

UNIVARIATE AND MULTIVARIATE ANALYSIS OF THE BOVESPA AND PETROBRAS INDICES BETWEEN 2005-2015

Maria Eugênia de Oliveira BATISTA¹
Rafaela da Silva GOMES¹
Luciene Resende GONÇALVES¹

- **ABSTRACT:** The Bovespa Index is the most important indicator of the average performance of quotations on the Brazilian stock market, portraying the behavior of the main securities traded on the Bovespa. The Petrobras Index refers to Petrobras quotations on the São Paulo Stock Exchange (Bovespa). In view of the expressiveness of these indices, the objective of this work will be to perform a univariate and multivariate temporal analysis of the series of returns of Petrobrás and Ibovespa in two periods. The first period is comprised between 2005 and 2015, while the second between 2015 and 2019. This division was intended to divide the pre and post crisis periods in the face of complaints of corruption against Petrobras, which caused a greater variation in liquidity in the company's shares. The survey is ongoing and only the analysis of the first period is complete. The series data were taken from the IPEADATA website (ipeadata.gov.br), then deflated, and then their returns were calculated using software R. This is done in two steps. In the first, the data are modeled using the ARIMA model, and in the second, the GARCH is fitted to the squared residues. For the Bovespa index series, an ARIMA (3.0.3) and then GARCH (1.1) were fitted. For the Petrobras series of returns, first, an incomplete ARMA was fitted with the autoregressive parameters of significant order 6 and the order 3 of moving averages also statistically significant. The volatility, in this case, was also fitted by a GARCH (1.1). The rates of the two returns were useful. A multivariate analysis indicated that, during the analyzed lag period, the Petrobras index did not directly influence the Bovespa index, although in the analyzed period it was chosen before the denunciation and corruption scandal. After analyzing the second period, a comparison will be made between the fits.
- **KEYWORDS:** Economic indexes; stock exchange, time series.

1 Introduction

The São Paulo Stock Exchange is an important stock trading center. It is a company responsible for organizing and managing bonds, securities and derivative contracts, in addition to providing registration, clearing and settlement services. Through the stock market it is possible to offer various products and services. Trading is conducted electronically through its main total return index; the Ibovespa.

¹Universidade Federal de Alfenas - UNIFAL-MG, Campus Varginha, ICSA, CEP: 37048-395, Varginha, MG, Brasil. E-mail: eugenia34@gmail.com; rafadasilva43@yahoo.com.br; luliresende.lr@gmail.com

Ibovespa's objective is to be the average performance indicator for the prices of the most tradable and representative assets in the Brazilian stock market, such as Petrobras's shares. In addition, he is very susceptible to the economic risks that permeate the foreign and domestic markets, as an example, we can mention the uncertainties of investors about the future of the Brazilian political scenario.

The Petrobras Index corresponds to the price of Petrobras's shares in the São Paulo Stock Exchange (Bovespa). Lemos (2006) points out that Petrobras's preferred shares (PETR4) correspond to one of the largest shares in the Bovespa Index and can indicate and explain the true cause and trend of Ibovespa.

Lemos (2006) points out that the possibility of a relationship between the Ibovespa and the behavior of Petrobras's shares can be considered because Petrobras's shares is representative of the institutional changes observed in the Brazilian stock market, for example, respect for minority shareholders, improvement in the level of corporate governance etc. According to Lemos (2006), it can also be established that the Ibovespa trend determines the behavior of PETR4 shares, since, for example, political and international crises are captured by Ibovespa and affect PETR4 shares, without, however, such factors being related to the fundamentals of this role.

Thus, this work aims to analyze through the time series methodology the univariate form of the behavior of the returns of these indexes in relation to the years 2005 to 2015. And to obtain future forecasts with the objective of helping in the decision making of shareholders and those who are interested on the Ibovespa and the Petrobras index. Later, the multivariate analysis will be carried out with the purpose of studying if there are relations between the indexes.

The article is arranged in five sections, this being the first. The second section will present basic concepts about Ibovespa, Petrobras and about the univariate and multivariate procedures used in the series. Then, the development of these procedures and the result obtained with the forecasts and multivariate VAR analysis will be presented. In the fourth section, the conclusion will be presented, and, finally, the bibliographic reference.

2 Theoretical frameworks

2.1 Ibovespa

According to Fortuna (2008), the Bovespa Index is the most important indicator of the average performance of quotations on the Brazilian stock market, because it portrays the behavior of the main shares traded on the Bovespa. The Bovespa (São Paulo Stock Exchange), therefore, is the largest stock trading center in Latin America (Pimenta Jr. & Higuchi, 2008).

The calculation of the index, according to Fortuna (2008), is done in real time by Bovespa, considering all the prices of the deals carried out in the spot market, with shares that are part of its portfolio. "The calculation can be made at any time, and it is nothing more than the sum of the weights (theoretical quantity of the stock multiplied by the last price of the same) of the shares included in its theoretical portfolio" (FORTUNA, 2008).

As it is shown by Lemos (2006), the Ibovespa acts as a driver, evaluating the return on shares in its portfolios and reflecting the capital gains, dividends and benefits that each share provides. The Ibovespa is relevant because it maintains its integrity and has not changed since its implementation in 1968, because it reflects the general performance of

the economy and because it is an early indicator of national economic activity (PIMENTA JR. and HIGUCHI, 2008).

B3 (which refers to the initial letters of Brasil, Bolsa, Balcão) emerged in its current format after the merger of BM & FBOVESPA with the Central of Custody and Financial Settlement of Securities (CETIP), approved by the Brazilian Securities and Exchange Commission (CVM) and by the Administrative Council for Economic Defense (CADE) on March 22, 2017 (ALERIG, 2017). In 2017, it was the fifth largest capital and financial market in the world, with assets of 13 billion dollars (MOREIRA, 2017).

B3 is linked to all Brazilian stock exchanges, including the Rio de Janeiro Stock Exchange (BVRJ), where only securities are traded. On August 29, 2014, the companies listed on the Bovespa beat the highest record in history in terms of market value, all 328 companies listed on the Brazilian stock exchange were worth R\$ 2.595 trillion (VEJA, 2014).

