PORTFOLIO SELECTION AND MANAGEMENT USING A HYBRID INTELLIGENT AND STATISTICAL SYSTEM

Juan G. Lazo Lazo, Marco Aurélio C. Pacheco, Marley Maria R. Vellasco[1]

ABSTRACT

This paper presents the development of a hybrid system based on Genetic Algorithms, Neural Networks and the GARCH model for the selection of stocks and the management of investment portfolios. The hybrid system comprises four modules: a genetic algorithm for selecting the assets that will form the investment portfolio, the GARCH model for forecasting stock volatility, a neural network for predicting asset returns for the portfolio, and another genetic algorithm for determining the optimal weights for each asset. Portfolio management has consisted of weekly updates over a period of 49 weeks.

Key words: Genetic Algorithms, Neural Networks, GARCH, VaR, Volatility

INTRODUCTION

This paper makes use of a Genetic Algorithm (GA) [1][2] to select the assets that will comprise the portfolio, based on a subset of assets that are traded on the São Paulo Stock Exchange - Brazil (BOVESPA). A Neural Network [3][4] contributes to portfolio management by predicting the asset returns for the next period of portfolio evaluation [5]. The GARCH model [6][7] is employed for predicting the volatility of each asset. Another GA is used for optimizing the assets within the portfolio by estimating the Value at Risk (VaR) [8] on a weekly basis. The portfolio is managed for a period of 49 weeks and its evaluation consists of comparing its behavior with that of the market - the BOVESPA Index. In this paper, Markowitz's model and the Efficient Frontier model [9][10] have been used to perform the selection and to determine the percentage represented by each asset in the portfolio.

The system is evaluated with the use of the return series of 137 Brazilian assets traded on the BOVESPA between July 1994 and December 1998. One part of the data was used for training the model and the other part (January 1998 to December 1998) was used for testing, i.e., for portfolio management.

The system that has been developed is of an academic nature and does not purport to be a tool to be used by individual investors. Its purpose is to evaluate the performance of GA and of Neural Networks in portfolio construction and management in emerging markets (the Brazilian Market).

Section 1 describes the modeling of the GA that has been employed for selecting the assets that make up the portfolio. Section 2 presents the volatility forecasts performed by the GARCH model. In section 3, the returns are predicted by the neural networks. Section 4 describes the modeling of the GA that has been developed for weight optimization and for portfolio management via the results obtained by the neural nets, while section 5 presents the calculation of the VaR for the portfolio. The results obtained with the proposed hybrid model are analyzed in section 6, and finally, section 7 presents the conclusions that have been drawn from this work.

CONSTRUCTION OF THE INVESTMENT PORTFOLIO BY GENETIC ALGORITHMS

In this model, it is the genetic algorithm that, by means of the Efficient Frontier Criterion, selects and determines which and how many of the 137 assets traded on the BOVESPA will form the investment portfolio.

To this end, the monthly return rates and risk are calculated for each one of the 137 assets in accordance with the Mean-Variance criterion proposed by Markowitz [9], where the return estimate is represented by the mean and asset risk is represented by the variance.

Basically, the problem contemplates an initial portfolio comprised of 137 assets and the GA must determine the percentage to be invested in each asset, which is also called asset weight. The GA must confer a significant weight to all the assets that are expected to yield profits and will confer zero weight to assets that do not comply with this condition. In other words, it will attempt to obtain the optimal portfolio according to the Efficient Frontier Criterion in order to minimize the risk for the portfolio. The following constraints must be met: the sum of all the weights must be equal to 1, and the weight attributed to an asset must be greater than or equal to zero.

The representation of the chromosome comprises 137 genes, (Figure 1), where each gene represents the weight of the asset in the portfolio.

1	2	3		136	137
Asset N	Asset A	Asset Q	.....	Asset D	Asset X

Figure 1 - Chromosome for asset selection.

