AUTOREGRESSIVE NEURAL NETWORK PREDICTORS IN THE BRAZILIAN STOCK MARKET

Fábio Daros de Freitas[*]
freitas@computer.org

Alberto Ferreira De Souza[†]
alberto@inf.ufes.br

Ailson Rosetti de Almeida[‡]
ailson@ele.ufes.br

[*]Programa de Pós Graduação em Engenharia Elétrica - PPGEE,
Departamento de Engenharia Elétrica, Universidade Federal do Espírito Santo - UFES
Secretaria da Receita Federal - SRF

[†]Departamento de Informática, Universidade Federal do Espírito Santo - UFES

[‡]Departamento de Engenharia Elétrica, Universidade Federal do Espírito Santo - UFES

ABSTRACT

This work compares the performance of autoregressive neural network predictors with that of the mean and *random walk* predictors. A new autoregressive neural network predictor is proposed, in which the differences between the values of the series of returns and a determined past value are the regression variables. A large set of experiments with real data from the Brazilian stock market was used in the comparison, which included the evaluation of the normality of the errors of prediction. Our results shown that it is possible to obtain Normal prediction errors with non-Normal series of stock returns. A predictive-based portfolio optimization model that uses Normal prediction errors as risk measure is proposed to take advantage of such predictors.

KEYWORDS: Time Series Prediction, Stock Markets, Portfolio Optimization

1 INTRODUCTION

The forecast of investment returns on financial assets plays a central role in finance theory, being a decisive issue on investment selection. Practitioners use several forecast methods, each one with a specific application profile. Four such methods are currently more prominent: Fundamental Analysis, Technical Analysis, Statistical Modeling, and Time Series Analysis. The *Fundamental Analysis* (Fischer and Jordan, 1995) is a method used for inferring about the relation of the value of a company with its current stock prices and making inferences about discrepancies that may indicate higher future prices. The *Technical Analysis* is used for searching patterns in the historical series of stock prices with the goal of estimating future prices (Fischer and Jordan, 1995). *Statistical Modeling* is used for hypothesizing about the probability distributions of the historical series of stock prices and making predictions based on the moments of these distributions (Petruccelli et al., 1999). Finally, the *Time Series Analysis* (Box et al., 1994) is used for estimating stochastic processes associated with the historical series and using these processes for forecasting.

The Fundamental Analysis demands examining and interpreting huge amounts of economic and accounting data and, despite its usefulness when historical series of prices are not available or are too short, its interactive nature restricts its application to a small number of stocks. The Technical Analysis, on the other hand, has been supported by automated methods that allow its application to large number of stocks (de Freitas and da Silva, 1999). However, both methods fail in providing consistent quantitative measures of prediction assurance, making it difficult to use their predictions for comparing investment alternatives.

Statistical Modeling can be used for providing uncertainty measures of realization of the expected investment return. These uncertainty measures, named *risk*, make Statistical Modeling suitable for comparing investment alternatives. This feature is central to the Modern Portfolio Theory (MPT), introduced by Markowitz through his celebrated model of portfolio[1] selection (Markowitz, 1952). In this model, the total risk of an investment in various stocks is minimized by the optimal selection of pairs of stocks with low joint risk, which provides a mechanism of loss compensation known as efficient diversification. The portfolio selection process, then, consists of finding, in a large collection of stocks, a group of pairs that minimizes the

[1]A portfolio is a collection of stocks, bonds, or other forms of investment. In this paper a portfolio is always a collection of stocks only.

portfolio's risk at a given portfolio return, or maximizes the portfolio's return at a given risk. The model assumes that the historical series of returns[2] of each stock follows a Normal distribution, uses the mean of the series as a prediction of the stock future returns, the variance as a measure of the risk of the stock, and the covariance between each pair of stocks as a measure of joint risk (the standard deviation, semi-variance and absolute deviation may also be used as a measure of risk in the model).

After the Markowitz model, many other models that use its basic assumptions (that the series of returns is Normal and that the moments of this series can be used as measure of future return and risk) appeared (Elton et al., 2003). In these models, known today as classic models, the portfolio's expected return is given by the linear combination of the expected returns of its stocks (the mean return), weighted by the participation of each stock in the portfolio. The portfolio risk, in turn, varies, but it is often related with the central moments of the joint distribution.

Despite the wide adoption of the classical methods of portfolio selection, it is important to mention that the distributions of the series of returns often exhibit kurtosis and skewness (Kon, 1984), challenging the assumption of normality of the series of returns. In addition, the realization of mean returns tends to be verified only in the long term. This has stimulated the development of predictive models based on *Time Series Analysis* and other non-linear methods, like the artificial neural networks, as a way of supporting the needs of investment on short horizons.