2.2 Petrobras

The Petrobras Index refers to Petrobras' quotations on the São Paulo Stock Exchange (Bovespa), this index started on Bovespa at the time the public offering of Petrobras shares was traded. According to Lemos (2006), Petrobras's preferred shares (PETR4) correspond to one of the largest shares in the Bovespa Index. In addition, PETR4 can implicitly indicate an evolution in the institutional structure and microstructure of the Brazilian stock market, which is the real reason for explaining the Ibovespa trend (LEMOS, 2006).

The story began in 1953, when President Getúlio Vargas signed Law 2004 that instituted the monopoly of exploration of oil and its derivatives to a new company, which received the name of *Petróleo Brasileiro S.A - Petrobras*. This creation occurred after many discussions about what would be the best policy to be adopted in relation to oil exports. At that time, oil consumption was greater than production, with this, there was an intensification of exploration and specialization of technicians in order to obtain expansion in a short period of time. For this purpose, in 1961, a new refinery was created in Rio de Janeiro, REDUC - Duque de Caxias Refinery. In 1963, the Research and Development Center (Cenpes) was created in Rio de Janeiro, where it became one of the most important centers of applied research in the world. In 1968, the first mobile oil platform was created, located in the Sergipe sea. Over time, several basins were found, and the expansion of oil exploration was taking place, allowing for increased production and export of oil. In the 1980s, Petrobras became a pioneer in deep water and the new Campo de Merlim Basin was discovered. In 1997, the monopoly fell, and the company became one of the largest oil companies in the world (PETROBRAS, 2015).

Beginning in the 2000s, Petrobras launched a new type of gasoline, considered the most advanced in the world; the company's first wind farm was inaugurated; Petrobras achieves self-sufficiency in the production of Oil and Gas; and in this decade oil production begins in the pre-salt deposits. There were, then, years of many investments and great importance in Petrobras's economic and sustainable history (PETROBRAS, 2015).

Currently, the company is going through a crisis of an internal and specific character that involves corruption scandals, overpricing of works and money laundering, which impacted the value of the company's shares. Due to the crisis, in January 2015, the

company's market value was 106.6 billion reais, less than the amount reached in the 2008 crisis. In addition to the internal crisis, Petrobras, as a state-controlled company, is constantly on the agenda of political and economic discussion. Government interference in the price of domestic oil products directly affects the company's cash flow and impacts on the country's indicators (FRAGA *et al.*, 2015). The scandals suffered brought different consequences for Petrobras, which revealed not only a serious distortion in the Brazilian representative practice and a failure in the mechanisms of power control. But also, the extensive renewal of political leaders in the National Congress and the alternation of power in the Executive in the 2018 elections. This has left Petrobras shares in decline, but there has been an increase in stock prices over the years, after the scandal.

2.3 Time Series

A time series is any set of observations ordered in time (MORETTIN, 2006). Sandes (2003) points out that the data of a time series are dependent, that is, the neighboring observations are important for its modeling, and with that, the temporal order of the facts becomes crucial.

With the study of the time series it is possible to analyze graphs, describe its behavior, as an example, check trend and seasonality and also make predictions of future values of the series so that, through this study it is possible to carry out the construction of models and a better description of these series. Among these models that are used for time series, there are the parametric - have a finite number of parameters and is performed by the temporal domain - and the non-parametric, conducted by the frequency domain (MORETTIN, 2006).

A time series can be written as a sum or multiplication of components, as in equations (1) and (2):

$$Z_t = T_t + S_t + \alpha_t \quad (1)$$

$$Z_t = T_t S_t \alpha_t \quad (2)$$

where Z_t = time observation at time t ; T_t = trend component that represents smoothed series values around a positive or negative slope; S_t = seasonal component that refers to an oscillatory movement linked to periodic variations, and α_t = zero mean random component and constant variance σ^2 .

According to Morettin and Tolo (2006), the main interest in considering a model of type (1) and (2) is to estimate the seasonality (S_t) and the trend (T_t) present in a series. These two components make the series behave non-stationary. Stationarity is a required condition for modeling this type of data and is achieved when the components mentioned are eliminated.

Since most series found in practice have some form of non-stationarity, a very useful function that helps to identify non-stationary components is the autocorrelation function.

In time series, the autocorrelation function is used to identify its properties, measuring the degree of correlation between the observations. It is given by the ratio between autocovariance and variance. According to Gujarati and Porter (2011), autocorrelation, or

sample correlation function (FAC), is one of the main ways of identifying the model, that is, it portrays the presence of stationarity in the series, together with its partial autocorrelation function (FACP). “FACP measures the correlation between time series observations that are separated by k periods, after checking the correlations in the intermediate lags (lags less than k)” (GUJARATI and PORTER, 2011). The graphical result of the autocorrelation is called a correlogram, and it can also be an important tool for identifying characteristics of the time series.

There are different models to suit a time series, and these are divided according to the number of parameters involved. In other words, parametric and non-parametric models. According to Morettin & Tolo (2006), in parametric models, the analysis is done in the domain of time. So, the most used models are the autoregressive and moving averages (ARMA), the integrated autoregressive and moving averages models (ARIMA), long memory models (ARIMA), structural models and non-linear models (ARCH) -GARCH). Now, in the class of non-parametric models, the most used are the autocovariance function and its Fourier transform, the spectrum.

The authors also emphasize that when the original data do not present significant results to be used in the model it is necessary to perform some transformations in these data so that it can find an adequate model for the research. Transformations generally occur when there is a need to stabilize the variance or make the seasonal effect additive. In some cases, doing a logarithmic transformation of the data may be appropriate. When the series is not stationary, the best way to correct this problem is to apply its differentiation in economic data, the best way is to use the differentiation in the original data. Each series has a specific transformation, so it is necessary to make a study on what will be necessary for the model to adapt in the future.