When the chromosome is evaluated, an attempt is made to minimize the portfolio's risk defined by equation (1), which represents the standard deviation of the portfolio's returns.

$$Min \quad \sum_{i=1}^{N} X_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} X_i X_j \sigma_{ij} \qquad (1)$$

where Xi is the weight for the asset that corresponds to gene i of the chromosome; Xj is the weight for the asset that corresponds to gene j of the chromosome; si is the risk that pertains to the asset that corresponds

[1] ICA : Applied Computational Intelligence Laboratory. Department of Electrical Engineering.
Pontifícal Catholic University of Rio de Janeiro, PUC-Rio, Rua Marquês de S. Vicente 225, Gávea, Rio de Janeiro, CEP 22453-900, RJ, Brazil. Email: {juan, marco, marley@ele.puc-rio.br}

to gene i of the chromosome, i.e., the standard deviation of asset i; and sij is the covariance of the asset that corresponds to gene i with the asset that corresponds to gene j of the chromosome.

Results Obtained

In every run, the GA converges to the same result, giving significant weights to the same 13 assets listed below:

Aços Vilares Pn, Albarus On, Bemge On, Brahma On, Brahma Pn, Cemig Pn, Ciquine Pna, Docas Pn, Electrolux Pn, FrasLe Pna, Light On, Telerj On, Unibanco On

In order to evaluate the results of this GA, the selected assets were used for setting up and managing a portfolio with a view to the maximization of portfolio returns. This portfolio was managed over a period of 49 weeks, from February 1998 to March 1999, and the asset weights were updated each week with the use of the Mean-Variance Criterion and the Efficient Frontier [9]. In other words, the estimated returns on each asset for the next period are given by the mean, and the risk is the variance-covariance matrix of the returns. The sample window for the re-estimations was of 6 months. The performance of the portfolios over the managed period is compared with the BOVESPA market index. Table 1 presents a few comparative measurements of the performance of the managed portfolios, such as the Mean Return (mean weekly portfolio returns) and the variance (risk for the portfolio).

	Market Portfolio	Market Portfolio Maximizes Return
Mean Return (%)	-0.46	3.12
Variance (%)	56.302	165.233

Table 1 - Comparison between the performance of the managed portfolio and the market portfolio.

Table 1 demonstrate that the managed portfolio has obtained better returns than the market portfolio, although at a much higher risk. However, this was espected because, once the objective was to maximize returns, there were no restrictions on risk. Besides, the market in this period presented a negative return and a high risk. This all suggests that the selected assets may generate profits.

VOLATILITY FORECASTING BY THE GARCH MODEL

The term volatility denotes the temporal variability of the degree in which the data scatter around their central trend. Hence, volatility is the variation, along the time horizon, of conditional mean. Volatility measures the risk that is associated with a specific series.

This paper has made use of the GARCH model (Generalized Autoregressive Conditional Heteroscedasticity Model) [6][7] to perform the volatility forecasts. On account of the results obtained, the representation of the GARCH(1,1) was elected, in agreement with the type of representation that is most commonly suggested for financial series in the bibliography [8][11][12].

FINANCIAL ASSET RETURN FORECASTING

In order to manage the portfolio, it is necessary to have estimates of the returns for the next period to be managed. This paper has made use of Neural Networks to make predictions of the returns by considering historical information on the returns of each asset and their volatility, which has been calculated by means of the GARCH model. Since the results obtained by the Backpropagation neural network [3] were satisfactory in terms of producing smaller prediction errors, it was decided that neural nets would be employed for obtaining return forecasts.

The Neural Networks Approach to Financial Asset Return Modeling

The neural network is formed by 11 inputs (the 10 previous values of the return series of the asset plus a value that corresponds to the volatility value that has been calculated with the GARCH model), with one hidden layer and one output; since each series presents distinct features, the quantity of neurons in the hidden layer is different for each return series. The activation function of the neurons in the hidden layer is the hyperbolic tangent, while the output activation function is linear; the forecast is made one step ahead.

The number of inputs for the neural network was determined experimentally, based on the auto-correlation analysis performed on the series, of the square of the returns, which presented several significant lags among the first 10 lags. This auto-correlation analysis was performed because the other input of the neural network is volatility. Since correlation is a linear measurement and neural networks have a nonlinear nature, it is expected that, in the tests performed, the neural network will find some type of nonlinear relation between the past data of the return series and the past data of asset volatility.

