



An Endogenous Feedback and Entropy Analysis in Machine Learning Model for Stock's Return Forecast

Edson Vinicius Pontes Bastos^{1,*}, Jorge Junio Moreira Antunes², Lino Guimarães Marujo¹, Peter Fernandes Wanke² and Roberto Ivo da Rocha Lima Filho¹

¹Instituto Alberto Luiz Coimbra de Pós-Graduação e Pesquisa de Engenharia-COPPE, Universidade Federal do Rio de Janeiro (UFRJ), Rio de Janeiro, 21941-598, Brasil

²Instituto de Pós-Graduação em Administração-Coppead, Universidade Federal do Rio de Janeiro (UFRJ), 21941-918, Brasil

*Corresponding Author: Edson Vinicius Pontes Bastos. Email: edson.bastos@coppe.ufrj.br

Received: 17 July 2022; Accepted: 04 November 2022

Abstract: Stock markets exhibit Brownian movement with random, non-linear, uncertain, evolutionary, non-parametric, nebulous, chaotic characteristics and dynamism with a high degree of complexity. Developing an algorithm to predict returns for decision-making is a challenging goal. In addition, the choice of variables that will serve as input to the model represents a non-triviality, since it is possible to observe endogeneity problems between the predictor and the predicted variables. Thus, the goal is to analyze the endogenous origin of the stock return prediction model based on technical indicators. For this, we structure a feed-forward neural network. We evaluate the endogenous feedback between the predicted returns and technical analysis indicators based on the generated residues. It is possible to predict the return. The high accuracy of the model indicates that, during the test period, there is a hit rate close to 76%. Regarding endogeneity, the term of interest and the return are the variables that influence the largest number of indicators. The results will help investors build investment strategies based on this expert system applied to forecasting.

Keywords: Forecast; endogeneity; neural networks; differential evolution; stochastic optimization

1 Introduction

To predict and trade the BOVA11, an exchange-traded fund (ETF) from Brazil, we apply a new approach using technical indicators. The main trading system is based on a neural network forecast method. The results are highly favorable to the hypothesis of abnormal returns in a successful forecast of one-day-ahead returns ($D + 1$) since the forecasting strategy proved to be feasible with an accuracy close to 76%.

According to [1], the efficient market hypothesis, suggests that it is impossible to predict stock value, since these values have random behavior. However, new research [2–5] shows that most share prices reflect previous records and information, so movement trends are vital for effectively predicting values [3]. Classic economic-financial literature defends the informational efficiency of the stock exchange. In a



This work is licensed under a Creative Commons Attribution 4.0 International License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

so-called perfectly efficient market, abnormal returns are unattainable, as prices have already incorporated all relevant information for future results. Thus, the efficient market is defined by [6] as one in which the assets prices fully reflect all available information at a given time. Assuming that all information is reflected in the price, investors expect to obtain a normal rate of return, which rules out the possibility of earning an abnormal return. In other words, in a so-called efficient market, it is expected that all the actors involved have the same access to available information and that these are already priced. Therefore, there would be no possibility of arbitrage. As a result, it would be impossible for investors to buy an undervalued asset or sell an overvalued asset to make an abnormal profit, as each stock would always be traded at its fair value. Thus, the theory questions the predictability of prices and opportunities for profitable and consistent long-term trades. The idea that the market cannot be conquered motivates a long-standing controversy between academics and market professionals.

Financial managers always seek to maximize the return on their investments, as the objective is to “win”. However, [7] confirm that hitting the stock market direction is considered an uncertain and high-risk strategy, since many external factors can affect these returns. Furthermore, according to the Efficient Market Hypothesis (EMH) theory, it is impossible to predict a series of returns. Such a time series follows a Brownian and independent movement with random characteristics. It is in this scenario of chaos and controversy that this article develops. The entropy concept can be applied in finance for stock pricing and forecasting. This concept can help to solve the general problem of determining probability distributions of financial market assets, which is characterized by volatility, uncertainty, limited and incomplete information.

Regarding entropy, we will apply the Stochastic Structural Relationship Programming (SSRP) model, based on the methodology of neural network residuals. With this model, we will evaluate the endogenous relationship between the predictive variables of the neural network using the maximum entropy functions according to the Principle of Maximum Entropy (MEP). From this method, it is possible to produce a general mapping of the causal relations between predictors and the Return. From this context, the following research questions arise:

1. How accurately is it possible to hit the one-day-ahead return forecast using technical indicators as input to the neural network predictive model?
2. What is the direction of the endogenous relationship between the technical indicators used as input to the neural network model for predicting returns?

The goal is to investigate the entropy and endogenous origin in the prediction of stock returns, with the predictive model inputs given by the technical signals. To that end, we will carry out an endogenous analysis between the technical indicators and the dependent variable Return. We will build a machine learning algorithm using artificial neural networks (ANN) to predict market returns. This research seeks to contribute to the research gap evidenced by the literature [8], which highlights the importance of investigating the link between performance and entropy on predictive stock models.

Studies like this are justified by the empirical evidence of utility, meaning the ability to satisfy a need. Which is contained in an artificial intelligence algorithm applied to emerging markets. Such discussion has the potential to empirically contribute to the literature, concerning the behavior analysis of asset returns forecast. This applies especially in countries with volatile financial markets, as advocated by [9], which are subjected to political-economic uncertainties and frequent changes in the rating, as seen in Brazil. Furthermore, with our work, it will be possible to outline a strategy capable of predicting the future return of the asset. And then, investigate the performance of an investment strategy made with a machine learning algorithm.

Reference [10] highlights the importance of developing and improving predictive machine learning algorithms applied to the financial market. With these algorithms, it is possible to analyze multiple assets

in different regions and countries. In this way, the operation becomes more efficient and diversified. This research contributes by bringing the possibility of abnormal returns, making clear the need to create better regulatory mechanisms to reduce informational inefficiencies.

2 Materials and Methods

We will work with the assumption that there is a pattern in the database. That pattern cannot be described mathematically by deterministic methods. So, we used machine learning to identify this hidden pattern, according to the learning theory [11]. The objective is to obtain an approximation function to better describe the behavior of the return that is unknown.

