NOVA IMS Information Management School

MEGI

Mestrado em Estatística e Gestão de Informação
Master Program in Statistics and Information Management

Forecasting the Stock Market Using ARIMA and ARCH/GARCH Approaches

Felipe Bardelli Dinardi

Dissertation presented as partial requirement for obtaining the Master's degree in Statistics and Information Management

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

NOVA Information Management School
Instituto Superior de Estatística e Gestão de Informação
Universidade Nova de Lisboa

FORECASTING THE STOCK MARKET USING ARIMA AND ARCH/GARCH APPROACHES

by

Felipe Bardelli Dinardi

Dissertation presented as partial requirement for obtaining the Master’s degree in Statistics and Information Management, with a specialization in Information Analysis and Management.

Advisor / Co-Advisor: Professor Jorge Mendes, PhD

November 2019

ACKNOWLEDGMENTS

I would like to take the opportunity to thank several people who helped me during the whole process of the thesis, as well as the curricular part of the program. First of all, Professor Jorge Mendes, who was my tutor in the research methodologies class and since the very beginning, provided all the help I needed in order the develop the work, including suggestions and comments essential for the development of the dissertation. Finally, my family and friends, who offered all the support I needed throughout the path.

ABSTRACT

Forecasting stock returns forecasting is a crucially important topic in the study of finance, econometrics, and academic studies, and involves an in-depth study on time series. This thesis aims to examine the most representative companies on the São Paulo Stock Exchange, and based on that data, predict the behavior of future stock returns using several different forecasting methods.

In time series analysis, ARIMA models are used in many situations and usually present good results; nevertheless, to determine which model best suits the data, others must be tested. When considering the high volatility of the data and factoring in the economic situation of the country that is being analyzed, other techniques must be considered, especially the ARCH family ones.

Those techniques are primarily used to predict data involving Stock Markets worldwide. An accurate prediction can bring advantages for the companies who make those predictions and benefit the stakeholders directly since it provides enough information to make better decisions towards the future.

KEYWORDS

Stock Market; Forecasting; Time Series; ARIMA models; Stock Returns.

INDEX

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS AND ACRONYMS

ACF	Auto Correlation Function
AI	Artificial Intelligence
AIC	Akaike Information Criterion
AR	Autoregressive
ARCH	Autoregressive Conditionally Heteroscedastic
ARIMA	Autoregressive Integrated Moving Average
DF	Degree of Freedom
GARCH	Generalized Autoregressive Conditionally Heteroscedastic
JB	Jarque-Bera
MA	Moving Average
PACF	Partial Auto Correlation Function
RMSE	Root Mean-square Error

1. INTRODUCTION

1.1 CONTEXT

The study of time series has been developed in many areas such as physics, engineering, anthropology, among others. Nevertheless, in recent years, the attention of many researchers in the field is focused on the study of financial time series. The reasons for this phenomenon can be easily understood if one considers the global scenario for the past several years, with a massive economic crisis in big markets such as the US and Europe. In this framework, the in-depth study of economic/financial time series, especially those involving modeling and statistical analysis applied to economics, becomes paramount.

Since understanding and predicting future conditions are crucial, traditional forecasting techniques have been used to predict stock market returns. Such linear techniques, called ARIMA, usually produce good results for other fields. However, for financial time series, they must be studied carefully because, in most cases, they are not the best models to understand the past to predict the future. In the forecasting field, financial markets can be considered as a special case, due to volatility, which happens because financial markets do not follow a pattern, and usually, present large variability over time for many different reasons. Hence the main reason for the development of other methods such as ARCH models and its derivations, this family of models imputed as heteroscedastic models for modeling time series to predict the future behavior of the returns. Heteroscedasticity in statistical terms can be defined as a statistical phenomenon in which observations exhibit different variances over time, and it fits perfectly in financial time series behavior.

Although volatility is in financial market data, the linear models are also going to be tested, since some companies might present a more stable time series and the heteroscedastic models might not produce better results for those. After carefully studying the behavior of each time series and based on the results of each the models, the best will be chosen, and the predictions will be made.