A characteristic of most financial time series is that in their level form they are random walks; that is, they are not stationary (GUJARATI and PORTER, 2011). The author also points out that a non-constant variance can have a high regressive structure. Where, over the different periods observed, heteroscedasticity can be self-correlated. Morettin and Tolo (2006) also argue that, in some economic and financial series, the use of returns becomes crucial, as these, as well as time series, have problems such as trend, seasonality and non-linearity. But a linear model, such as ARIMA, is not suitable to describe such behavior.

With this, the returns of these series are used so that the model can be fitted. The models used are non-linear (ARCH and GARCH). In this work the ARIMA linear model will be presented, and the ARCH and GARCH nonlinear models, which will be used as fits to the chosen return series

2.4 Univariate process

2.4.1 Returns

Normally, R_t is expressed as a percentage relative to the period and is also called as rate of return. Morettin (2011) also states that it is preferable to use returns as they are free of scale and have more interesting statistical properties, such as stationarity and ergodicity so that they can be modeled.

The log-return, in a period k , is given by

$$R_t[k] = \log \frac{P_t}{P_{t-k}}. \quad (3.1)$$

According to Rosa (2011), the return series has some specific characteristics that are called stylized facts, such as: absence of autocorrelation; asymmetry; excess kurtosis (heavier tails, compared to normal distribution); volatility grouping; nonlinearities.

Morettin (2011) states that a series of feedback observed in an instant of equally spaced times can be considered part of a stochastic process. And for any instant of time t_1, \dots, t_n and any $n \geq 1$, their distributions must satisfy certain conditions that, in practice, are difficult to meet. So, for this, it is necessary to characterize the process up to a certain order, such as the mean and the autocovariance function

Other assumptions can be introduced, such as conditions of stationarity, ergodicity or normality of the process "(MORETTIN, 2011). Prices in general are not stationary, but their log-returns are, but the condition of normality is not valid for Ergodicity is the most difficult to establish, since characteristics of interest are estimated from a single trajectory of the process.

2.4.2 Linear model: ARIMA

The methodology of Box (JENKINS and REISEL, 1994) is widely used in the analysis of parametric models and basically consists of fitting the ARIMA model (p, d, q), which means, integrated autoregressive models of moving averages to a set of data. (MORETTIN and TOLOI, 2006). Such models are also characterized by being simple and parsimonious; forecasts are quite accurate, comparing favorably with other forecasting methods. They are classified in stationary and non-stationary linear models.

The Box-Jenkins methodology also reports that the models are chosen through some strategy based on an iterative cycle where the structure of the model is based on the data itself. In this cycle, a general class of models is used first which will be considered in the analysis; then, there is identification of which model best fits those data, taking into account the analysis of autocorrelation, partial autocorrelation and other criteria; in the next step, the parameters are identified and estimated; and, finally, the model is diagnosed and verified through an analysis of residues to see if there is adequacy of the model so that it can be used in the research. (MORETTIN and TOLOI, 2006).

The ARIMA model (p, d, q) corresponds to an ARMA model, but the difference is that its time series will be differentiated d times. In other words, it is an integrated autoregressive series of moving averages. Where p are the autoregressive terms, d the number of times the series will be differentiated before it becomes stationary and q the number of moving averages. If $d = 0$, the series becomes stationary for the ARMA model (p, q) (GUJARATI and PORTER, 2011).

The ARIMA model equation (p, d, q) is given by:

$$\phi(B)\Delta^d Z_t = \theta(B)a_t \quad (4)$$

where $\phi(B) = (1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)$ is the polynomial of moving averages of order q , B is the delay operator being $B^m Z_t = Z_{t-m}$, where m is any lag, Δ is the difference operator defined by $\Delta^d Z_t = (1 - B)^d Z_t$, and d is the number of differences needed to make the series stationary.

The conditions of existence of the ARIMA model (p, d, q) are stationarity and invertibility, that is $\phi(B) = 0$ e $\theta(B) = 0$ and must be outside the unit circle.

An ARIMA process $(p, 0, 0)$ means a purely stationary AR (p) process; an ARIMA $(0, 0, q)$ means a purely stationary MA (q) process. Given the values of p, d and q , it is possible to say which process is being modeled (GUJARATI and PORTER, 2011).

2.4.3 Non-linear models: ARCH and GARCH

Nonlinear models are more suitable for financial series, which have conditional variance evolving over time. That being, the ARCH (“autoregressive conditional heteroscedasticity”) and GARCH (“generalized ARCH”) models will be used. The goal will be to model what is called volatility, which is the conditional variance of a variable, commonly a return (Morettin and Toloi, 2006, p. 361).

The ARCH model - auto-regressive with conditional heteroscedasticity - has as its basic idea that the return X_t is not correlated serially, but the volatility depends on returns passed through a quadratic function. The first step in identifying the ARCH model is to try to fit an ARMA model, so that the serial correlation of the series is removed if there is any (MORETTIN and TOLOI, 2006). ARCH models allow errors to be uncorrelated but dependent, that is, $Cov(r_t^2, r_{t-1}^2)$ could be non-null.

An ARCH(m) model is defined as

$$r_t = \sqrt{h_t} \varepsilon_t \quad (5.1)$$

$$h_t = \alpha_0 + \alpha_1 r_{t-1}^2 + \dots + \alpha_m r_{t-m}^2 \quad (5.2)$$

where ε_t is independent with a mean of zero, $\alpha > 0$, $\alpha_i \geq 0$, $i = 1, \dots, m-1$; $\alpha_m > 0$. But, in practice, we assume that $\varepsilon_t \sim N(0,1)$, $\varepsilon_t \sim t_\nu$ or a distribution that best writes the heavy tails of financial series.

In the case of financial series, the more restrictions the returns process imposes, the more restrictions there are for the model coefficients (MORETTIN, 2011).

The author also points out that, when we admit that r_t follows an ARCH model, the tails will be heavier than the normal ones, characterizing an advantage to the model. But its disadvantage is that the model treats both positive and negative returns in a similar way, as the squares of returns fall into the same volatility formula, and some large, isolated returns can lead to over-predictions.