Analogous to the market features, we can mention the human brain because they share some characteristics such as nonlinearity, uncertainty, evolution, cloudiness, chaotic and dynamic systems with a high degree of complexity. According to [12], the brain can be considered a highly complex, non-linear and parallel computing, capable of organizing its constituents (neurons) for processing. In this sense, neural networks present themselves as a processor capable of stock and process the acquired knowledge and making it available for use. Regarding the learning components, the dependent variable (Y) is Return (Ret). This is the response variable to check how the algorithm behaves to predict day t with the lagged at t-1, that is, a lagged day. Technical and volatile VIX indicators will be the X_i inputs.

The ANN mathematical model is formed by a set of n inputs (x_1, x_2, \dots, x_n), to which weights (w_1, w_2, \dots, w_n) are attached indicating each corresponding neuron effect, by a linear combination $\sum_i^n w_i x_i$. The sum of the product between the weights w_1 and x_1 of input signals passes through the activation function f and induces the transformation. Every neuron transforms the input vector into an output O (x). We have included the hidden layer to increase flexibility. The general mathematical representation of a single hidden layer network with J hidden neurons was recommended by [13].

The database is BOVA11, a Brazilian ETF with more than 60 companies traded on B3 (Brasil, Bolsa, Balcão). The BOVA11 includes companies such as Petrobras, Vale, Itaú Unibanco, Bradesco Bank, Ambev and Banco do Brasil. We downloaded the dataset from Yahoo's finance website as done by [14], where there is historical stock time-series data such as open, high, low, close, adjusted close and volume information.

The period under review is 2010 to 2020. We divided the dataset into training and testing. The training stage comprises 80% of the oldest observations (1520 observations) and the test stage, the most recent 20% of the database (380 observations). We remove samples with null or missing information as a cleaning operation process.

We use the ANN tool to predict the BOVA11 returns (Ret) one-day-ahead. This network is structurally similar to biological neural structures and has computational capacity acquired through learning and generalization. We structured the learning algorithm with feed-forward neural networks with R software [15]. Therefore, step zero was to structure a single-hidden-layer Multilayer Perceptron (MLP), the feed-forward neural network. We then apply the quasi-Newton method (Broyden-Fletcher-Goldfarb-Shanno algorithm, BFGS) [15] for optimization in the neural net step.

The first model selection problem is represented as a two-class classification to predict the Return Class more accurately. If the Return value is positive, the Return class takes the value 1, indicating a buy signal. Otherwise, the value will be zero. Thus, the objective is to evaluate the endogeneity of the neural network model to predict the return of the BOVA11 index one-day-ahead. This model uses traditional technical analysis variables, such as inputs presented in Table 1.

Table 1: Feed-forward neural networks inputs

Indicator	Abbrev.	Definition	Referential
VIX volatility index	VIX	It is a volatility indicator, designed to serve as a metric for market expectations of future volatility implied in option prices (derivatives).	[16]
Commodity channel index	CCI	Represents the current price positioning in relation to the period's moving average. It is a metric for price deviations from the moving average.	[17,18]
Moving average convergence divergence	MACD	It is an oscillator-type indicator, which varies centered on zero. It turns two trend indicators, a fast-moving average and a slow-moving average into a momentum oscillator by subtracting the higher moving average from the lower moving average. The 9-period exponential moving average serves to identify changes in trend.	[18,19]
Williams %R	WILL	Larry Williams is a momentum indicator and it means the closing level relative to the peak of the period, the asset maximum.	[18]
Stoch stochastic oscillator	STOCHK	It is a momentum indicator and it represents the price position related to the amplitude of variation relative to the maximum and the minimum of a period n.	[19]
Triple smoothed exponential oscillator	TSEO	The Triple Smoothed Exponential Oscillator is a boost indicator that indicates the percentage change of the exponential average. It calculates the rate of change within a triple	[19]
Bollinger bands	BB	It is a volatility indicator built on a 20-period moving average, where two bands with 2 standard deviations above the average and 2 standard deviations below the moving average are added.	[18]
Chande momentum oscillator	CMO	Indicates the momentum change with the total movement divided by the net movement.	[18]
Detrended price oscillator	DPO	It is the price time series without the trend component and does it by subtracting the moving average from the price over the price.	[18,19]
Rate of change	ROC	It is an oscillator-type indicator based on the rate of change calculated from the time series, the percentage change in prices over n periods. It is a zero-centered indicator floating at this point.	[18,19]

Indicators are known to give an idea of a specific metric, so they should always be read in context, using other tools to avoid false signals. The increase in subjectivity is evident when using only technical indicators for decision-making, without applying a structured and subjectivity-free methodology that can analyze such indicators. The need to remove human judgment is increasingly emphasized to make good use of indicators. It is necessary to structure a machine learning model to improve the extraction of information from an extensive set of indicators. According to [14], the ability to promote decision-making by induction and

the behavior of the human brain were the main motivators for the development of neural networks, which can be seen as intelligent computerized systems. The neural network main model follows the following specification:

Model 1: $Ret \sim f(VIX + CCI + MACD + WILL + STOCHK + TSEO + BB + CMO + DPO + ROC)$

Additionally, we structured 10 feed-forward neural networks, according to the specified models (2 to 11). The objective was to detail the endogenous developments between the variable Return and its predictors (the technical signs) to identify significant causal relationships. Therefore, we verify the hypothesis of endogeneity between the predicted return variable and the predictors for the technical analysis. As defined by [20], endogeneity exists when the explanatory variables have some correlation with the error term of the predictive model. Once the non-correlation assumption is violated, one or more regressors will be endogenous. This endogeneity adds bias to the estimators obtained, negatively affecting the outcome of the forecasting model. For [21], in finance it is possible to observe endogeneity due to measurement errors of regressors, simultaneity or even by variables omitted from the model.