The predictability of the stock markets is still an open question in financial theory, and the theoretical framework that guides this discussion, known as *Efficient Market Hypothesis (EMH)*, is still suffering revisions (Fama, 1998). The EMH states that the stock's prices move according to a *random walk* (Fischer and Jordan, 1995), in which all new information available are immediately assimilated in the prices. Therefore, the EMH suggests that the series of prices (and returns) would not contain useful information for predictions. However, many tests found anomalies in the time series, in the EMH sense, that may be used for predictive purposes (Elton et al., 2003), and successful applications are steady appearing in literature (White, 1988; Sharda and Patil, 1992; de Freitas et al., 2001; Hellström, 2000; **?**; Hung et al., 2003; Cao and Tay, 2003).

This work compares the performance of autoregressive neural network (ARNN) predictors with that of the mean and *random walk* predictors, in the Brazilian stock market. A new ARNN predictor is proposed, in which the differences between the values of the series of returns and a determined past value are the regression variables, instead of simply the returns, as in the traditional prediction methods. The results obtained with this predictor are analyzed and compared with some traditional prediction methods. In addition, with the goal of obtaining risk measures suitable for developing new predictive-based portfolio selection models, we investigate the normality of the errors of prediction and, supported by our results, we suggest and discuss one such model. An earlier version of this work appears in (de Freitas et al., 2005).

This paper is organized as follows. After this Introduction, in Section 2 we formulate the problem of predicting stock returns. In Section 3 we describe our experimental methods. Following, in Section 4 we present our experimental results and, in Section 5 we formulate our predictive portfolio optimization model based on the hypothesis of normality of the errors of prediction. Section 6 closes this paper with our conclusions and proposals of future works.

[2]The difference between the price of the stock at time t and the price at time $t-1$, divided by the price at time $t-1$

2 PREDICTION OF STOCK RETURNS

The one-period stock return in time t, r_t, is defined as the difference between the price of the stock at time t and the price at time $t-1$, divided by the price at time $t-1$, as shown in Eq. 1.

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \tag{1}$$

where, p_t and p_{t-1} are the stock prices at times t and $t-1$, respectively.

The series of N past returns of a stock, r', is defined as:

$$r' = (r_{t-N+1}, \ldots, r_{t-1}, r_t) \tag{2}$$

The time-series-based prediction of a future return of the stock can be defined as the process of using r' for obtaining an estimate of r_{t+l}, where $l \geq 1$. The value of l directly affects the choice of prediction method. For $l = 1$, or one-period prediction, the standard choice are the autoregressive methods (Box et al., 1994), while for $l > 1$, or multiple-period prediction, techniques such as the method of Temporal Differences (Sutton, 1988) are used; one can also make l recursive one-period predictions in this case. This work investigates only the one-period prediction problem.

2.1 Random walk predictor

The *random walk* predictor, also known as *naive predictor* (Hellström, 1999), is defined as:

$$\hat{r}_{t+1} = r_t \tag{3}$$

As the Eq. 3 shows, in this predictor the predicted stock return $\hat{r}_{t+1}$ is equal to its return at time t, r_t. The rationale behind this predictor is the verified first-lag autocorrelation found in stock returns data (Hellström and Holmström, 1998), in which r_t strongly correlates to r_{t-1}.

2.2 Statistical mean predictor

The statistical mean predictor is defined as the mean of the n past observations of the series of stock returns, as given by Eq. 4.

$$\hat{r}_{t+1} = \bar{r}_n = \frac{1}{n} \sum_{i=t-n+1}^{t} r_i \tag{4}$$

If n is equal to the number of available observations, $\hat{r}_{t+1}$ is the statistical expected value of the series of returns.

2.3 Autoregressive neural network (ARNN) predictors

We use an autoregressive neural network based predictor (White, 1988), $\mathcal{S}_{\mathcal{R}}$, whose inputs are the present value and the $p-1$ past values[3] of the series of returns, and the output is the value for the next time period, as shown in Eq. 5. The number of input values, p $(p > 0)$, is the *regression order*, and some techniques for obtaining it are shown in (Box et al., 1994).

$$(r_{t-p+1}, \ldots, r_{t-1}, r_t) \rightarrow \mathcal{S}_{\mathcal{R}} \rightarrow \hat{r}_{t+1} \qquad (5)$$

Thus, after trained with its input-output pairs, the neural predictor $\mathcal{S}_{\mathcal{R}}$ implements a fully operational non-linear multiple regression model for the series of returns.

In this work, we propose a new ARNN method, inspired in the way one observes a time series plot to guess its future value. In this task, one tends to concentrate the visual attention in the last points of the plot, creating an imaginary frame that delimits a region of the plot, and offers an image with sufficient visual information for extrapolating the next value of the series. In this prediction scheme, one uses the value of some point inside the region as a reference value. We mimic that in our method by subtracting a reference value from the returns presented to the neural network. This reference value is equal to one of the elements of the series inside the observation frame $t-p-k+1, \ldots, t$ $(p, k > 0)$. This new method was named *autoregressive moving reference AR-MR(p,k)*, where p is the regression order and k is the reference index. Then, the autoregressive moving reference neural network (AR-MRNN) inputs and outputs become:

$$(r_{t-p+1} - z, \ldots, r_t - z) \rightarrow \mathcal{S}_{\mathcal{R}} \rightarrow \widehat{r_{t+1} - z} \qquad (6)$$
$$z = r_{t-p-k+1}$$