When we assume that $\varepsilon_t \sim N(0,1)$ we see that X is a random variable with distribution $X^2(1)$, which shows that there is an AR (1) model for r_t^2 , but with non-Gaussian errors. For the ARCH model (m), we have

$$r_t^2 = \alpha_0 + \sum_{i=1}^m \alpha_i r_{t-i}^2 + v_i, \quad (5.3)$$

where we have an AR (m) model for r_t^2 , with non-Gaussian innovations, if ε_t is gaussian.

In addition, it can be demonstrated that the r_t returns also form a white noise with variance given by

$$\text{Var}(r_t) = \frac{\alpha_0}{1 - \sum_{i=1}^m \alpha_i}. \quad (5.4)$$

To verify whether the series has conditional heteroscedasticity, two tests can be used, examining the series r_t : Box-Pierce-Ljung test for r_t^2 , and Lagrange multiplier (ML) test.

These tests will be performed with the purpose of examining the F statistic, that is, if it is significant, it is concluded that there is conditional heteroscedasticity in the series. In addition, the estimators of the model parameters are obtained by the conditional maximum likelihood method.

Morettin (2011) points out that for an ARCH (m) model, with ε_t normal or t-Student, residuals

$$\tilde{r}_t = \frac{r_t}{\sqrt{h_t}} \quad (5.5)$$

(standardized) are independent random variables with standard or t-Student normal distribution.

After performing all the necessary procedures of the ARMA and ARCH models, that is, trying to fit the ARMA model to remove the serial correlation from the series, finding a suitable parameter for the ARCH model (m) and removing the conditional heteroscedasticity in the residues, finally, the forecasts can be made.

"The volatility forecasts using the ARCH (m) model are obtained recursively" (MORETTIN, 2011).

This way

$$\hat{h}_t(1) = \alpha_0 + \alpha_1 r_t^2 + \dots + r_{t-m+1}^2 \quad (5.6)$$

is the forecast of h_{t+1} , with origin fixed in the instant t . The forecasts ℓ steps forward, originating in t , are given by

$$\hat{h}_t(\ell) = \alpha_0 + \sum_{i=1}^m \alpha_i \hat{h}_t(\ell - 1), \quad (5.7)$$

where $\hat{h}_t(\ell - 1) = r_{t+l+1}^2$, se $\ell - i \leq 0$.

Morettin (2011) states that the GARCH (generalized ARCH) model can be used to describe volatility with fewer parameters than an ARCH model. And is defined by

$$r_t = \sqrt{h_t} \varepsilon_t \quad (5.8)$$

$$h_t = a_0 + \sum_{i=1}^m \alpha_i r_{t-1}^2 + \sum_{j=1}^n \beta_j h_{t-j} \quad (5.9)$$

where ε_t are random independent variables, with mean zero, $a_0 > 0, a_1 \geq 0, i = 1, \dots, m - 1, \beta_j \geq 0, j = 1, \dots, n - 1, a_m > 0, \beta_n > 0, \sum_{i=1}^q (a_i + \beta_i) < 1, q = \max(m, n)$.

The author also states that the advantages and disadvantages are the same in relation to the ARCH model. Identifying the order of a GARCH model to be fitted to a real series is usually difficult, so it is recommended to use models with a low order.

Morettin (2011) points out that the volatility forecasts can be calculated similar to the ARMA model, and are given by

$$\hat{h}_t(\ell) = a_0 + (a_1 + \beta_1) \hat{h}_t(\ell - 1), \ell > 1 \quad (6)$$

ARCH models have many advantages in analyzing returns on financial assets. The main advantages are (ANDERSEN *et al.*, 2014):

The model can produce volatility clusters (ANDERSEN *et al.*, 2014).

The model shocks, represented by ε_t , have heavy tails; (ANDERSEN *et al.*, 2014):

However, there are also some weaknesses, which are:

The model assumes that positive and negative shocks have the same effect on volatility because it depends only on squares of past shocks. In practice, we know that the prices of financial assets respond differently to positive or negative shocks (COSTA, 2019).

The ARCH model is quite restrictive. In practice, such a restriction limits the ability of the ARCH model with errors to distribute itself normally in capturing the excess kurtosis commonly found in the distribution of financial asset returns (COSTA, 2019).

The ARCH model does not provide any insight to understand the source of variation in a financial time series. It only provides a way to describe the behavior of the conditional variance (COSTA, 2019).

3 METHODOLOGY

The methodology to be used is that of time series. The data correspond to daily returns of the Ibovespa and Petrobras series between the years 2005 and 2015, taken from the database of the *ipeadata* and *economática* sites. First, the original data was deflated in order to remove the effect of inflation from the results. Then, the series returns, as well as all univariate analysis, were performed using the R software.

The autoregressive vector model (VAR) is a methodology in which it is an extension of a univariate regression to a multivariate environment, where each equation defined by the VAR is a regression by ordinary least squares of a given variable in variables that are outdated by itself and by other variables that make up the model, in which this method will be very important for the understanding of the research done.

The univariate procedure initially requires the construction and analysis of the graphs of the original series, the autocorrelation function and the periodogram. These graphs will be useful in identifying the components present and identifying the model to be fitted. The returns were used because they have better characteristics in the analysis of financial data.

Thus, the ARIMA model (p, d, q) will be the first procedure to be fitted, where different models of autoregressive orders and moving averages will be tested to obtain the best result. The diagnosis of the fits will be made through the residual autocorrelation test, criterion of the lowest AIC value and the Box Pierce test, which will then allow the fit of the second model, since the serial correlation of the series will be removed.

The GARCH model, used with the objective of modeling the volatility present in returns, will remove the conditional heteroscedasticity in the residues, allowing the best fit of the model, based on the Box-Pierce-Ljung test, of the squared residues. Thus, through the fits of the models, it is possible to carry out forecasts of the returns of the Ibovespa and the Petrobras index.