2.1 First Endogeneity Analysis: Minimal Endogenous Relationship Variance

Technical indicators are correlated and a correlation between their residues is expected. The existence of this correlation suggests evidence of endogenous relationships between the predictors. Thus, this research presents an approach based on the Stochastic Structural Relationship Programming (SSRP) model, from the residues generated by 10 specifications of neural network models. The objective is to clarify the endogeneity and significant structural cause-and-effect relationships that exist between Return and technical analysis variables. For this, we use the residues obtained by the following 10 models:

Model 2: $VIX \sim f(Ret + CCI + MACD + WILL + STOCHK + TSEO + BB + CMO + DPO + ROC)$

Model 3: $CCI \sim f(VIX + Ret + MACD + WILL + STOCHK + TSEO + BB + CMO + DPO + ROC)$

Model 4: $MACD \sim f(VIX + CCI + Ret + WILL + STOCHK + TSEO + BB + CMO + DPO + ROC)$

Model 5: $WILL \sim f(VIX + CCI + MACD + Ret + STOCHK + TSEO + BB + CMO + DPO + ROC)$

Model 6: $STOCHK \sim f(VIX + CCI + MACD + WILL + Ret + TSEO + BB + CMO + DPO + ROC)$

Model 7: $TSEO \sim f(VIX + CCI + MACD + WILL + STOCHK + Ret + BB + CMO + DPO + ROC)$

Model 8: $BB \sim f(VIX + CCI + MACD + WILL + STOCHK + TSEO + Ret + CMO + DPO + ROC)$

Model 9: $CMO \sim f(VIX + CCI + MACD + WILL + STOCHK + TSEO + BB + Ret + DPO + ROC)$

Model 10: $DPO \sim f(VIX + CCI + MACD + WILL + STOCHK + TSEO + BB + CMO + Ret + ROC)$

Model 11: $ROC \sim f(VIX + CCI + MACD + WILL + STOCHK + TSEO + BB + CMO + DPO + Ret)$

We use the residues from models 1 to 11 to generate sets of conditional probability distributions of the residuals, thus obtaining ten probability distributions to investigate the entropy of such residuals. In information theory, entropy refers to the probabilistic uncertainty related to a given probability distribution. Different degrees of uncertainty was associated with different distributions, since each distribution has an intrinsic degree of uncertainty. The principle of Maximal Information Entropy for Directional Weighted Residuals establishes that the probability distribution most adherent to the variable is the one with the highest entropy.

The conditional distributions of the residues show the direction of the relationship between the variables under study. There are two steps to run the Stochastic Structural Relationship Programming (SSRP) method that reveal the endogeneity and identify significant cause and effect relationships. The first step, called Minimal Endogenous Relationship Variance, consists of exploring the degree of relative importance

between the variables of predictive models from 1 to 11, through the variance of each model. To investigate the existence of endogeneity we used the covariance between the models.

We simultaneously minimize the covariance and variance terms of the residues of the 11 models by a nonlinear stochastic optimization problem, as presented in Eq. (1), according to [22].

$$\min \left[\text{Var} \left(\sum_{i=1}^{11} w_i * R_i \right) + \left(2 * \sum_{i,j=1}^{11} \text{Covar} (w_i * w_j * R_i * R_j), i \neq j, j < i \right) \right]$$

subject to

$$\sum_{i=1}^{11} w_i = 1 \quad (1)$$

where w_i stands for the weights, which range from 0 to 1 ($0 \leq w_i \leq 1 \forall i$), assigned, respectively, to the residual vectors of each one of the 11 models described; R_i is the models' residues where i varies from 1 to 11.

We solve Eq. (1) with the differential evolution method (DE). It is a stochastic method with the assumptions of natural selection and evolution. This algorithm uses a randomly generated initial population from a uniform distribution with probability crossover, differential mutation and selection operators. The DE method belongs to the genetic algorithms (GA) field and is biologically inspired, just like the structures of neural networks. According to [10], the optimization process in the GA is based on a randomly guided process. In this process, a group of specific parameters is randomly generated for a fixed number of so-called populations. More details on DE methodology can be seen in [23], which discusses the DEoptim R software package.

We solve models 1–11 with a bootstrap technique, with 100 repetitions each, generating 100 residues for each model. Subsequently, we optimize these 100 residues of models 1–11 by optimization Eq. (1), generating a W distribution for an assertive prediction of the coefficients. In this way, the W values obtained by the optimization represent the optimal solution and, therefore, the minimum points of variance (Var) and Covariance (Covar) of the Residues grouped according to Eq. (1) considered.

2.2 Second Endogeneity Analysis: Maximal Information Entropy for Directional Weighted Residuals

From the Principle of Maximum Entropy, the probability distribution that best represents the current stage of knowledge is the one with the highest entropy. The second step is the use of the Maximal Information Entropy for Directional Weighted Residuals algorithm. In this step, we get a set with all possible combinations of the Conditional Residual distributions (CR_k). We use the bootstrap results with 100 replications, obtained in the first step of the non-conditional distributions of residuals (R_i), as a starting point for the calculations of the second step, where $CR_k \sim f(R_i/R_j)$ for all i and j , $i \neq j$ and $K = i * j - i = 11 * 11 - 11 = 110$. Likewise, we apply the DE method [23] to solve the entire following non-linear programming model:

$$\max \left[\left(\sum_i^{11} \sum_j^{11} H \left(f \left(\frac{R_i}{R_j} \right) * w_i * w_j \right) \right) \text{OR} \left(\sum_i^{11} \sum_j^{11} H (g(R_i, R_j) * w_i * w_j) \right), i \neq j \right]$$

subject to

$$\sum_{i=1}^{11} w_i = 1 \quad (2)$$

$$0 \leq w_i \leq 1,$$

where $H(\cdot)$ represents the information entropy function, $g(R_i, R_j)$ is the unconditional marginals of the residuals from models (1)–(11), $\forall i, j, i \neq j, f(R_i/R_j)$ is the conditional distribution of the residuals from models (1)–(11), $\forall i, j, i \neq j$.