After training $\hat{r}_{t+1}$ is obtained by:

$$\hat{r}_{t+1} = \widehat{r_{t+1} - z} + z \qquad (7)$$
$$z = r_{t-p-k+1}$$

With the autoregressive moving reference neural network method the values encoded into the neural network weights are usually smaller than those of the series of returns, which minimizes the possibility of saturation at the neuron's output, enhancing the dynamic range of the network and its ability of representing the series of returns. This alleviates the needs of pre-processing, like normalization and detrending. This feature is especially important for modeling trended time series, like the series of stock prices and indexes. Early results comparing this new method with the traditional autoregressive method show a peak improvement of 250% in the root mean squared error (RMSE) of the training set convergence.

3 EXPERIMENTAL METHODS

A set of experiments was designed with the goal of (i) comparing the performances of the ARNN predictors against the mean and *random walk* predictors, (ii) verifying the normality of the distributions of the series of returns, and (iii) verifying the normality of the errors of prediction. We conducted 60 one-period predictions of return of 46 stocks for each predictor (60×46 predictions).

3.1 Neural network predictors

We used ARNN predictors of order $p = 4$, both for the traditional predictor - ARNN(4), and for the autoregressive moving reference predictor - AR-MRNN(4,1). All ARNN predictors were implemented with a fully-connected feedforward neural network with 2 hidden layers, sigmoidal activation function, and 4:30:15:1 topology (4 input neurons, 30 neurons in the first hidden layer, 15 neurons in the second hidden layer, and 1 output neuron). The training was conducted during 200.000 epochs using the backpropagation algorithm with learning rate of 0.009 and inertia of 0.95. These topological and training parameters were determined empirically. We trained a total of 5520 ($60 \times 46 \times 2$) networks using the 64 ATHLON XP 1800 nodes Enterprise cluster of the Departamento de Informática of UFES.

3.2 Data

We selected 46 stocks from the current (the first quarter of 2005) IBovespa index portfolio of the Brazilian Bovespa stock exchange. For each stock, we computed the weekly returns sampled at Wednesdays in the period between 16-feb-2000 and 09-mar-2005, obtaining 46 series of 265 returns as the data set of our analysis and experiments.

We have trained a separated neural network for each of the 60 predictions, between 21-jan-2004 and 09-mar-2005, of each stock. To reduce overfitting, we employed a simplified form of cross validation (Haykin, 1999), which consisted of dividing the training data set for each of the 60 predictions into two segments: a training segment and a validating segment. The training segment for the first of the 60 predictions, the 200 weekly returns from 16-feb-2000 to 10-dec-2003, was used for updating the neural networks weights during the 200.000 training epochs, while the validating segment, the 5 weekly returns from 17-dec-2003 to 14-jan-2004, was used for choosing the best weights observed during the training. In order to select the best weights, at each 1.000 epochs the neural network of each stock was used for predicting the returns within the validating segment. The weights responsible for the smaller root mean squared prediction error (RMSE) observed with the validating segment were used for the prediction of the return of 21-jan-2004. This procedure was repeated for each of the remaining 59 predictions, advancing one week in all dates mentioned. The sizes of the segments were obtained with the heuristic proposed in (Haykin, 1999, pg. 217)

We used the same data set in much the same way for performing the 60 predictions of mean and *random walk* predictors. To calculate the mean we have used the complete training set (the training and the validating segments). The prediction of the random walk predictor was the last element of the validating segment.

[3] The notation used in this paper differs slightly from that used in (Box et al., 1994; White, 1988), where the predictions are made at time $t-1$ for the "future" time t.

3.3 Error measures

The comparison among the methods described in Section 2 was conducted using the measures of the prediction error described in this subsection. They are standard measures in the literature (Hellström, 1999; Armstrong and Collopy, 1992). In the following equations, r_i and $\hat{r}_i$ are the observed and the predicted values of prediction i, respectively, and P is the number of predictions.

3.3.1 RMSE error

The root mean squared error (RMSE) is a standard measure for evaluating the differences between two variables, and is defined here as:

$$RMSE = \sqrt{\frac{1}{P}\sum_{i=1}^{P}(r_i - \hat{r}_i)^2} \tag{8}$$

The RMSE error may be interpreted as the standard deviation of the $r_i - \hat{r}_i$ errors with respect to a zero mean value, that is, it shows the distance of the errors from the ideal situation of zero mean error. The RMSE has low *outlier*[4] protection, good sensitivity for small changes in data, and does not display data asymmetry (Armstrong and Collopy, 1992).

3.3.2 MAPE error

The Mean Absolute Percentage Error (MAPE) is defined as:

$$MAPE = \frac{1}{P}\sum_{i=1}^{P}|\frac{r_i - \hat{r}_i}{r_i}| \tag{9}$$

The MAPE error is a unit-free measure, has good sensitivity for small changes in data, does not display data asymmetry and has very low outlier protection.