In order to evaluate the forecasting results, the following error measurements have been employed:

MAD: Mean absolute deviation;
NRMSE: Normalized root mean square error;
MSE: Mean square error;
U-Theil: Metric that measures the extent to which a result is better than one obtained by means of naive prediction.

Table 2 below presents the prediction error statistics for a few of the predicted series.

	A.Vilares Pn	Albarus On	Bemge On	Brahma On
MAD	0.1159	0.0999	0.5206	0.0564
MSE	0.0254	0.0195	0.7611	0.0070
RNMSE	1.4201	1.1556	3.8683	1.0578
U-THEIL	0.6697	0.7517	0.7005	0.8154

Table 2 - Prediction errors.

In table 2, it may be observed that the prediction errors of the neural net are small, and the U-Theil statistic indicates that the forecasts obtained are much better than those of the naive prediction.

PORTFOLIO MANAGEMENT BY GENETIC ALGORITHMS

The portfolio has been managed by means of the evaluation of its performance over a period of 49 weeks with weekly updates of asset weights and of the return and risk estimates for each week with the use of the predictions of returns and volatility. A GA has been used for the purpose of determining the percent-tage of the fund to be invested in each of the assets in the portfolio.

This paper has opted for the management of two portfolios. Thus, the basic problem is to find the asset weights that will allow one portfolio to maximize the return and the other, to minimize the risk. Both portfolios must meet the following constraints: The sum of the portfolio weights must be equal to 1 and the weight of each asset must be greater than or equal to zero.

The chromosome for managing the investment portfolio is formed by 13 genes, which represent the weights of the assets in the portfolio (Figure 2).

1	2	3	4	......	11	12	13

Figure 2 – Chromosome.

Since the two portfolios to be managed have different objectives, each portfolio presents its own evaluation function. In the first case, the evaluation function of the chromosome attempts to maximize the portfolio's returns and is defined by equation (2):

$$MAX \ \ \theta = \frac{\sum_{i=1}^{N} X_i (\bar{R}_i - R_F)}{\left[\sum_{i=1}^{N} X_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} X_i X_j \sigma_{ij} \right]^{1/2}}$$

where Xi is the weight for the asset that corresponds to gene i of the chromosome; Xj is the weight for the asset that corresponds to gene j of the chromosome; Ri represents the predicted return on the asset that corresponds to gene i of the chromosome; RF is the return of a risk-free asset; si is the risk that pertains to the asset that corresponds to gene i of the chromosome (predicted volatility for asset i); sij is the covariance of the asset that corresponds to gene i with the asset that corresponds to gene j of the chromosome.

This function attempts to maximize portfolio returns by maximizing the Sharpe ratio [10]. The numerator of the formula expresses the portfolio return that exceeds the return of a risk-free asset, which is the return obtained from an asset that pays known interest rates. This paper has used the Brazilian CDI (Interbank Deposit Certificate) as the risk-free asset.

For the second case, the evaluation function of the chromosome tries to minimize the portfolio's risk for a given return; the function is defined by equation(3), which represents the standard deviation of the portfolio's returns:

$$MIN \sigma_P = \left[\sum_{i=1}^{N} X_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} X_i X_j \sigma_{ij} \right]^{1/2} \quad (3)$$

here σ_P represents the risk for the portfolio or its volatility.

CALCULATION OF THE VAR FOR THE PORTFOLIO

The VaR is a measure of exposure which attempts to quantify the maximum potential loss possible for a given portfolio (or asset) within a time horizon and with a specific confidence interval [8].