A nonlinear integer programming model makes it possible to identify whether the conditional distributions of each pair of residuals have significantly different directions. For example, the weights assigned to $f(R_i/R_j)$ could produce higher entropy than those assigned to levels of $f(R_j/R_i)$, compared to the unconditional residues analyzed in the first step, called Minimal Endogenous Relationship Variance for endogeneity investigation. This non-integer linear programming methodology returns the structural relationship of the dependent variables defined in Eq. (1), for which the information entropy is maximum. This method ensures uniqueness and provides consistent support for the probabilistic weight profile calculated in the Minimal Endogenous Relationship Variance step, where the overall residual variance is minimal. We used the weights calculated in the Minimal Endogenous Relationship Variance step as an initial database for the optimization of the Maximal Information Entropy for Directional Weighted Residuals algorithm, called step 2. We also used the DE methodology to find an optimal solution regarding the maximum entropy for each pair of variables ij .

Thus, the output is whether i cause j (or the other way around) or whether the relationship is endogenous, for each pair ij . Through Table 2 it is possible to see a summary of the methodological step-by-step and the pseudo-code used to calculate the estimates of the function $f(\cdot)$ and $g(\cdot)$ in the optimization of the step of Maximal Information Entropy for Directional Weighted Residuals.

Table 2: Pseudo code used for computing $f(\cdot)$ and $g(\cdot)$

Inputs:	n_model: Number of models fitted
	Residuals vectors from each model fitted
Outputs:	Conditional distribution of all n_models pairs combination
	Unconditional margin distribution of all n_models pairs combinations
1	Fit the statistical distributions for each model residuals
2	Calculate the correlation matrix between all model residuals
3	Fit the best distribution of residuals for each model (cf. Table 3)
4	Generate multivariate Copulas preserving the correlation and distribution structure for all model residuals
5	Select percentile thresholds to stress directional relationships under extreme Copulas distribution
6	for p in percentiles do
7	for i in 1 to n_models do
8	for j in i to n_models do
9	Evaluate the conditional distribution of copula ij
10	Select conditional distribution greater than p
11	Evaluate the unconditional marginal distributions of copula ij
12	Select unconditional distribution greater than p
13	end do
14	end do
15	end do

3 Results

The VIX indicator had the second smallest dispersion metric observed by the value of its standard deviation, second only to the indicator ROC. The highlight of the VIX value considered at the neural network input is the difference between the VIX of day t and the VIX of $t-1$, all divided by the value of the VIX index in $T-1$. The index VIX is quoted in percentage points. The higher the index, the greater the risk perception. It is known as the fear index, as it manages to capture investor sentiment.

After establishing the input connections, in the learning process, the objective was to find the fit of the weights vector π . Thus, the training objective aimed at convergence was achieved. For the first model, the neural network algorithm converged at the end of 740 iterations. At this point, we concluded that learning has occurred. This was associated with the ability of the neural network to adapt the parameters as a result of its interaction with the database. The learning process is interactive, and through it, the ANN should gradually improve its performance as it interacts with the variables.

The performance criteria that determine the ANN quality and the training breakpoint were pre-established by the training parameters, usually associated with measures of accuracy or error. In this way, we adjusted the hidden-layer neural network parameters with 2000 iterations at maximum with an iterative loop for each number of units in the hidden layer and weight decay, using the size (number of neurons units in the hidden layer) between 2 and 30 for layout test, and the decay test values between 1, 0.1, 0.01, 0.001, 0.0001, 0.00001 and 0 to converge to the best layout.

To this end, we structured a matrix to accumulate the accuracy values of the 10-fold cross-validation for each layout and decay. We validated the hyperparameter search of the neural network models (neurons, decay and error) by comparing the lowest values of the Mean Squared Error (MSE). We capture the predicted values through the prediction function with R software. If the variables are correlated, is expected a relationship between their residues. This correlation indicates endogenies between the predictors and, as a result, was the analysis of residuals of the 11 models obtained by the neural net combinations, where each predictor variable assumes the role of a dependent variable in separate models.

With this, it was possible to estimate the residuals of the 11 models. Using the minimum endogenous relationship variance methodology that minimizes the variance, we used Olden's criterion to capture the importance ranking between the models. Regarding Olden's criterion of importance over the main model, the most important variable was the ROC, while the least important variable with a negative value was the DPO indicator. In addition, the TSEO variable had repeatedly high negative importance in models 3, 5, 8 and 9, where the dependent variable was CCI, WILL, BB and CMO, respectively. When the dependent variable was the CMO indicator of the neural network, the high and positive degree of importance of the MACD variable stands out. On the other hand, the variables DPO, ROC, Ret, TSEO and VIX presented a high negative index of importance. Finally, for models with ROC as the dependent variable, there were lower degrees of Olden's importance for all predictors.

To answer the research objective, we present the performance analysis of the first model of neural networks through the confusion matrix. The dependent variable was the return and the inputs of the technical indicators.

The McNemar's Test is a metric to assess the performance of the predictive model through the analysis of the confusion matrix [24]. According to Table 3, the McNemar's Test had a p -value equal to 0.14, so it was not possible to reject the null hypothesis (H_0) given a significance level of 5%. Only the classifiers had a similar proportion of errors in the test dataset. The network correctly classified 1439 observations from 1899, an accuracy of approximately 76%. This confirmed the research hypothesis regarding the possibility of trade success. Thus, the application of neural network to predict return signals was profitable and consistent, showing that the investor who makes decisions based on neural network outputs would hit 76% of the days on average. Along with the risk management of the allocated capital, this hit

rate enabled profitable long-term trades, presenting evidence regarding the use of technical indicators as inputs of the neural network model.

Table 3: Confusion matrix analysis for the first model to predict return by ANN

Prediction	FALSE	TRUE
FALSE	675	246
TRUE	214	764
Accuracy 95% CI	0.7578 (0.7378, 0.7769)	
Positive pred value:	0.7329	
Negative pred value:	0.7812	
Balanced accuracy:	0.7579	
Sensitivity:	0.7593	
Specificity:	0.7564	
Mcnemar's test <i>P</i> -value:	0.1484	

The results indicated that the strategy allows higher gains than the buy and hold method. As a result, the second hypothesis was also valid: technical indicators can predict market movements. This result is in line with the works presented by [25], which evidenced that it is possible to have a winning predictive strategy in the stock market based on machine learning models, as applied in this article.