In the multivariate procedure, the GRETl software will be used to structure the VAR model. First, the number of lags in the model will be chosen according to the criteria of Akaike and Bayesian information. Once the bilateral causality is verified, the VAR will be used to verify the effect between the series, that is, to verify whether the Ibovespa variations influence the Petrobras index or vice versa.

In the initial hypothesis, a great similarity between the forecasts analyzed by the univariate process is expected, through the fits of the models, with the real values found in the Ibovespa and Petrobras indexes. In addition, in the multivariate analysis, a direct relationship between the indices is expected, that is, the Petrobras index is expected to directly influence the Ibovespa values, mainly because PETR4 corresponds to one of the largest shares traded on the Stock Exchange.

4 RESULTS AND DISCUSSION

4.1 Univariate Analysis of the Bovespa and Petrobras indexes

Analyzing graphically the original data of the Bovespa and Petrobras Indices, they have a similar structure. This is because they are related, that is, Petrobras as one of the largest companies listed on the Stock Exchange, has great influence within it. Thus, variations in its index can influence variations in the Bovespa index, respectively.

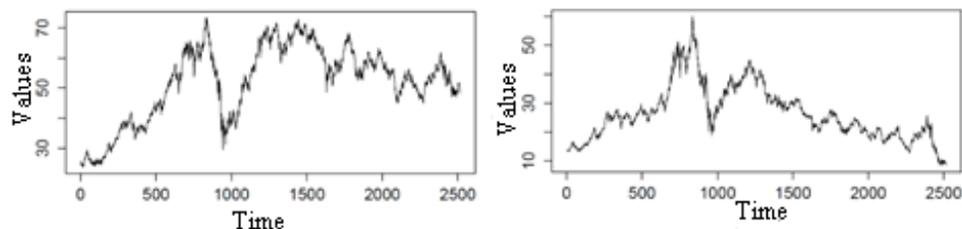


Figure 1 - Original graphics of the Bovespa and Petrobras index series.

In addition, it can be said that the stock market is extremely influential in relation to everything that occurs in the world such as external and internal crises, political events and also shocks in the economy, in Brazil or in any other country.

According to the behavior of the graphs above, between 1994 and 2007, both are on the rise; there are some decreases, but none truly relevant. The most noticeable point of decline in both charts is in 2008, when the housing crisis in the United States was occurring.

The crisis had a direct impact on several economies in the world due to the decrease in imports from the USA. There was an outflow of capital and an increase in the prices of goods traded internationally. This caused a devaluation in the exchange rate and, consequently, impacted inflation, mainly in economies with overvalued currencies, as is the case of Brazil (Carcanholo, *et al.*, 2008).

Brazil was impacted by the crisis for two reasons: the price of commodities (goods), due to its high dependence on exports from the United States. And capital flows, which determines the current account deficit necessary to keep rates high investments and Brazilian economic growth (PASTORE and PINOTTI, 2008). In addition, according to Alberini and Boguszewski (2008), the first impact was registered by the São Paulo Stock Exchange, which dropped by almost 10 over the course of a day in the second half of 2007, and, consequently, Petrobras also suffered the impacts of the crisis.

Another noticeable point in the Petrobras index graph corresponds to the end of the analyzed period, between the end of 2014 and the beginning of 2015, which corresponded to a drop in the index, a period in which the accusations of corruption against the company intensified. In January 2015, Petrobras indicated that it may not pay dividends from 2014 to shareholders due to the financial and political crisis in the company, and outlined a difficult outlook for the near future, frustrating expectations of analysts and investors and causing a fall in shares (FRAGA *et al.*, 2015).

To use the necessary statistical model in economic series, as in the case of the Bovespa and Petrobras indexes, some changes in their original series are important. In this work, the data were deflated and the rates of return will be calculated, which are free of scale and have more interesting statistical properties, as well as the variances over a certain period, so that the ARCH-GARCH model can be used.

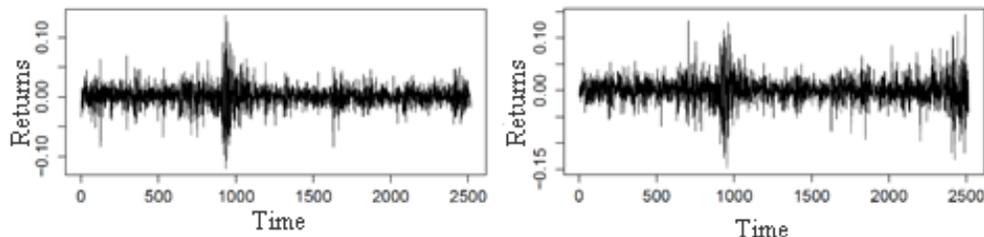


Figure 2 - Returns from the Ibovespa and Petrobras series.

After the returns are calculated, the index fitting process begins. First, the autocorrelation and partial autocorrelation graphs are generated to detect the autoregressive and moving average parameters and to identify the possible fits to be made in the ARIMA model.

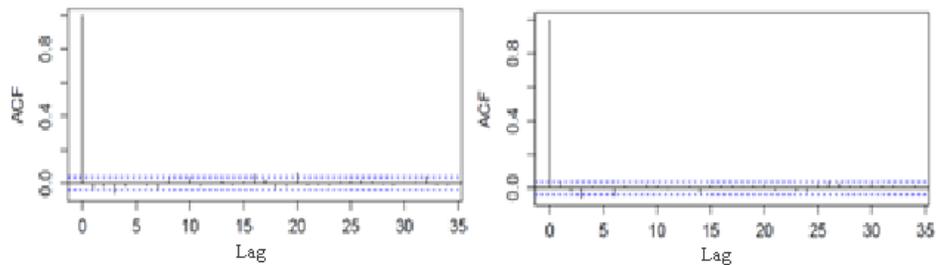


Figure 3 - Autocorrelation Ibovespa and Petrobras.

The graphs generated show significant lags at orders 3, 16 and 20 for the Bovespa Index, and orders 3 and 6 for Petrobras. These values will be used to make possible fits to moving averages of the models.

The partial autocorrelation graphs, which reproduce the autoregressive values to be fitted, result in the possible values of order 3, 7, 16 and 20 for the Ibovespa, and 3 and 6 for the Petrobras return.