3.3.3 Mean error

The Mean Error (ME) is the average difference between the real and predicted values, defined as:

$$ME = \frac{1}{P}\sum_{i=1}^{P} r_i - \hat{r}_i \tag{10}$$

The ME error is the central point of the distribution of prediction errors.

3.3.4 Hit rates

The Hit Rates (H_R, H_{R+} and H_{R-}) measure the percentages of predictions in which the signals of $\hat{r}$ and r coincides. H_R measures the percentage of predictions which the signal coincides with the signal of the observed value, for the non-zero observed and predicted returns, and is defined as:

$$H_R = \frac{Count_1^P(r_{t+1}\hat{r}_{t+1} > 0)}{Count_1^P(r_{t+1}\hat{r}_{t+1} \neq 0)} \tag{11}$$

[4]Outliers are unusual values observed in the series, typically generated by transient effects and sampling errors.

where, the function $Count_1^P(\cdot)$ counts the occurrences in which its argument are true in P predictions.

H_{R+} and H_{R-} were designed with the goal of detecting asymmetries in the predictor's performance. They measure the percentage of the predicted positive and negative returns, respectively, which the signal coincides with the signal of the observed value:

$$H_{R+} = \frac{Count_1^P(r_{t+1} > 0 \ AND \ \hat{r}_{t+1} > 0)}{Count_1^P(\hat{r}_{t+1} > 0)} \tag{12}$$

$$H_{R-} = \frac{Count_1^P(r_{t+1} < 0 \ AND \ \hat{r}_{t+1} < 0)}{Count_1^P(\hat{r}_{t+1} < 0)} \tag{13}$$

4 EXPERIMENTS AND RESULTS

4.1 Normality of the series of returns

The normality of the series of returns was investigated with the Chi-square test (Papoulis, 1991), and the results for the 46 series are shown in Tab. 1. The tests were conducted for the standard significance levels (α) of 0.01, 0.05 and 0.10, and the normality was accepted (i.e. not rejected) for 14 (30%), 6 (13%) and 5 (11%) series, respectively. All the 46 (100%) distributions of returns have shown positive kurtosis, with an average value of 9.5, for a two standard errors of kurtosis ($sek = \sqrt{24/P}$) of 1.26, displaying, therefore, central values much higher than those expected in a Normal distribution. Also, 44 (96%) of the distributions have shown significant positive skewness, with average value of 1.2, for a two standard errors of skewness ($ses = \sqrt{6/P}$) of 0.632, displaying a significantly higher probability of occurring positive returns than negative ones. The minimum and maximum values of kurtosis and skewness are also shown in Tab. 1.

4.2 Prediction errors

Table 2 summarizes errors of prediction of all predictors studied. The table shows the minimum, maximum, average and the standard deviation of each error measure presented in Section 3.3. As Table 2 shows, all predictors produced near zero average ME errors, as expected for non-biased predictors. The ARNN predictors showed similar ME errors, while the mean and *random walk* predictors produced significantly lower values. The low values of this error for the *random walk* and mean predictors suggest that the losses

Table 1: Distributional analysis for the IBovespa stocks

Chi-square Test for the 46 series of returns		
α	H_0 not rejected	H_0 rejected
0.01	14 (30%)	32 (70%)
0.05	6 (13%)	40 (87%)
0.10	5 (11%)	41 (89%)
Kurtosis and skewness for the 46 series of returns		
	kurtosis	skewness
min	0.468249	-0.402678
max	93.479614	7.394180
avg	9.514798	1.246658
> 0	46 (100%)	44 (96%)
< 0	0 (0%)	2 (4%)

Table 2: Summary of the errors of prediction

Summary of the 60 prediction errors of the 46 series of returns				
ARNN(4) PREDICTOR				
	min	max	avg	std. dev.
ME	-0.014225	0.009038	-0.003735	0.005195
RMSE	0.037974	0.088681	0.054420	0.012041
MAPE	0.786097	307.092668	20.474236	64.231617
H_R	0.390000	0.660000	0.536522	0.069641
H_R+	0.350000	0.660000	0.531304	0.083336
H_R-	0.110000	1.000000	0.478043	0.149229
AR-MRNN(4,1) PREDICTOR				
	min	max	avg	std. dev.
ME	-0.020726	0.009740	-0.002170	0.006187
RMSE	0.042013	0.119054	0.066950	0.017264
MAPE	0.914974	1004.082783	57.390787	202.008827
H_R	0.410000	0.650000	0.514783	0.060175
H_{R+}	0.370000	0.660000	0.521522	0.081322
H_{R-}	0.250000	0.630000	0.447826	0.097660
RANDOM WALK PREDICTOR				
	min	max	avg	std. dev.
ME	-0.003031	0.002325	-0.000220	0.001235
RMSE	0.035464	0.083520	0.052723	0.011053
MAPE	1.175320	2422.404885	139.987180	467.397381
H_R	0.330000	0.660000	0.479565	0.071148
H_{R+}	0.300000	0.710000	0.505652	0.089509
H_{R-}	0.200000	0.600000	0.401087	0.105013
MEAN PREDICTOR				
	min	max	avg	std. dev.
ME	-0.011019	0.010480	0.000562	0.004249
RMSE	0.000013	0.009942	0.004319	0.002498
MAPE	0.000862	124.281414	8.802414	25.721641
H_R	0.440000	0.640000	0.560652	0.051872
H_{R+}	0.000000	0.620000	0.473261	0.192250
H_{R-}	0.000000	0.600000	0.071304	0.186876

and gains with these predictors cancel themselves out in the long run.