Regarding the verification of predictor variables endogeneity, the initial results refer to the analysis of the probability distribution of the model's residuals. Most models presented the logistical distribution as more adherent to their respective residues. The endogenous analysis was discussed based on the results of the Stochastic Structural Relationship Programming SSRP methodology.

The relative importance of the models to obtain the minimum residual variance was low in 10 variables including the return, as seen in Fig. 1. Such attributes accounted for almost all total median weights.

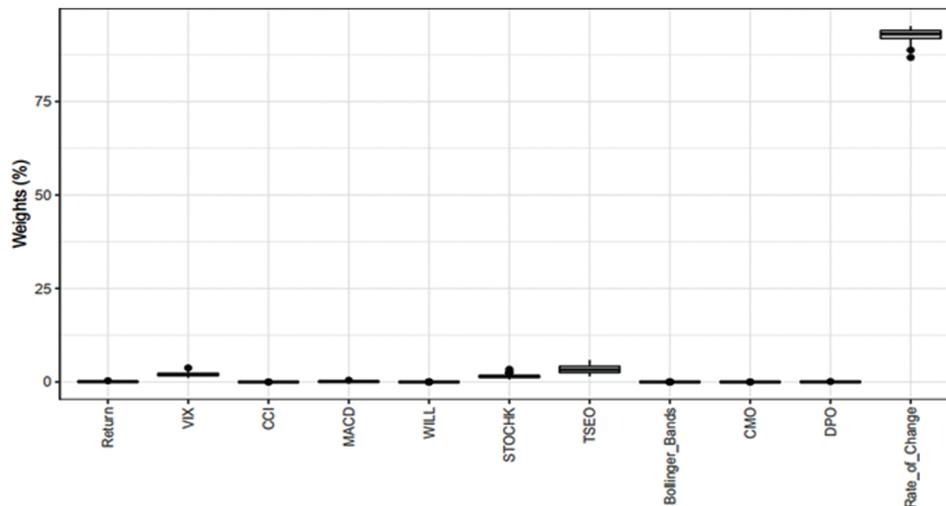


Figure 1: Relative importance of models 1–119

In a normal situation, where there was a balance between the variables, the expected importance value for each model would be an equivalent weight for all 11 models, that is, $(100\%)/11 = 9.09\%$. However, there were indicators with unbalanced weight distribution. Initially, some indicators showed the strongest causal relationships linked to the ROC variable. The Rate of Change has the property of measuring the price percentage change in a given period. This means that the greater the difference between the contemporary price and the price of the period considered, the greater the value of the ROC indicator. Such causality can be explained by the herd effect, usually present when the stock market undergoes large fluctuations around the mean. Analyzing the causality pairs, Fig. 2 confirmed that the ROC variable was associated with a cause-and-effect process that could occur with TSEO, VIX, and STOCHK. In Fig. 2 it is possible to see the results for the relative importance of the main interaction pairs in explaining the general variance of the residuals.

Regarding the main combinations of endogenous pairs, the effect of joint feedback on the residual's variances of the TSEO and ROC, VIX, ROC, STOCHK, and ROC pairs stands out. As it is a momentum indicator, the ROC variable indicates the percentage of variation during a time window, and it was possible to observe that it was controlled by the ROC, VIX, and STOCHK indicators.

As a robustness test, we presented the results of the entropy of information for conditional and unconditional distribution (Tables 4 and 5).

4 Discussion

As seen in Tables 4 and 5, the unconditional distribution is symmetric and the conditional distribution is asymmetric. The cause-and-effect relationship is perceptible through these tables obtained by the Maximal Information Entropy for Directional Weighted Residuals method for the 0.975 percentile. Also evidenced in Fig. 3 is the framework of the relationship between technical indicators. Such sign and weight are derived from Olden's criterion for ranking the importance of the neural network inputs, which provides the absolute importance.

In this second analysis, regarding conditional and non-conditional distributions, the objective was to find a maximization in the information entropy, as a test of robustness to verify if in a scenario with worse uncertainty (principle of maximum entropy) this type of behavior is repeated.

For each model, through bootstrap, we collect 100 combinations of residues and then find the non-conditional and conditional distribution of residues from one model to another. This is the second analysis detailed in the methodology by Eq. (2), to maximize the entropy of the information. It is considered a robustness test to confirm the first stage results. For the non-conditional simulation, we used the copulas method, following the same distribution and correlation between residuals found originally.

If the entropy of the information of the original (non-conditional) residuals is greater than conditional, the variable under analysis is independent and no other variable rules it. However, if the entropy of the conditional information is higher, then it is stated that there is a relationship between the two variables (there is endogeneity). In this case, the row variable influences the column variable in Tables 4 and 5. For example, in the first row of the Return variable in Table 4, the comparative analysis is done first between the Return and the VIX index. Comparing the pair of values 0.67043 and 0.666955 with the value of the unconditional matrix of the variable in the respective first line, in this case, the value is 0.463661. As the pair of values of the conditional distribution is higher (0.67043 and 0.666955) than the value of the non-conditional distribution (0.463661). It denotes that the Return (row variable) commands the VIX indicator.