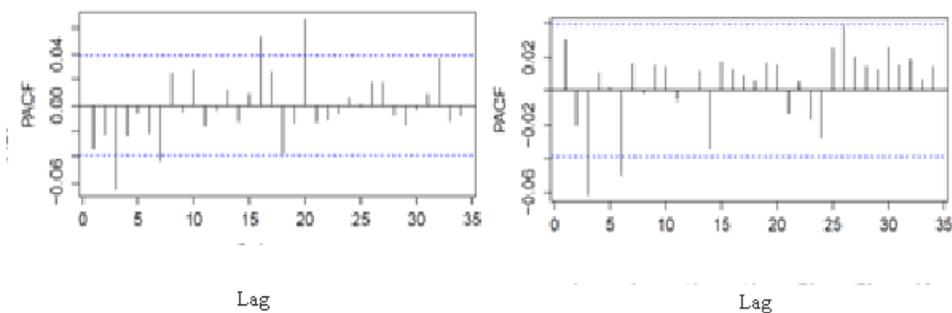


Figure 4 - Partial Autocorrelation Ibovespa and Petrobras.

From the figures, attempts to fit the models begin.

Regarding the fit of Ibovespa returns, the best model tested was ARMA (3.0.3), that is, it has the third autoregressive parameter and significant moving averages. Petrobras' returns were fitted by an incomplete ARIMA model with autoregressive parameters of order 6 and order 3 of moving averages. The fits of the models are proven through the graphs of the residuals of the returns, as it is through them that the loss of memory of the significant parameters is perceived. These are willing to follow.

The tests performed that confirmed the best fits were through the lowest values of AIC and Ljung-Box. The next step is to calculate the residuals squared so that the volatility of the financial series can be fitted using GARCH models.

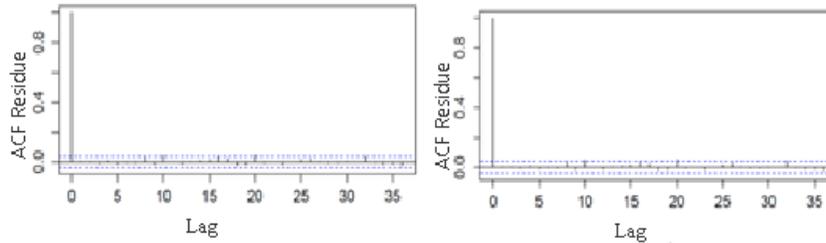


Figure 5 - Ibovespa and Petrobras waste.

The squared waste graphs in financial series have the structures shown below, as they have specific characteristics.

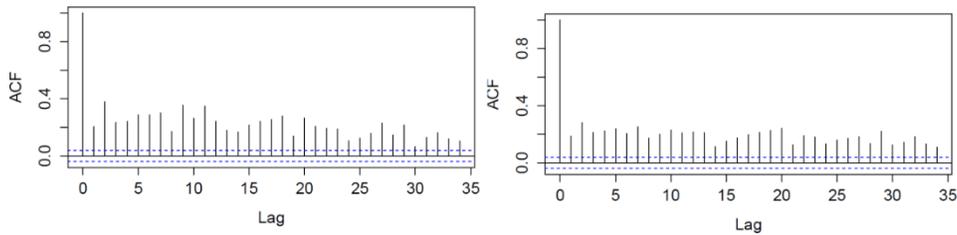


Figure 6 - Ibovespa and Petrobras squared waste.

The best fits obtained in the GARCH models were those corresponding to the volatility (1.1) of the two series, Ibovespa and Petrobras, that is, both series were fitted by the same GARCH model.

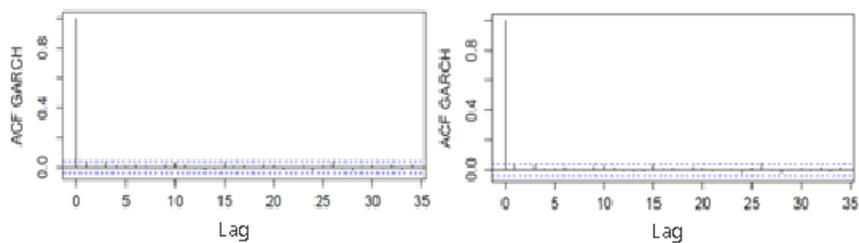


Figure 7 - Fit of GARCH models (1.1) for Ibovespa and Petrobras.

Through the fits made, it is clear that the series have lost their memories, are stationary and are suitable for the last procedure of the univariate process: calculating the forecasts, which are shown in Tables 1 and 2.

First, the return calculations of the Ibovespa and Petrobras series were performed. After finding the values, six values of the series found were reserved so that it was possible to carry out the forecast, as well as comparing these values to obtain the results.

To calculate the predictions of the original series, the reverse process of calculating the return was performed, that is, the procedure done to calculate the return was inverted, so it was possible to obtain the original values. Predictions were obtained using the same procedure as returns.

Table 1- Values of returns and originals calculated and forecasted on the Bovespa Index

Calculated Returns	Expected Returns	Original Values	Expected Values
0.005551009	0.0005809004	50.6333	55.03871
-0.01642925	0.0005809004	51.2783	55.03871
-0.002042983	0.0005809004	51.2923	55.03871
-0.007653556	0.0005809004	51.2353	55.03871
-0.01613556e	0.0005809004	51.2783	55.03871
-0.01822075	0.0005809004	51.8723	55.03871

According to Table 1, it can be concluded that the predictions of the original values are optimistic compared to the real values of the series. There was an overestimation of the values, but there was no big difference to the point of discarding the analysis made and concluding that the software used is not suitable for making the predictions.