For the RMSE errors, all predictors produced similar average values, about 6%, except the mean predictor, which produced significantly lower values.

The levels of MAPE errors were very high, because of the low magnitude of the values that appear in the series of returns ($r \ll 1$), and also because of the very low outlier protection of this error measure.

The average of the total hit rates H_R of the ARNN predictors were about 52%, displaying a predictive capability slightly higher than chance (50% hit). The H_R reported in literature are typically 55% (Hellström, 1999), and in our experiments, 18 ARNN(4) predictors and 12 AR-MRNN(4,1) predictors outperformed this value, with 6 ARNN(4) predictors and 7 AR-MRNN(4,1) predictors exceeding the 60% hit rate. Also, the H_{R+} results were closer to H_R results than to H_{R-} results, which were 6% below H_R, showing a superior ability for predicting positive returns. The H_R for the *random walk* predictor was also dominated by H_{R+}, and near the expected 50% value. The mean predictor displayed its filtering properties and hit almost only the positive returns.

Figures 1 and 2 show a sample of the results obtained with the ARNN(4) and AR-MRNN(4,1) neural predictors. They show the 60 predictions of the PETR4 (Petrobras) stock returns. The time series ("Tgt" caption - dotted line) and the predictions ("Out" caption - solid line) are shown in the top plots. The weeks for which there are predictions are shown in the magnified plot, on the top right. The bottom plots show the Autocorrelation Function coefficients (ACF) and the Probability Density Function (PDF) of the 60 prediction errors, on the left and right sides, respectively. The autocorrelation coefficients are lower than their standard errors (dotted line), suggesting that the predictions are not biased (i.e. independent), and the results of the normality tests (captioned "H0" in the PDF plot, showing the α value and the test result, accepted - AC, or rejected - RE) shows that the errors are normal for a 0.01 Chi-square significance level. In these plots, the AR-MRNN(4,1) predictor performed better than the ARNN(4) predictor, however, this superiority was not verified in the summarized results shown in Tab. 2, where the ARNN(4) predictors has beaten the AR-MRNN(4,1) in all error measures. This fact suggests that, despite the AR-MRNN(4,1) predictor displays a higher dynamic range than the ARNN(4), its signal fitting is poorer, which demands further investigation.

Table 3: Normality of the errors for the ARNN predictors

Chi-square test for the prediction errors		
ARNN(4) Predictor		
α	H_0 not rejected	H_0 rejected
0.01	45 (98%)	1 (2%)
0.05	40 (87%)	6 (13%)
0.10	39 (85%)	7 (15%)
AR-MRNN-RM(4,1) Predictor		
α	H_0 not rejected	H_0 rejected
0.01	44 (96%)	2 (4%)
0.05	39 (85%)	7 (15%)
0.10	36 (78%)	10(22%)

4.3 Normality of prediction errors

The normality of the errors of prediction of the ARNN(4) and AR-MRNN(4,1) predictors was tested and the results are shown in Tab. 3. The Chi-square tests of the 46 series, each one with 60 observations, were conducted for the standard significance levels (α) of 0.01, 0.05 and 0.10. The normality of the prediction errors was accepted (i.e. not rejected) for 45 (98%), 40 (87%) and 39 (85%) of the ARNN(4) predictors, and for 44 (96%), 39 (85%) and 36 (78%) of the AR-MRNN(4,1) predictors, respectively.

The normality of the series of returns (Section 4.1) was accepted for 14 (30%), 6 (13%) and 5 (11%) of the series, respectively and for the same α values. These results suggest that it is possible to produce Normal prediction errors from non-Normal series of returns. This fact is specially useful in developing new predictive portfolio selection models (de Freitas et al., 2001) with risk measures based on these prediction errors, which takes the benefit of the theoretical developments of the Normal framework.