1.2 BACKGROUND

Stock price forecasting is a popular and important topic in financial and academic studies (Yue & Uc, 2012), due to the number of factors that can influence the behavior of the series over time. High volatility is one of the main characteristics of this type of time series, and this phenomenon might occur due to several reasons:

- Market factors
- Economic environment
- Political scenario
- Crisis (financial mostly)
- War

Due to the factors enumerated above, forecast data involving financial markets is a tough task, and although several methods might be used for this purpose, not all of them are going to show effective results.

For data with less volatility, studies usually consider ARIMA, this approach was introduced by Box and Jenkins, and develops a systematic class of models called autoregressive integrated moving average (ARIMA) models to handle time-correlated modeling and forecasting (Shumway & Stoffer, 2011). In several fields, this method is good enough to describe phenomena and make good predictions as well. Nevertheless, for this study, other methods must be tested, for the reasons already mentioned, and also to have other approaches to compare results. Although financial markets are mostly predicted by other models, ARIMA might have some validity, and good results can come from this class of models, especially if the data does not present high volatility, which is not usual, but possible.

A problem in the analysis of financial time series is to forecast the volatility of future returns. Models such as ARCH and GARCH and stochastic volatility models have been developed to handle these problems (Shumway & Stoffer, 2011).

Both ARCH and GARCH models that were proposed by Engle (1982) and Bollerslev (1986), can capture volatility simultaneously (AL-Najjar, 2016). Therefore, they are the most recommendable for the forecasting of this particular type of time series.

The specific objectives and relevance of the study will be discussed in the next session.

1.3 OBJECTIVES AND STUDY RELEVANCE

The main goal of the study is to identify the best methods to make predictions in the stock market. In terms of the models, a salient point is that different models can be applied depending on the company business since variability directly depends on the stability of the market to which the company belongs to. Considering the general steps to develop the predictions, the diagram below describes the steps briefly:

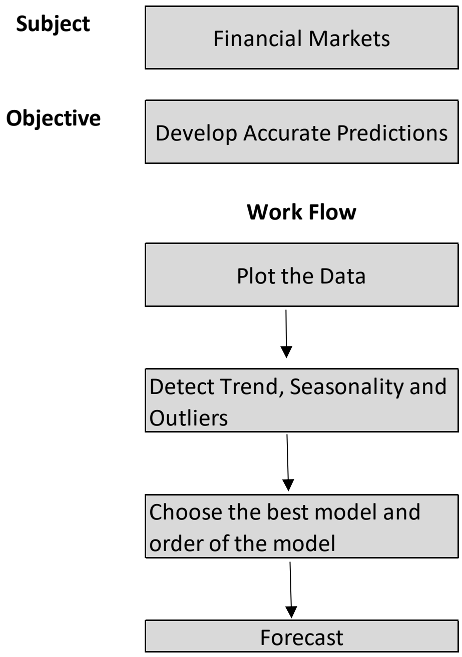

Figure 1.1 - Process

The diagram process is the general procedure to make decisions about the model. Some decisions must be taken in each step. For instance, in the first step, a plot of the time series allows us to see if the series is stationary, and it detects trend or seasonality as well. In the case of ARIMA models, the data must be detrended before developing the model. All this information can be identified in the plots, and later it will be helpful to determine the order of the model.

The main question to be answered with this thesis is: **Which one of the forecasting methods produces better results in the financial time series context?**

In order to achieve this objective, some questions, such as the time frame, need to be decided beforehand. Given the 2008 global crisis, the best choice is to skip this period, since it might have a substantial influence on the time series. In this scenario, the idea is to start with weekly data from 2010, and study the behavior of the time series until 2016, then based on that, perform the predictions for five years.

With regard to exploratory data analysis, in financial time series, it is particularly complex to determine if a certain value is considered as an outlier due to the volatility. The same applies to seasonality and trend.