Table 2- Values of returns and originals calculated and predicted from the Petrobras Index

Calculated Returns	Expected Returns	Original Values	Expected Values
-0.0507394	0.0003668681	9.139865028	25.09522
-0.0502927	0.0003668681	9.615583363	25.09522
-0.0129290	0.0003668681	10.11154503	25.09522
0.037246	0.0003668681	10.2431267	25.09522
0.0082389	0.0003668681	9.86862503	25.09522
0.0187897	0.0003668681	9.787651697	25.09522

Table 2 reveals that the predictions, compared to the values of the original series, have a very large difference. This result can be explained due to the time when Petrobras was passing in early 2015, values that are being used in the analysis. Thus, it can be justified that the software is not capable of detecting moments of crisis in the company, and, for this reason, they overestimate future values based on past values.

4.2 Multivariate analysis of the Bovespa and Petrobras indexes

The multivariate analysis of the indices was calculated using the GRETL software. First, a VAR gap selection test was performed to decide the number of lags used in the analysis. The Engle Granjer cointegration test together with the Akaike criterion (AIC)

presented the value of five lags as significant according to the information criteria, and the results are presented in Tables 3 and 4:

Table 3 - Ibovespa Vector Autoregression

Variable	Coefficient	Standard error	P-value
Constant	0.18919	0.0786852	0.01627**
Ibovespa_1	0.97724	0.0312868	0.00001***
Ibovespa_2	0.0125223	0.0443482	0.77769
Ibovespa_3	0.0166101	0.0443387	0.70798
Ibovespa_4	-0.0548449	0.0442183	0.21497
Ibovespa_5	0.0449224	0.0311544	0.14945
Petrobras_1	-0.0342424	0.0407489	0.40081
Petrobras_2	0.0311249	0.0597011	0.60217
Petrobras_3	-0.10021	0.0598524	0.09420*
Petrobras_4	0.16149	0.0596743	0.00685***
Petrobras_5	-0.0579366	0.0407501	0.15522

The significance level of 10 percent was used, as the results showed up to this approximately significant value and to summarize the results clearly and concisely this level was used.

Table 3 contains information regarding the relationship between the Bovespa index and past values of the index itself and also of the Petrobras index. It can be concluded that the lagged values of three and four days prior to the daily returns of the Petrobras index influence the Ibovespa with a significance of 10 percent. And Ibovespa itself is influenced by outdated values of one day by its own index.

Table 4 depicts the influence of the lagged values of the Bovespa and Petrobras index on Petrobras' values. It is concluded that most of the lagged values of the Petrobras index influence the values of the index itself, but only influence the fourth day lagged in the results of the Bovespa index, at a significance of 10 percent.

The explanation for the results obtained in this work may be related to the moment when Petrobras was passing in early 2015, a period in which the accusations of corruption against the company intensified, causing a greater variation in the liquidity of its shares. Therefore, it cannot be said that Petrobras' shares systematically influence the market to the point of generating major impacts, on the contrary, the market acts independently of the behavior of these shares, with regard to the period from January 2005 to March 2015.

Table 4 - Petrobras' Vector autoregression

Variable	Coefficient	Standard error	P-value
Constant	0.126253	0.0603865	0.03665**
Ibovespa_1	-0.0264295	0.0240108	0.27112
Ibovespa_2	0.0433242	0.0340348	0.20316
Ibovespa_3	0.0341191	0.0340274	0.31611
Ibovespa_4	-0.0661942	0.0339351	0.05121*
Ibovespa_5	0.0131586	0.0239092	0.58212
Petrobras_1	1.05114	0.0312725	0.00001***
Petrobras_2	-0.0904043	0.0458172	0.04859**
Petrobras_3	-0.0724014	0.0459333	0.1151
Petrobras_4	0.163001	0.0457966	0.00038***
Petrobras_5	-0.0522141	0.0312734	0.09512*

Conclusions

Ibovespa's objective is to be the average performance indicator for the prices of the most tradable and representative assets in the Brazilian stock market, such as Petrobras' shares. The Petrobras Index corresponds to Petrobras' quotes on the São Paulo Stock Exchange (Bovespa). Lemos (2006) points out that Petrobras' preferred shares (PETR4) correspond to one of the largest shares in the Bovespa Index and can indicate and explain the true cause and trend of Ibovespa.

Based on these indices, univariate and multivariate analyzes were performed to find predictions from these series and to search for relationships between them. First, the returns were calculated to improve the fit of the econometric models in obtaining the forecasts. The returns are used because they are free of scale and have more interesting statistical properties than the original series. In the modeling of these returns, fits from the ARCH family are used, which at first receive an ARIMA fit and then the GARCH is modeled.

The fit of Ibovespa was based on the ARIMA (3.0.3) and GARCH (1.1) models; and the Petrobras index obtained an incomplete ARMA fit with autoregressive parameters of significant order 6 and that of moving order 3 also significant. The volatility, in this case, was also fitted by a GARCH (1.1). The forecasts obtained from the fits were not exact, but in the case of Ibovespa there was a closer approximation to reality. In the Petrobras index, the results were not expressive, as there is a considerable overestimation between the forecasts and the original values obtained.

Regarding the multivariate analysis, it can be seen that the results obtained come into conflict with several ideas that directly relate the Petrobras index to the Ibovespa and the initial hypothesis, which predicted a direct relationship between the Petrobras index and the Ibovespa.

It is concluded that, in the analyzed lag period, the lack of influence of Petrobras shares in relation to the Bovespa index, is that in the period from 2005 to 2015 there were no significant changes in the shares they grew according to the economy, like other companies

that were on the stock exchange in that period and all scandals and corruptions have occurred since 2015 involving Petrobras. So, in view of this situation, we intend to analyze from 2015 to the end of 2020 and see if there is any influence, as the literature mentions this period more than what was studied in this research.

Acknowledgements

We thank FAPEMIG for the scientific initiation scholarship for the development of the article, and reviewers and editors for their comments e suggestions.

BATISTA, M. E. O., GOMES, R. S., GONÇALVES, L. R. Análise uni e multivariada dos índices Bovespa e Petrobras entre 2005-2015. *Rev. Bras. Biom.* Lavras, v.39, n.1, p.139-157, 2021.