The time horizon is one week. The VaR for the portfolio is recalculated each week according to the updates of the portfolio's weights. Therefore, the portfolio's variance or risk ($\sigma^2{}_P$) for each period depends on the weight that has been allocated to each asset and this variance is calculated each week when the it is time to optimize the weights, equation (4). The correlation matrix needed for calculating the VaR was calculated for each week, based on the return forecasts.

$$\sigma_P^2 = \sum_{i=1}^{N} X_i^2 \sigma_i^2 + \sum_{i=1}^{N} \sum_{\substack{j=1 \\ j \neq i}}^{N} X_i X_j \sigma_{ij} \quad (4)$$

The VaR for the portfolio is calculated with the use of equation (5) for a confidence interval of 95% (a = 1.65) and for a portfolio value (VP) of 100 thousand US dollars.

$$VaR_P = -\alpha \sigma_P V_P \quad (5)$$

RESULTS

Both of the portfolios created were managed over a period of 49 weeks with weekly updates of the asset weights, and with the use of the weekly return forecasts and the volatility forecasts. The results obtained by portfolio management with the proposed model are shown in table 3.

Figures 3 and 4 present the performance of each portfolio comparing with the returns of the market portfolio (BOVESPA Index)

It may be observed that on the

	Market Portfolio	Portfolio that Maximizes Return	Portfolio that Minimizes Risk for Given Return R > 5%
Mean Return (%)	-0.221	5.227	0.54
Variance (%)	59.395	142.391	12.2094
Beta	1	0.549	0.203

Table 3 – Comparison of the results for the two managed portfolio.

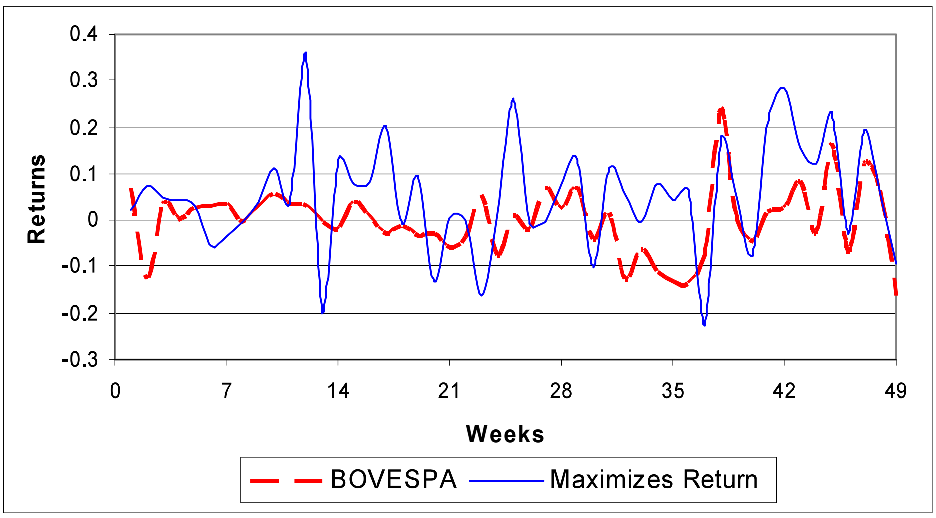

Figure 3 - Comparison between the performance of the portfolio that maximizes the return and the return and the market portfolio.

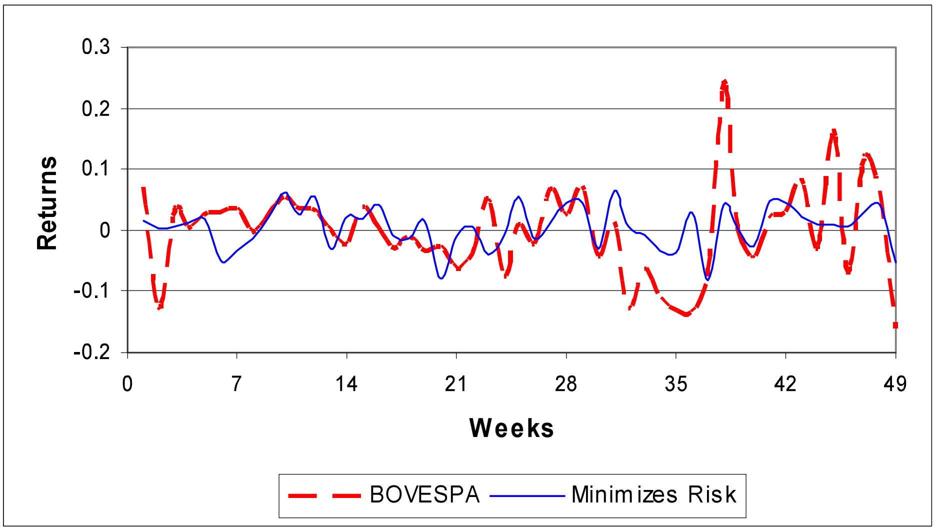

Figure 4 - Comparison between the performance of the portfolio that minimizes the risk for a given return of more than 95% and the market portfolio.

