section 4 shows methodology. Results and the summary and conclusions are shown in the last section.

2. Research Subjects and Lecture Review.

2.1. Research subjects and periods. The research subject is Taiwan’s open-end equity fund and used the investment area as the international equity fund of the multinational market. In addition, following the standards of the international fund evaluation institutions such as Standard & Poor’s Fund Services and Lipper, this study only used the funds that are established for more than three years. The sources of data are the basic information of funds from the website of the Securities Investment Trust & Consulting Association of Taiwan and the database of Taiwan Economic Journal (TEJ). The research period is from January, 2001 to September, 2007. The data from January, 2001 to July, 2007 was used as training sample in establishing forecasting model and the data from August and September, 2007 was used as the validation sample. If there are new funds and fund transformation found within the period of the study, they will not be included in the subject of the study because of incomplete data.

2.2. Related researches on evaluating mutual fund performance. Markowitz [6] proposed that Mean-Variance Portfolio Model (MV model) initiated portfolio theory. In 1960, Sharpe, Lintner and Treynor proposed the Capital Asset Pricing Model (CAPM). $E[R_i] = R_f + \beta_i(E[R_m] - R_f)$, where $R_i$ = return of $i$th fund, $R_f$ = risk free interest rate, $R_m$ = returns of market portfolio, $\beta_i$ = systematic risk of $i$th fund, $R_m - R_f$ = market risk premium (represents the compensation of the investor’s load in market risk), $\beta_i(R_m - R_f)$ = risk premium of $i$th fund.

In the foundation of CAPM theory, Treynor [7], Sharpe [8] and Jensen [5] separately establish performance measures among which, the performance measures of Treynor and Sharpe can be the relative measures of performance and not an absolute measures of performance evaluation. The performance measure model proposed by Jensen [5] is $R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + \varepsilon_i$, where $\alpha_i$ = Jensen measure of the $i$th fund portfolio, $\varepsilon_i$ = random errors of $i$th fund return. For example, if all portfolio are under complete risk segregation, Jensen measure that are greater than 0 means good forecasting ability and it can earn higher returns otherwise, the forecasting ability is bad.

Droms and Walker [1] used the cross-section/time series analysis to analyze long-run mutual fund investment performance. This study considered the relationship among the factors asset size, expense ratios, portfolio turnover, and load status and investment performance. The results in the study discovered that only expense and returns are related and higher expenses are associated with higher returns.

2.3. Related researches on factors influencing mutual fund performance. There are a lot of factors that influence returns such as market factors, fund size, performance before funds and different expenses. Black, Jensen, and Scholes [9] found the factors that influence securities portfolio. The results showed that the risk factors that influence returns were market factors, fund size and book-to-market equity. The three-factor model of Fama and French [10] is $R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + s_iSMB + h_iHML + \varepsilon_i$, where $SMB(Small Minus Big)$ = the difference between the return on a portfolio of small stocks
and the return on a portfolio of large stocks, \( HML \) (High Minus Low) = the difference between the return on a portfolio of high-book-to-market stocks and the return on a portfolio of low-book-to-market stocks. Carhart [11] used the three-factor Fama-French model as the foundation and added the factor momentum turning the model into a four-factor model

\[
R_i - R_f = \alpha_i + \beta_i(R_m - R_f) + \gamma_iSMB + \delta_iHML + p_iPR1YR + \varepsilon_i,
\]

where \( PR1YR \) = one-year momentum in stock returns.

Grinblatt and Titman [12] aimed at the monthly Jensen measure of USA’s 279 funds from 1974 to 1984 and analyzed the relationships among net asset value, load, expenses, turnover and management fee and fund performance using Cross-Sectional Regression Model. The results showed that turnover is significantly and positively related to the ability of fund managers to earn abnormal returns.

Volkman and Wohar [13] aimed at the USA’s 332 funds from October, 1980 to December, 1989 as monthly data and used regression model to analyze the relation between persistent fund performance and past fund performance, size, goal, load, and management fee. The results showed that excessive fund size would produce low efficiency. Small fund size doesn’t possess Economies of Scale which makes risk exist easily. Therefore, the performance of medium size factor model is persistent and good performance means that it can continue to the next term. In addition, management fee and fund performance are negatively related. The results of Carhart [11] showed that expenses and turnovers are negatively related to fund performance.

There were scholars who researched on the effect of fees towards the mutual fund from different angles. For example, Shu, Yeh and Yamada [14], with Taiwan’s mutual funds as subjects and performance measure, average size of account per investor, average turnover ratio and management fee ratio as control variables, established a regression model to forecast mutual fund flows. Barber, Odean and Zheng [15] researched on the mutual funds investors’ behavior. The results of their study showed that investors are also influenced by fees such as in-your-face fees, operating expense, front-end loads and commissions.

2.4. Forecasting methods. Many scholars studied on the forecast of mutual funds performance and the methods used were correlation analysis, cross-section regression, time series analysis, logistic regression or regression analysis with Bayesian approach. Ciccotello and Grant [16] believed that mutual funds with superior historical returns easily attract investors to buy so the growth of the mutual fund size speeds up. Therefore, studying the relationship between equity fund size and performance can be used to forecast the performance of mutual funds.