- *RESUMO: O Índice Bovespa é o mais importante indicador do desempenho médio das cotações do mercado de ações brasileiro, retratando o comportamento dos principais papéis negociados na Bovespa. O Índice Petrobras diz respeito às cotações da Petrobras dentro da bolsa de valores de São Paulo (Bovespa). Tendo em vista a expressividade desses índices, o objetivo deste trabalho será fazer uma análise temporal uni e multivariada das séries de retornos da Petrobrás e do Ibovespa em dois períodos. O primeiro período está compreendido entre 2005 e 2015, enquanto o segundo entre 2015 e 2019. Essa divisão teve por finalidade dividir os períodos pré e pós crise frente as denúncias de corrupção contra a Petrobrás o que provocou uma maior variação de liquidez nas ações da empresa. A pesquisa está em andamento e somente a análise do primeiro período está concluída. Os dados das séries foram retirados do site IPEADATA (ipeadata.gov.br), em seguida foram deflacionados, e, depois, foram calculados seus retornos no software R. Na modelagem de um retorno, modela-se a volatilidade por meio da classe dos modelos ARCH. Isso é feito em duas etapas. Na primeira, modelam-se os dados por meio do modelo ARIMA, e, na segunda, ajusta-se o GARCH aos resíduos ao quadrado. Para a série do índice Bovespa, foi ajustado um ARIMA (3,0,3) e então o GARCH (1,1). Já para a série de retornos da Petrobrás, ajustou-se, primeiramente, um ARMA incompleto com os parâmetros autoregressivos de ordem 6 significativo e o de ordem 3 de médias móveis também significativo. A volatilidade, neste caso, também foi ajustada por um GARCH (1,1). As previsões dos dois retornos foram então obtidas. A análise multivariada indicou que, no período de defasagem analisado, o índice Petrobras não influenciou diretamente no índice Bovespa. Após a análise do segundo período será realizada uma comparação entre os ajustes.*
- *PALAVRAS-CHAVE: Índices econômicos; bolsa de valores; séries temporais.*

REFERENCES

- ALERIGI JUNIOR, A. *CVM aprova fusão entre BM&FBovespa e Cetip* [Versão Eletrônica]. *Investing.com*. 22 mar., 2017. Available in: <<https://br.investing.com/news/mercado-de-a%C3%A7%C3%B5es-e-financeiro/cvm-aprova-fus%C3%A3o-entre-bm-amp;fbovespa-e-cetip-234503>>. Access in: January 20, 2020.
- ANDERSEN, T.; BOLLERSLEV, T.; HADI, A. *ARCH and GARCH models*. New York: John Wiley & Sons, 2014. 244p.
- BOX, G. E.; JENKINS, G. M.; REINSEL, G. C.; LJUNG, G. M. *Time series analysis: forecasting and control*. 3.ed. New York: John Wiley & Sons, 1994. 612p.
- BUENO, R. D. L. S. *Econometria de séries temporais*. Porto Alegre: Cengage Learning, 2008. 285p.
- CARCANHOLO, A. *Crise financeira Internacional: Rio de Janeiro: Natureza e impacto*, 2008.
- COSTA, H. C. *Modelos ARCH e GARCH*. *RPubs*, c2019. Available in: <<https://rpubs.com/hudsonchavs/archegarch>>. Access in: January 14, 2021.
- FORTUNA, E. *Mercado Financeiro: produtos e serviços*. 17.ed. Rio de Janeiro: Qualitymark. 2008.
- GRETLL.2020. *Acrônimo de GNU regression, econometrics and time-series library. Econometric software for the GNU generation*. Available in: <<http://gretl.sourceforge.net/pt.html>>. Access in: January 20, 2020.
- GUJARATI, A.; DAMONAR, N.; PORTER, A.; DAWN, C. *Econometria básica*. 5.ed. Porto Alegre: AMGH, 2011. 920p.
- HILL, C. et. al, *Econometria*. 1.ed. São Paulo: Saraiva, 1999.
- LEMOS, B. P. Comportamento Recente do Índice Bovespa. *Revista Economia & Tecnologia*, v.7, n.2, p. 163-173, 2006.
- MOREIRA, M. *Fusão entre BM&FBovespa e Cetip cria a B3, 5ª maior bolsa de valores do mundo*. [Versão Eletrônica]. Agência Brasil. 30 mar., 2017. Available in: <<http://agenciabrasil.ebc.com.br/economia/noticia/2017-03/fusao-entre-bmfbovespa-e-cetip-cria-b3-5a-maior-bolsa-de-valores-do-mundo>>. Access in: January 23, 2020.
- MORRETIN, P. A. *Econometria financeira*. 2.ed. São Paulo: Egard Blucher, 2011.
- MORRETIN, P. A.; TOLOI C. M. *Análise de séries temporais*. 2.ed. São Paulo: Egard Blucher, 2006.
- PASTORE, A. C.; PLNOTTI, M. C. *A crise de 2008 e o Brasil*. Rio de Janeiro, 2008.
- PIMENTA JUNIOR, T.; HIGUCHI, R. H. *Variáveis macroeconômicas e o ibovespa: um estudo da relação de causalidade*. 60.ed. São Paulo: READ, 2008.
- R CORE TEAM. *R: A Language and environment for statistical computing*. Vienna, Austria: R Foundation for Statistical Computing. Available in: <<http://www.R-project.org/>>. Access in: January 24, 2020.

ELLERS, R. S. *Análise de séries temporais*. Universidade Federal do Paraná: Laboratório de Estatística e Geoinformação, 2003.

SILVEIRA, V. G. et.al, The scandals of the petrobras and the Brazilian market: is there dynamic correlation? *Revista Business and Management Review*, v. 05, n.1, 2015.

VEJA. *Empresas brasileiras batem recorde de valor de mercado*. [Versão Eletrônica]. *Veja*. 1set. 2014 São Paulo: Editora Abril. Available in: <<https://veja.abril.com.br/economia/empresas-brasileiras-batem-recorde-de-valor-de-mercado/>>. Access in: January 22, 2020.

Received on 31.08.2020

Approved after revised on 17.02.2021