5 PREDICTIVE PORTFOLIO MODEL

Since the original Markowitz's mean-variance proposition, computational feasibility, model simplifications and development of risk measures has received considerable research attention. This effort strongly contributed to establish the portfolio theory and practice, producing remarkable results (Markowitz, 1956; Sharpe, 1963; Konno

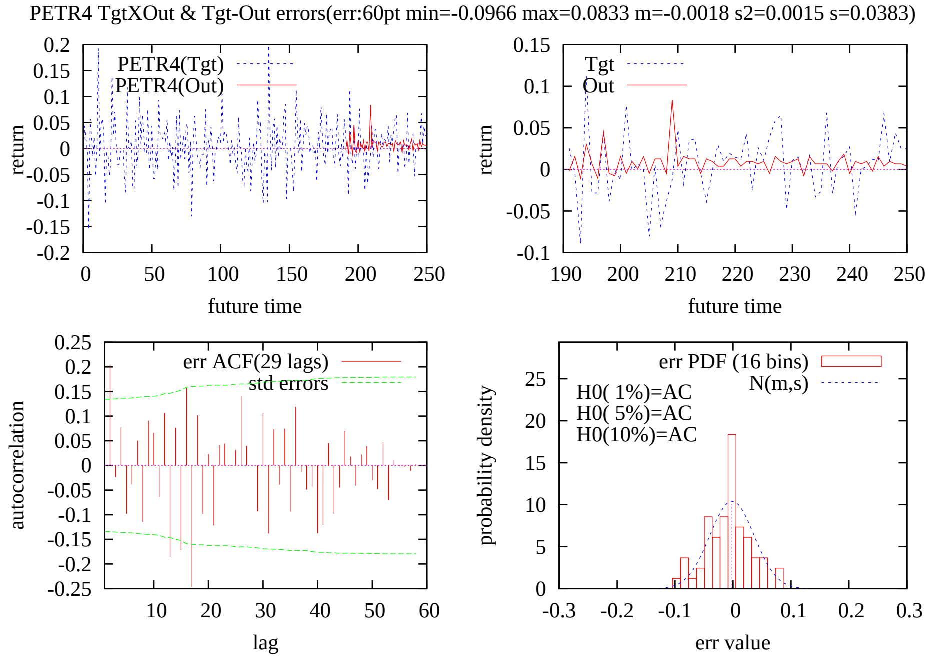

Figure 1: ARNN(4) predictions for the PETR4 stock

and Yamazaki, 1991; Sortino and van der Meer, 1991; Hamza and Janssen, 1996). However, the mean estimates for the future returns are expected to be verified only in the long horizons, therefore, not totally suitable for using in active portfolio management and other short-horizon strategies. The using of better prediction methods for obtaining the estimates of future short-horizon returns, with suitable risk measures, may derive new predictive portfolio models for these short-horizon applications. In an early work (de Freitas et al., 2001), we modified the mean-variance model to use predicted returns, measuring the prediction risk with an at-target variance, in which we calculated the mean quadratic deviation of the series of returns with respect to the predicted return. In preliminary investigations with the prediction-quadratic deviation model, we achieved portfolio returns 12% higher than the Markowitz model, for investment simulations with real data of Brazilian market.

This section formulates a probabilistic-predictive portfolio optimization model, based on risk measures developed with the results of Section 4.3. The model follows the safety-first class of portfolio models, first introduced by (Roy, 1952).

Let the real stock return r_t in time t be given by:

$$r_t = \hat{r}_t + \epsilon_t \tag{14}$$

where, $\hat{r}_t$ is the predicted return for time t, obtained at time $t-1$, ϵ_t is the prediction error for time t, and r_t is the real value for time t.

For a non-biased predictor, the series $\varepsilon' = (\epsilon_1, \epsilon_2, \ldots, \epsilon_n)$ of the n prediction errors, defined by:

$$\epsilon_t = r_t - \hat{r}_t \tag{15}$$

must be statistically independent and identically distributed (iid), with mean and variance given by:

$$\mu_\epsilon = \bar{\epsilon} = 0 \tag{16}$$

$$\sigma_\epsilon^2 = \frac{1}{n-1} \sum_{t=1}^{n} {\epsilon_t}^2 \tag{17}$$

5.1 Portfolio return and risk

The portfolio return is the linear combination of the predicted returns of its stocks, weighted by the participation of each stock in the portfolio, given by:

$$\hat{r}_p = \sum_{i=1}^{N} X_i \hat{r}_i \tag{18}$$

The portfolio risk is the variance of the joint Normal distribution of the combination, expressed by:

$$\hat{\sigma}_p^2 = \sum_{i=1}^{N} \sum_{j=1}^{N} X_i X_j \hat{\gamma}_{ij} \tag{19}$$

where, the covariances $\hat{\gamma}_{ij}$ obtained with the *ex-ante* prediction errors of the stocks i and j as:

$$\hat{\gamma}_{ij} = \gamma_{\epsilon ij} = \frac{1}{n} \sum_{t=1}^{n} \epsilon_{it} \epsilon_{jt} \tag{20}$$

In Eqs.18, 19 and 20, X_i is the share of the stock i in the portfolio, $\hat{r}_i$ is the prediction of the future return for stock i, $\hat{\gamma}_{ij}$ is the covariance between the i and j series of predicted returns, ϵ_{it} is the prediction error for the stock i in time t, N is the number of considered stocks, and n is the length of time series of errors.