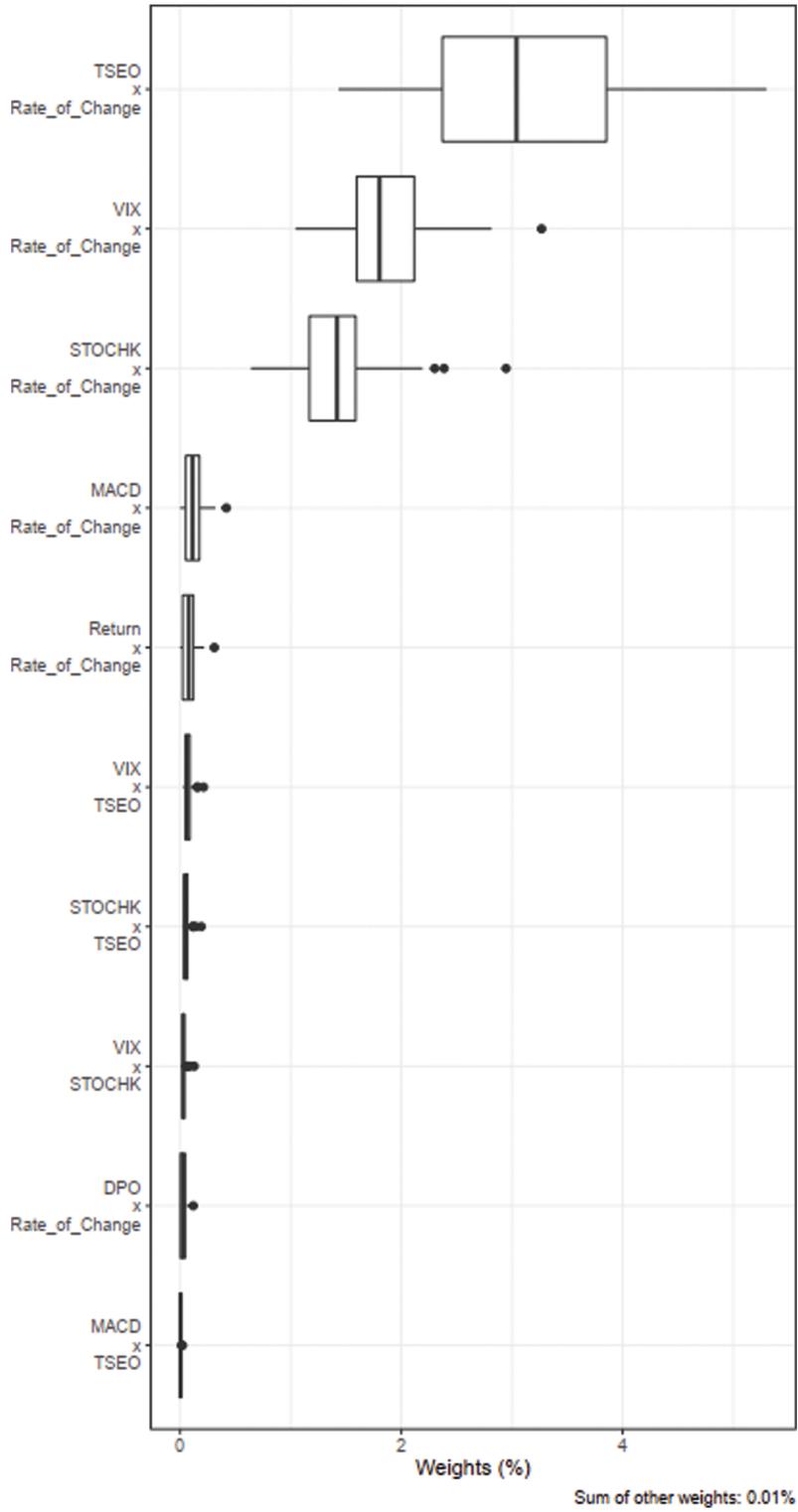


Figure 2: Major combinations of endogeneity weights for pairs

Table 4: Information entropy of the variables conditional distribution in a row by column variable (0.975 percentile)

Models	Return	VIX	CCI	MACD	WILL	STOCHK	TSEO	BB	CMO	DPO	ROC
Return		0.67043	0.671837	0.668645	0.670899	0.67038	0.672423	0.671562	0.67257	0.672148	0.669543
VIX	0.666955		0.67198	0.666432	0.670171	0.666054	0.663709	0.6724	0.671557	0.667841	0.6656
CCI	0.662021	0.66411		0.664263	0.671254	0.664399	0.662636	0.669693	0.670844	0.670168	0.663566
MACD	0.662722	0.660727	0.667484		0.671692	0.667577	0.665411	0.668615	0.670383	0.664149	0.661624
WILL	0.667713	0.670715	0.672397	0.671107		0.670921	0.672256	0.672064	0.672364	0.670875	0.669914
STOCHK	0.667811	0.66333	0.670817	0.66489	0.670503		0.665509	0.66999	0.669676	0.666369	0.662102
TSEO	0.669401	0.665986	0.671867	0.667043	0.671425	0.666335		0.671482	0.669443	0.667345	0.663775
BB	0.667419	0.662676	0.667669	0.666219	0.669436	0.662898	0.66341		0.670151	0.668615	0.668677
CMO	0.671495	0.669992	0.671355	0.671206	0.671542	0.672097	0.67098	0.671555		0.672991	0.672384
DPO	0.66378	0.664539	0.672385	0.666075	0.666029	0.664457	0.663712	0.670263	0.667739		0.66383
ROC	0.665457	0.665984	0.669618	0.663598	0.67095	0.663595	0.663228	0.668207	0.670552	0.66935	

Table 5: Information entropy of unconditional distributions (0.975 percentile)

Models	Return	VIX	CCI	MACD	WILL	STOCHK	TSEO	BB	CMO	DPO	ROC
Return		0.463661	0.480923	0.505515	0.527556	0.49639	0.450283	0.431047	0.510379	0.540525	0.512695
VIX	0.463661		0.49396	0.447094	0.508872	0.452949	0.463278	0.467673	0.511832	0.458567	0.441099
CCI	0.480923	0.49396		0.453365	0.489638	0.450432	0.482986	0.469742	0.509596	0.501697	0.490528
MACD	0.505515	0.447094	0.453365		0.497722	0.500555	0.478295	0.482133	0.530984	0.447742	0.462585
WILL	0.527556	0.508872	0.489638	0.497722		0.513575	0.528153	0.506991	0.493019	0.51471	0.526744
STOCHK	0.49639	0.452949	0.450432	0.500555	0.513575		0.528718	0.440973	0.510817	0.459325	0.480576
TSEO	0.450283	0.463278	0.482986	0.478295	0.528153	0.528718		0.471555	0.46035	0.501732	0.464979
BB	0.431047	0.467673	0.469742	0.482133	0.506991	0.440973	0.471555		0.462222	0.493179	0.505712
CMO	0.510379	0.511832	0.509596	0.530984	0.493019	0.510817	0.46035	0.462222		0.489453	0.534978
DPO	0.540525	0.458567	0.501697	0.447742	0.51471	0.459325	0.501732	0.493179	0.489453		0.479412
ROC	0.512695	0.441099	0.490528	0.462585	0.526744	0.480576	0.464979	0.505712	0.534978	0.479412	

Analogously, it is possible to analyze cause and effect relationships through the diagram shown in Fig. 3 where the arrows indicate the direction of causal relationships between the variables. The conclusion is that the return influences the variables CCI, MACD, WILL, STOCHK, TSEO, BB, DPO, ROC and VIX. Except for CMO indicators, we observed the predominant return domain.