Ahmed [17] pointed out that to segregate risk, the investors of mutual funds usually buys two or more mutual funds from similar industry. Therefore, the correlation between mutual funds plays quite an important role. Ahmed used eight models to forecast the relationship between mutual finds and discovered that Fama-French three-factors model have the lowest forecasting errors. Avramov and Wermers [18] used the factors manager skill, fund risk loadings and benchmark returns to forecast the returns of the US domestic equity mutual funds. If there is information of business cycle, the effect of the forecast will be better.

Pastor and Stambaugh [19] believed that investors of mutual funds would consider the highest Sharpe ratio in selecting mutual funds. In addition, management skill is also an important factor in investment decision. These two are the prior information. Therefore, Pastor and Stambaugh used multivariate regression and added the prior information in a Bayesian angle to find the optimal portfolios.

Chen, et. al. [4] The purpose is to establish a fund performance classification model. In this study is to use the model proposed by Jensen [5] as a foundation and the CHAID
(Chi-Square Automatic Interaction Detection) analysis to find the interactions among the fund performance and possible influencing factors. Thereafter, a new fund performance model is established using logistic regression. The model will be used to forecast the probability of positive or negative return of the funds. The positive/negative return of fund performance considered by this study is the return of the $i$th fund subtracted by the risk free interest rate.

3. **Variables.** This study used the ”equity return of mutual fund” to represent the portfolio return. Fund size and transaction cost are considered as the factors that influence fund performance. The fund size of this study first underwent the stratification method of Dalenius and Hodges [20] to find the appropriate cutoff points and separate the fund size into big size, medium size and small size.

With regards to the factors of transaction cost, the variables used are the variables organized from the studies of past scholars. The variables considered by this study are 1. Charge rate ($Cr$) = (Charge Fee/Net Assets)$\times$ 100%, 2. Exchange rate ($Ec$) = (Transaction Tax/Net Assets)$\times$ 100%, 3. Management rate ($Mg$) = (Management fee/Net Assets)$\times$ 100%, 4. Storage rate ($Sr$) = (Storage Fee/Net Assets)$\times$ 100%, 5. Buy Turnover rate ($BTurn$) and 6. Sell Turnover rate ($STurn$).

4. **Methodology and Results.** Using correlation analysis to find important factors that have influence mutual fund performance have statistically and significantly related variables. Then, these variables are used as independent variables in the regression model. The obtained regression model can be used to forecast the return rate in the future.

4.1. **Correlation analysis.** The Pearson Correlation coefficients between each variable were first computed and the Pearson Correlation matrix is shown in Table 1. Table 1 shows that the Pearson Correlation of the index return of S&P 500 ($R_i$), the ten-year government bond yield in U.S.A. ($R_f$), Management rate ($Mg$), Storage rate ($Sr$) and Buy Turnover rate ($BTurn$) and fund’s net worth return are statistically significant. The correlation coefficients between management rate, storage rate and fund’s net worth return are $-0.248$ and $-0.157$ respectively showing that the two factors are statistically and negatively related to fund’s net worth return. The correlation coefficients among the ten-year government bond yield in U.S.A., index return of S&P 500 and fund’s net worth return are $-0.140$ and $0.621$ respectively showing that the two factors are statistically related to fund’s net worth return. In Table 1, ** indicates $p$ – value < .01; * indicates $p$ – value < .05.

<table>
<thead>
<tr>
<th></th>
<th>$R_i$</th>
<th>$Cr$</th>
<th>$Ec$</th>
<th>$Mg$</th>
<th>$Sr$</th>
<th>$BTurn$</th>
<th>$STurn$</th>
<th>$R_m$</th>
<th>$R_f$</th>
</tr>
</thead>
<tbody>
<tr>
<td>$R_i$</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Cr$</td>
<td>-0.030</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Ec$</td>
<td>-0.008</td>
<td>0.542**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Mg$</td>
<td>-0.248**</td>
<td>0.138**</td>
<td>0.197**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$Sr$</td>
<td>-0.157**</td>
<td>-0.043*</td>
<td>0.020</td>
<td>0.421**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$BTurn$</td>
<td>0.021</td>
<td>0.465**</td>
<td>0.309**</td>
<td>-0.036</td>
<td>-0.056**</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>$STurn$</td>
<td>-0.022</td>
<td>0.524**</td>
<td>0.385**</td>
<td>0.027</td>
<td>-0.017</td>
<td>0.900**</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>$R_m$</td>
<td>0.621**</td>
<td>-0.045*</td>
<td>-0.048*</td>
<td>-0.165**</td>
<td>-0.105**</td>
<td>-0.019</td>
<td>-0.047*</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>$R_f$</td>
<td>-0.140**</td>
<td>0.044*</td>
<td>-0.002</td>
<td>-0.047*</td>
<td>-0.039*</td>
<td>-0.028</td>
<td>-0.030</td>
<td>0.196**</td>
<td>1.000</td>
</tr>
</tbody>
</table>
4.2. Regression analysis. Using the model proposed by Jensen [5] as foundation and considering the significant variables of correlation analysis. The data from January, 2001 to July, 2007 were used to undergo regression analysis where \( R_i - R_f \) is the dependent variable and \( R_m - R_f \), \( M_g \), \( Sr \), \( BTurn \) are the independent variables and the \( R^2 \) is 99.2%. Therefore, the model was used to forecast the return rate of August and September 2007. The regression model is

\[
R_i - R_f = 0.72 + .998(R_m - R_f) - .562(M_g),
\]

where \( R_m - R_f = \) market risk premium and \( M_g = Management\ rate.\)