average, the returns produced by the managed portfolios are higher than the market returns and that the portfolio's risk (variance) is lower than the market's risk for the portfolio that minimizes risk, in contrast with the portfolio that maximizes return which presents a higher risk. However, this was expected once there were no restrictions on risk and also considering that the return reached was 24 times the market return.

The Beta values of the portfolios reveal that both portfolios are of a defensive nature. It may also be observed that the managed portfolios perform well during the deepest market dips. Therefore, given its constraints, this model may generate gains for the investor and possibly increase such gains by allowing a risk-free asset to be incorporated into the portfolio.

In figure 5 below, the weekly forecasts of the VaR and the value of the portfolio after the management period are compared.

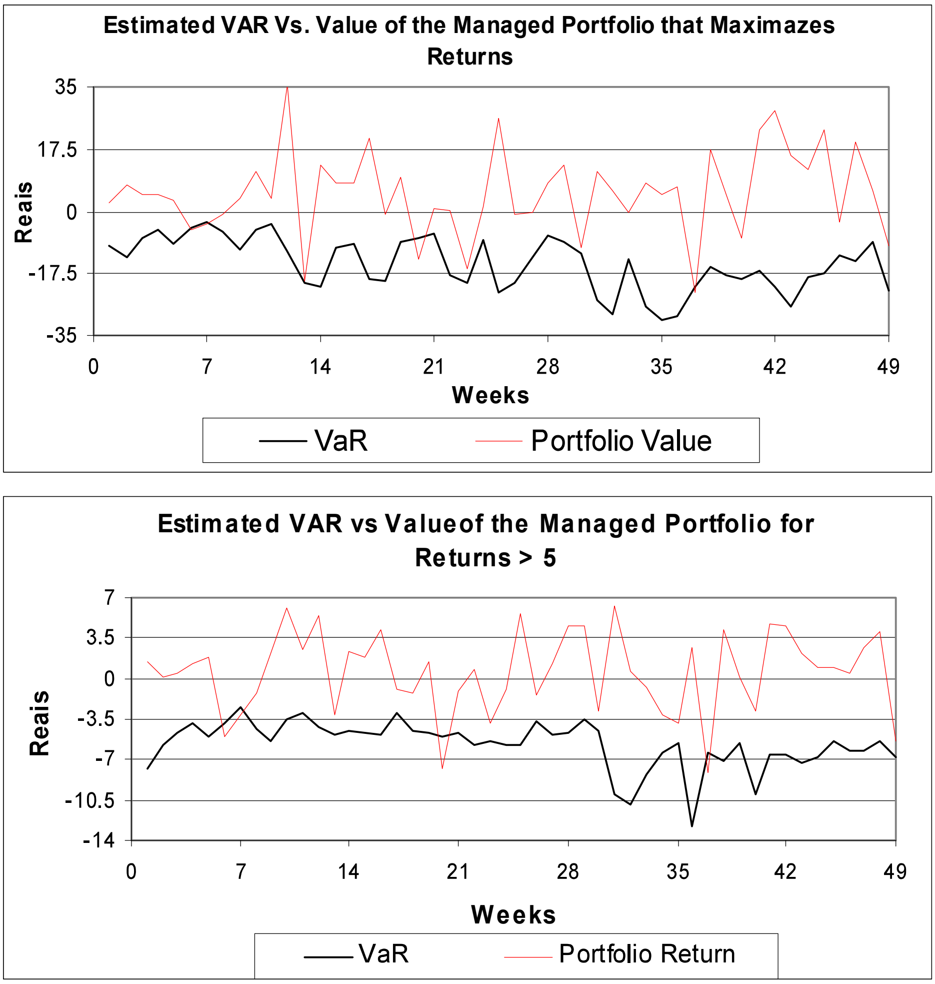

Figure 5 - Comparison between the predicted VaR for the portfolio and the value of the portfolio at the end of the management period, for each of the two managed portfolios.