When analyzing the VIX, the volatility controls all the other indicators except WILL, TSEO, ROC and Return. The VIX index known as the fear index is an exogenous variable calculated from the maturity of options and serves as a proxy for market risk.

The CCI controls only the variable BB and is therefore controlled by all others. Such a weak command relationship can be justified by the fact that the CCI signal is used to detect the initial and final trends. This signal has the characteristic of storing the lowest value compared to the other indicators. To interpret CCI indicator results, we use the concepts of overbought and oversold. It can be understood as if the market is overbought when it is above +100 and oversold when it is below -100. However, some investors use the movement of these values to understand the market strength. Breaking the value of +100 upwards can represent strength in the uptrend. However, when it returns below +100, could mean that the market is correcting the recent high, and the same goes for values below -100. We calculate CCI using 20 periods

The BB only has a direct influence on the ROC. If the stock's volatility increases, the BB tends to widen. Otherwise, the BB tends to narrow. It serves as an overbought or oversold indicator. For example, when the price is close to the upper band, there are signs of reversal. We built this indicator considering 20 periods for the moving average and 2 standard deviations for each band.

The CMO influences CCI, TSEO, BB, DPO, and Return. The Chande Momentum Oscillator is useful for determining the beginning of trends and is considered a modified relative strength index (RSI). The RSI is an oscillator-type technical indicator that measures the relationship between buying and selling forces of a given paper, ranging from 0 to 100. RSI signals regions indicate overbought and oversold. When the indicator is below a threshold, we have RSI_{Low}, commonly equal to 30. In this scenario that the share price is in an oversold zone, and the selling force is losing strength. This can be a sign that the share price will rise.

The DPO influences only the CCI and BB indicators. The Detrended Price Oscillator removes the trend component from the price time series by subtracting the moving average from the price over the price. We used ten for the number of periods of the moving average, and six for the number of periods of change in the moving average. Finally, the ROC can influence VIX, CCI, MACD, WILL, CMO, and DPO.

5 Conclusion

The present study aimed to verify the performance of the model of neural networks to predict returns in the Brazilian market. In addition, it investigated the information entropy from the variables of the neural network predictive model.

Confusion matrix analysis confirms hypothesis H_1 since the main model has an accuracy of about 76%. In other words, during 100 trading days, the model would have settled in 76 days, thus fulfilling the first research objective. The predictive ability of the model is slightly better for the shorts position (negative returns) compared to the predictive ability of positive return scenarios (long position).

Since the predictor variables present correlated residues, there is evidence of endogeneity. However, considering only the correlation analysis, it is not possible to support this statement. So, that's why we investigate the endogenous origin and information entropy in a prediction system of stock returns with the inputs given by technical indicators. We carried out an endogenous analysis between the technical indicators and the dependent variable Return.

From the residuals bootstrap of the 11 neural networks models combinations, we optimized the smallest possible variances using the minimum endogenous relationship variance methodology. We applied the minimization method through a nonlinear stochastic optimization process as presented in Eq. (1) within the methodology. Based on this method, we assign an importance degree to the models. Based on the theory of information entropy, it is expected that the model with the lowest variance has the highest degree of importance. In a normal situation, the expected value of importance for each model would be an equivalent weight for all 11 models, i.e., $(100\%)/11 = 9.09\%$. However, there is a difference between the expected and actual weight, indicating how well this model behaves better than others, or how much this variable dominates the others. Based on the results, we conclude that there is no balanced weights distribution among the 11 models under study.

The model with the lowest variance indicates good behavior and, consequently, will have greater weight. We did this by minimizing covariance with the variances. The input is a residual matrix, for which we solved models 1–11 by applying a bootstrap technique with 100 repetitions, generating 100 residuals for each model. So, we optimized this resulting in the optimum weights. With that, we obtained a behavior profile of the weights.

In model 1, the Rate of Change has the highest positive Olden's criterion of absolute importance, which was confirmed by the robustness test (Fig. 1). Finally, the dependent term of the main model, the return, is the variable that influences the largest number of indicators. Thus, the endogenous origin of the main neural network based on the technical indicators presented is favorable for the predictive purpose of the interested output, the Return.

The findings will help the target public to build investment strategies and verify strategy adherence in different risk environments. This work presents a tool for decision-making and, therefore, a practical and applicable contribution.

Apart from that, it is known that the Brazilian stock market is small compared to the American one, with few companies covered by analysts. Therefore, the findings related to the prediction of the returns are especially important to investors without access to market analysts, and for other individuals who want to learn about investment strategies from machine learning algorithms. Moreover, with quantitative strategies, it is possible to significantly reduce human interference in the decision-making process, eliminating behavioral biases that negatively impact investment returns. As a direction for further research, we suggest investigating the application of artificial intelligence algorithms to make the prior selection of technical variables that will be input to the predictive model. For this, it is possible to use a random forest model beforehand to select the variable improvements to be used as input to the neural network models.

Funding Statement: The authors received no specific funding for this study.

Conflicts of Interest: The authors declare that they have no conflicts of interest to report regarding the present study.