In the past studies, the result of Volkman and Wohar’s [13] study showed that management fees and persistent fund performance are negatively related. The result of Carhart’s [11] study showed that expense ratios are significantly and negatively related to performance. These results and the results discovered in this study are consistent.

4.3. Genetic algorithm. This study used the most common binary code \{0, 1\} method to undergo Genetic Algorithm, also known as Binary Genetic Algorithms. The reason is that binary code is more suitable in solving numerical problems. Therefore, the initial population was represented by random binary string. For the detailed process of Genetic Algorithm, please refer to Hastie, Tibshirani, and Friedman [21] or Berthold and Hand [22]. Genetic Algorithm has been widely used in the researches and practices of different fields. For example, Wang [23] applied Genetic Algorithm in clustering analysis to prevent the problem of local optimal in the traditional method. Gao et al. [24] adopted Genetic Algorithm in the scheduled fuzzy controller. Xhafa and Carretero [25] also applied Genetic Algorithm to solve for problems regarding the schedulers for grid computing system. Nakajima et al. [26] applied Genetic Algorithm to solve the measurement of the value of translation and rotation between the corresponding teeth digital dental models before/after an orthodontic treatment.

According to the suggestions of Srinivas and Patnaik [27] and Gen and Cheng [28], the crossover rate should be between .5 and 1.0 and the mutation rate should be between .001 and .05. In this article, the set values for population size were 50, the crossover rate is .8, mutation rate is .05 and stopping conditions was 500. The fitness function and variable range of the Genetic Algorithm is

\[
\max R_{it} - R_{jf} = \alpha_i + \beta_{i1}(R_{mt} - R_{ft}) + \beta_{i2}(Mg_{it}),
\]

where: \(0 < \alpha < .5\), \(8 < \beta_{i1} < 1.2\) and \(-.85 < \beta_{i2} < -.55\).

4.4. Forecasting. This study used the 64 data in August and September, 2007 as the validation sample of the tested model’s forecasting ability. The forecasted values and actual values’ Root Mean Square Error (RMSE) and Mean Absolute Error (MAE) of the least square method and Genetic Algorithm was used to test the accuracy of the mutual fund return rate forecasting model established by this study and the forecasting abilities of the two methods. The equation of \( RMSE \) and \( MAE \) are

\[
RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^{n} (y_i - \hat{y}_i)^2}
\]

and

\[
MAE = \frac{1}{n} \sum_{i=1}^{n} |y_i - \hat{y}_i|,
\]

where: \( y_i = \) actual value, \( \hat{y}_i = \) forecasted value, and \( n = \) sample size.

The \( RMSE \) and \( MAE \) of Genetic Algorithm are .0126 and .0781, respectively, and they were lower than the \( RMSE = .0398 \) and \( MAE = .1654 \) of the least square method. The results showed that Genetic Algorithm has better forecasting ability with regards to the estimated parameter of the forecasting model.

5. Summary and Conclusions. The purpose of this study is to establish a mutual fund return rate forecasting model to forecast the return rate of mutual funds. The research results can provide references for investors during mutual fund selection. First, correlation analysis was used to find possible important factors that will influence mutual
fund performance and the "Jensen performance model" which Jensen [5] extend from CAPM. Regression analysis was used to establish the forecasting model with \( R_t - R_f \) as the dependent variable and \( R_m - R_f, M_g \) as the independent variable and separately used the least square method and Genetic Algorithm parameter estimation to find better mutual fund return rate forecasting model. In reality, using Taiwan's data from January, 2001 to July, 2007, two methods were used to estimate parameters of the forecasting model and the data from August to September, 2007 were used for validation.

The empirical results discovered that in the forecasted results of data from August to September, 2007, the calculated RMSE of the Genetic Algorithm was .0126 which was lower that the RMSE \( (= .0398) \) of the least square method. Therefore, the forecasting ability of Genetic Algorithm in the estimation of forecasting model was better than the least square method’s. In actual applications, adopting Genetic Algorithm to estimate parameters is suggested if the data has large variations or the user is not confident with the results of least square method.

Acknowledgment. The research of Juei-Chao Chen is support by Taiwan’s National Science Council under the contract number NSC 95-2118-M-030-003. The authors also gratefully acknowledge the helpful comments and suggestions of the reviewers, which have improved the presentation.

REFERENCES