It may be observed in the graphs above that the predicted value of the potential loss for the portfolio (VaR) for a confidence interval of 95% was surpassed by the portfolio, in the worst case, only in 3 weeks throughout the entire management period, in other words, the model is able to make good predictions of the VaR for the portfolio.

CONCLUSIONS

In general, the results of the tests performed with the proposed model that uses GA for selecting the assets in the portfolio proved to be satisfactory (table 1) since the selected assets were capable of generating profit for the investor.

It should be pointed out that, in the course of the 49 weeks that were selected for portfolio management, the Brazilian market presented periods of greater volatility. It is more difficult to make forecasts during such periods.

The GARCH model once more proved to be efficient in volatility forecasting since this type of forecast plays an important part in the modeling of the neural nets that provide return forecasts and of the GA for weight optimization, both of which are employed when calculating the VaR for the portfolio.

It has been demonstrated that the return forecasts were considerably improved by the fact that the asset volatility forecasts produced by the GARCH model were introduced into the modeling of the neural network for return forecasts.

This study has shown how the proposed system is able to perform good forecasts of the VaR for the portfolio over its management period. As a rule, the managed portfolios obtained a better performance during the deepest dips in the market.

It has been observed that on the average, the managed portfolio obtains higher returns at a lower risk than the market portfolio.

REFERENCIAS BIBLIOGRÁFICAS

1. Goldberg, David E. Genetic Algorithms in Search, Optimisation, and Machine Learning. Boston (MA): Addison-Wesley Publishing Company, Inc. 1989.

2. Michalewicz, Zbigniew. Genetic Algorithms + Data Structures = Evolution Programs. 3rd rev. ed. Springer-Verlag, USA. 1996.

3. Haykin, Simon. Neural Networks - A Comprehensive Foundation. 2nd ed. Upper Saddle River (NJ): Prentice Hall, 1998.

4. Zurada, Jacek M. Introduction to Artificial Neural System. St. Paul (MN): West Publishing Co., 1992.

5. Lazo, Juan G. L., Marley M.B.R. Vellasco and Marco Aurélio C. Pacheco. A Hybrid Genetic-Neural System for Portfolio Selection and Management. Proceedings of the Sixth International Conference on Engineering Applications of Neural Networks, 2000 July 17-19; Kingston Upon Thames, United Kingdom. Edited by Dimitris Tsaptsinos, School of Mathematics, Kingston University. 2000.

6. Engle, Robert F. Autoregressive Conditional Heteroscedasticity with Estimates of the Variance of the United Kingdom Inflation. Econometrica 1982 July, Vol.50, No 4, pp. 987-1007.

7. Bollerslev, Tim. Generalized Autoregressive Conditional Heteroscedasticity. Journal of Econometrics 1986, Vol. 31, pp. 307-327.

8. Jorion, Philippe. Value at Risk: The New Benchmark for Controlling Market Risk. New York. McGraw-Hill Companies, Inc., 1997.

9. Markowitz, Harry M. Portfolio Selection: Efficient Diversification of Investment. 2nd ed. Cambridge (MA): Black-Well, 1991.

10. Elton, Edwin J. and Martin J. Gruber. Modern Portfolio Theory and Investment Analysis. 5th ed. New York: John Wiley & Sons, Inc. 1995.

11. Drost, Feike C. and Theo Nijman E. Temporal Aggregation of GARCH Processes. Tilburg University, Center Discussion Papers, 1992, pp.9066 - 9240.

12. Nelso, Daniel B. Stationarity and Persistence in the GARCH(1,1) Model. Econometric Theory 1990, Vol. 6, pp.318-334.