5.1.1 Portfolio model

With the fundamental measures defined, the probabilistic-predictive portfolio optimization model is formulated as:

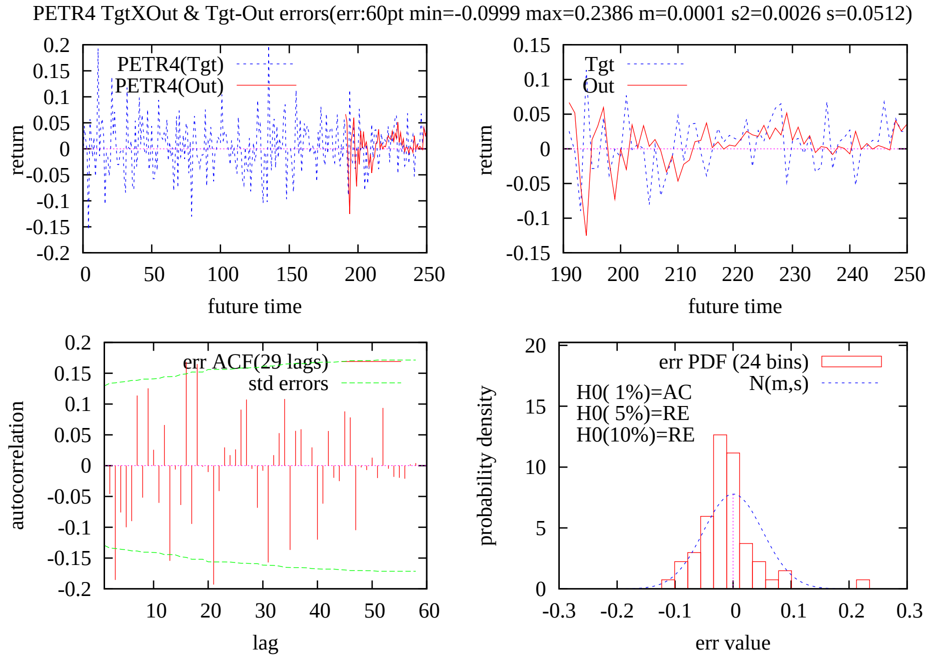

Figure 2: AR-MRNN(4,1) predictions for PETR4 stock

$$Maximize$$

$$P_{R_P}(\hat{r}_p \geq R_d) = \int_{R_d}^{\infty} \frac{1}{\hat{\sigma}_p^2\sqrt{2\pi}} e^{-(1/2)\,[(x-\hat{r}_p)/\hat{\sigma}_p^2]^2} dx, \tag{21}$$

$$subject\ to$$

$$\sum_{i=1}^{N} X_i = 1 \tag{22}$$

$$X_i \geq 0, i = 1, \ldots, N \tag{23}$$

where, R_d is the desired portfolio return, X_i is the stock i share in the portfolio, $\hat{r}_p$ and $\hat{\sigma}_p^2$ are the portfolio return and risk, respectively and defined in Eqs.18 and 19, and N is the number of considered stocks.

Despite the using of the joint Normal modeling, this model strongly differs from the Markowitz model because (i) the expected returns are the predicted returns instead of the mean of the series of returns, (ii) the individual distributions of the stock returns are obtained with the distributions of the *ex-ante* prediction errors centered in the predicted return, instead of with the historical series of returns, and (iii) the models optimize different utility functions. This result is based on the assumptions of Eq. 16 and of the normality of the errors of prediction, which are experimentally supported by the results shown in Tabs. 2 and 3.

The prime motivation in the proposal of this model is our belief that the qualification of portfolios in terms of the probability of obtaining the desired return, bounded in the $[0, 1]$ interval, has a more plausible interpretation to the investor than the unbounded dispersive risk measures.

The maximization of the probability of Eq. 21 is the *Roy's criteria* (Elton et al., 2003), defined as:

$$Maximize \qquad \frac{\hat{r}_p - R_d}{\hat{\sigma}_p} \tag{24}$$

This method explores the properties of the Normal curve, through the relation of the area below the curve (i.e. the probability), starting at the R_d abscissa, and the distance between the abscissas of R_d and $\hat{r}_p$, in standard deviation units. Thus, the optimization problem can be efficiently solved with the algorithms derived in (Elton et al., 2003), instead of directly use the stochastic programming methods.

6 CONCLUSION AND FUTURE WORKS

In this work we have investigated the performance of autoregressive neural network (ARNN) predictors and compared it against that of the mean and *random walk* predictors. We conducted 60 one-period prediction experiments for 46 series of weekly returns from the Brazilian stock market.

For the hit rate measures of the prediction signal, the performances of the ARNN predictors were substantively better than those of the reference methods. Our results contradict the Efficient Market Hypothesis, since they show that many of our predictors can achieve hit rates higher than 60%.

A new prediction method, named *autoregressive moving reference neural network*, was introduced. In this method, the regression variables are the differences of the values of the series and a determined past value. The prediction performance of the new method was compared with the traditional autoregressive neural network, and although it has exhibited a better representational capability, its numerical results do not confirm its superiority, which demands further investigation and development.