References

- [1] S. Akhterand and M. A. Misir, "Capital markets efficiency: Evidence from the emerging capital market with particular reference to Dhaka stock exchange," *South Asian Journal of Management*, vol. 12, pp. 35, 2005.
- [2] M. Ballings, D. Van den Poel, N. Hespels and R. Gryp, "Evaluating multiple classifiers for stock price direction prediction," *Expert Systems with Applications*, vol. 42, pp. 7046–7056, 2015. <https://doi.org/10.1016/j.eswa.2015.05.013>
- [3] M. Nabipour, P. Nayyeri, H. Jabani, A. Mosavi, E. Salwana *et al.*, "Deep learning for stock market prediction," *Entropy*, vol. 22, pp. 840, 2020. <https://doi.org/10.3390/e22080840>
- [4] D. Shah, H. Isah and F. Zulkernine, "Stock market analysis: A review and taxonomy of prediction techniques," *International Journal of Financial Studies*, vol. 7, pp. 26, 2019. <https://doi.org/10.3390/ijfs7020026>
- [5] C. -F. Tsai, Y. -C. Lin, D. C. Yen and Y. -M. Chen, "Predicting stock returns by classifier ensembles," *Applied Soft Computing*, vol. 11, pp. 2452–2459, 2011. <https://doi.org/10.1016/j.asoc.2010.10.001>
- [6] E. F. Fama and J. D. MacBeth, "Risk, return, and equilibrium: Empirical tests," *Journal of Political Economy*, vol. 81, pp. 607–636, 1973. <https://doi.org/10.1086/260061>
- [7] T. K. Lee, J. H. Cho, D. S. Kwon and S. Y. Sohn, "Global stock market investment strategies based on financial network indicators using machine learning techniques," *Expert Systems with Applications*, vol. 117, pp. 228–242, 2019. <https://doi.org/10.1016/j.eswa.2018.09.005>
- [8] X. Liu and D. D. Thomakos, ""Taps": A trading approach based on deterministic sign patterns," *Expert Systems with Applications*, vol. 175, pp. 114761, 2021. <https://doi.org/10.1016/j.eswa.2021.114761>
- [9] R. Garcia, D. Monte-Mor and N. Tardin, "Can accounting-based and market-based indicators predict changes in the risk rating of Brazilian banks?," *Revista Brasileira de Gest3o de Neg3cios*, vol. 21, pp. 152–168, 2019. <https://doi.org/10.7819/rbgn.v21i1.3968>

- [10] F. Ecer, S. Ardabili, S. S. Band and A. Mosavi, "Training multilayer perceptron with genetic algorithms and particle swarm optimization for modeling stock price index prediction," *Entropy*, vol. 22, pp. 1239, 2020. <https://doi.org/10.3390/e22111239>
- [11] Y. S. Abu-Mostafa, M. Magdon-Ismael and H. -T. Lin, *Learning from Data*, vol. 4. New York, USA: AMLBook, pp. 1–9, 2012.
- [12] S. Haykin and P. M. Engel, *Redes neurais: Princípios e prática*, 2nd ed., Porto Alegre, Brasil: Bookman, pp. 898, 2003.
- [13] D. P. Fonseca, P. F. Wanke and H. L. Correa, "A Two-stage fuzzy neural approach for credit risk assessment in a Brazilian credit card company," *Applied Soft Computing*, vol. 92, pp. 106329, 2020. <https://doi.org/10.1016/j.asoc.2020.106329>
- [14] A. Thakkar and K. Chaudhari, "A comprehensive survey on deep neural networks for stock market: The need, challenges, and future directions," *Expert Systems with Applications*, vol. 177, pp. 114800, 2021. <https://doi.org/10.1016/j.eswa.2021.114800>
- [15] B. Ripley, W. Venables and M. B. Ripley, "Package 'NNET': Feed-forward neural networks and multinomial log-linear models," *R Package Version*, vol. 7, pp. 3–12, 2016.
- [16] P. Carrand and L. Wu, "Variance risk premiums," *The Review of Financial Studies*, vol. 22, pp. 1311–1341, 2009. <https://doi.org/10.1093/rfs/hhn038>
- [17] S. L. Lambert, "Fundamental signals, future earnings and security analysts' efficient use of fundamental signals during 1991 through 2008," Ph.D. Dissertation, University of Texas, Arlington, Texas, USA, 2011.
- [18] J. Ulrich, "Package TTR: Technical trading rules," R package, 2021. <https://github.com/joshualrich/TTR>.
- [19] F. Lemos, "Análise técnica dos mercados financeiros," in *São Paulo*, Brasil: Saraiva Educação S.A., 2017.
- [20] A. D. Hill, S. G. Johnson, L. M. Greco, E. H. O'Boyle and S. L. Walter, "Endogeneity: A review and agenda for the methodology-practice divide affecting micro and macro research," *Journal of Management*, vol. 47, pp. 104–143, 2021. <https://doi.org/10.1177/0149206320960533>
- [21] L. A. B. de C. Barros, F. H. Castro, D. Silveira, A. D. Miceli and D. R. Bergmann, "Endogeneity in corporate finance empirical research (in Portuguese)," 2010. Available at Social Science Research Network: <https://doi.org/10.2139/ssrn.1593187>
- [22] A. B. Alves, P. Wanke, J. Antunes and Z. Chen, "Endogenous network efficiency, macroeconomy, and competition: Evidence from the Portuguese banking industry," *The North American Journal of Economics and Finance*, vol. 52, pp. 101114, 2020. <https://doi.org/10.1016/j.najef.2019.101114>
- [23] D. Ardia, K. Boudt, P. Carl, K. M. Mullen and B. G. Peterson, "Differential evolution with DEoptim: An application to non-convex portfolio optimization," *The R Journal*, vol. 3, pp. 27–34, 2011.
- [24] P. P. Shinde, K. S. Oza and R. K. Kamat, "Big data predictive analysis: Using R analytical tool," in *Int. Conf. on I-SMAC (IoT in Social, Mobile, Analytics and Cloud) (I-SMAC)*, Palladam, India, pp. 839–842, 2017.
- [25] M. M. Kumbure, C. Lohrmann, P. Luukka and J. Porras, "Machine learning techniques and data for stock market forecasting: A literature review," *Expert Systems with Applications*, vol. 197, 2022. <https://doi.org/10.1016/j.eswa.2022.116659>