The distributional aspects of the prediction errors produced by the ARNN predictors were examined, and its normality was verified in 45 of the 46 series studied. This fact suggests the possibility of producing Normal prediction errors with the non-Normal series of returns. These prediction errors support the development of predictive portfolio optimization models using the Normal framework facilities. A predictive-probabilistic portfolio optimization model that explores this assumption of normality was proposed, with a brief discussion of some aspects for its solution.

ACKNOWLEDGEMENTS

We are very thankful to Enfoque Informações Financeiras Ltda for providing the stock market data used in this work.

REFERENCES

Armstrong, J. S. and Collopy, F. (1992). Error measures for generalizing about forecasting methods: Empirical comparisons, *International Journal of Forecasting* **8**(1): 69–80.

Box, G. E. P., Jenkins, G. M. and Reinsel, G. C. (1994). *Time Series Analysis: Forecasting and Control*, third edn, Prentice Hall.

Cao, L. J. and Tay, F. E. H. (2003). Support vector machine with adaptive parameters in financial time series forecasting, *IEEE Transactions on Neural Networks* **14**(6): 1506–1518.

de Freitas, A. A. C. and da Silva, I. N. (1999). Análise técnica de títulos financeiros através de redes neurais artificiais, *IV Congresso Brasileiro de Redes Neurais*, São José dos Campos - SP, Brazil.

de Freitas, F. D., de Souza, A. F. and de Almeida, A. R. (2005). Avaliação de preditores neurais autoregressivos no mercado de ações, *VII Congresso Brasileiro de Redes Neurais*, Natal, Brasil, p. to be appeared.

de Freitas, F. D., de Souza, A. F., Negreiros-Gomes, F. J. and de Almeida, A. R. (2001). Portfolio selection with predicted returns using neural networks, *IASTED International Conference on Artificial Intelligence and Applications*, Marbella, Spain, pp. 99–103.

Elton, E. J., Gruber, M. J., Brown, S. J. and Goetzmann, W. N. (2003). *Modern Portfolio Theory and Investment Analysis*, sixth edn, John Wiley & Sons, Inc.

Fama, E. (1998). Market efficiency, long-term returns, and behavioral finance, *Journal of Financial Economics* **49**: 283–306.

Fischer, D. E. and Jordan, R. J. (1995). *Security Analysis and Portfolio Management*, 6 edn, Prentice Hall International.

Hamza, F. and Janssen, J. (1996). Linear approach for solving large-scale portfolio optimization problems in a lognormal market, *IAA/AFIR Colloquium*, Nürnberg, Germany, pp. 1019–1039.

Haykin, S. (1999). *Neural Networks: a Comprehensive Foundation*, 2 edn, Prentice Hall, Inc.

Hellström, T. (1999). Data snooping in the stock market, *Theory of Stochastic Processes* **5**(21): 33–50.

Hellström, T. (2000). Predicting a rank measure for portfolio selection, *Theory of Stochastic Processes* **22**(6): 64–83.

Hellström, T. and Holmström, K. (1998). Predictable patterns in stock returns, *Research and Reports Opuscula ISRN HEV-BIB-OP-30-SE 30*, Department of Mathematics and Physics, Mälardalen University, Sweden.

Hung, K.-k., Yiu-ming, C. and Xu, L. (2003). An extended ASLD trading system to enhance portfolio management, *IEEE Transactions on Neural Networks* **14**(2): 413–425.

Kon, S. J. (1984). Models of stock returns - a comparison, *Journal of Finance* **XXXIX**(1): 147–165.

Konno, H. and Yamazaki, H. (1991). Mean-absolute deviation portfolio optimization model and its applications to Tokyo stock market, *Management Science* **37**(5): 519–531.

Markowitz, H. M. (1952). Portfolio selection, *Journal of Finance* **VII**(1): 77–91.

Markowitz, H. M. (1956). The optimization of a quadratic function subject to linear constraints, *Naval Research Logistics Quarterly* (3): 111–1331.

Papoulis, A. (1991). *Probability, Random Variables, and Stochastic Processes*, third edn, WCB-McGraw-Hill.

Petruccelli, J. D., Nandram, B. and Chen, M. (1999). *Applied Statistics for Engineers and Scientists*, Prentice-Hall, Inc.

Roy, A. D. (1952). Safety-first and the holding of assets, *Econometrica* **5**(20): 431–449.

Sharda, B. and Patil, R. B. (1992). A connectionist approach to time series prediction: An empirical test, *Journal of Intelligent Manufacturing* .

Sharpe, W. F. (1963). A simplified model for portfolio analysis, *Management Science* (9): 277–293.

Sortino, F. A. and van der Meer, R. (1991). Downside risk, *Journal of Portfolio Management* pp. 27–31.

Sutton, R. S. (1988). Learning to predict by the methods of temporal differences, *Machine Learning* **3**: 9–44. *citeseer.ist.psu.edu/sutton88learning.html

White, H. (1988). Economic prediction using neural networks: The case of ibm daily stock returns, *Proceedings of The IEEE International Conference on Neural Networks*.