After the decisions involving the time frame and those involving the exploratory analysis, the determinations concerning the model need to be made. In the case of financial markets, each series needs to be carefully analyzed, since the same model is probably not going to be suitable for all the time series under study. The first approach will be the ARIMA, but in all cases, the heteroscedastic models (ARCH) are going to be tested since they are the most recommendable for the purpose of the study. Although practical experiences show that heteroscedastic models might be more appropriate, the results will be analyzed and compared.

The table below summarizes all the points exposed so far:

Question to be defined	Answer
Time Frame	Weekly data from 2010 to 2016.
Trend, Seasonality and Outliers	Each case will be individually analyzed due to the data peculiarities.
Best Model for the data	Several models (sucg as ARIMA & ARCH) will be tested.

Table 1.1 - Summary of the Questions

The relevance of the study is associated with the quality of the predictions since contributions about what theoretical model better fits data from a financial market might help investors make better decisions; furthermore, they and might help the companies as well since companies often invest in other companies to make profits.

2. LITERATURE REVIEW

Box and Jenkins published the ARIMA models in 1976, and this was undoubtedly the start of a new framework regarding forecasting. The method is popularly known as the Box-Jenkins (BJ) methodology but technically coined the ARIMA methodology. The emphasis of these methods is not on constructing single-equation or simultaneous-equation models, but on analyzing the probabilistic or stochastic properties of economics time series on their own under the philosophy, let the data speak for themselves (Gujarati, 2003). The ARIMA models were not developed based on economic theory, which is usually based on simultaneous-equations models. Therefore the model is sometimes called an atheoretic model.

Abundant research has been undertaken in numerous disciplines or subjects that involve ARIMA methodology to forecast the future value(s) of a given variable (Afeef, Mustafa, Ihsan, Anjum, Zada, 2018). Currently, many studies use the ARIMA methodology to predict variables, and apropos of the financial framework, it is even more useful. The study proves ARIMA to be a very robust model, although way more effective in the short term, which seems to be a consensus for this type of models. From the findings of the study, it has been construed that ARIMA has a very good capacity to forecast future values in the short run. Of course, the long-term prediction using lagged values of a variable will only make little sense, however (Afeef, Mustafa, Ihsan, Anjum, Zada, 2018). In relation to the financial market framework, where predictions in many situations must be more focused on the short term, the models are highly effective and have proven validity.

Although those models are eminently effective and used in other fields besides finance, it makes a strong assumption that the future data values are linearly dependent on the current and past data values (Büyükşahin & Ertekin, 2019). The fact that it makes such assumptions might be good for some problems, but in practical terms might not help in other situations, where the assumptions are violated. In this specific case, the authors point to the fact that even though it is an outstandingly robust model, it does have limitations, and for some specific situations, other models might be considered for forecasting.

With respect to financial time series features, ARIMA models will not always be the best to make predictions. In this context, many other works in the literature use the ARCH/GARCH class of models to address issues with which ARIMA models are not capable of dealing with. Essentially, ARCH models estimate volatility as a function of past volatility shown in the time series.

The ARCH regression model has a variety of characteristics that make it attractive for econometric applications. Econometric forecasters have found that their ability to predict the future varies from one period to another. McNees suggests that “the inherent uncertainty or randomness associated with different forecast periods seems to vary widely over time.” He also documents that “large and small errors tend to cluster together (in contiguous time periods).” This analysis immediately suggests the usefulness of the ARCH family, where the underlying forecast variance may change over time and is predicted by the past forecast errors. The results presented by McNees also show some serial correlation during the episodes of large variance (Engle, n.d.). Engle´s work had a high relevance in the field of finance, making room for a new branch inside econometrics. In the study mentioned, Engle studied inflation in the United Kingdom.

This is where the so-called autoregressive conditional heteroscedasticity (ARCH) model originally developed by Engle comes in handy. As the name suggests, heteroscedasticity, or unequal variance, may have an autoregressive structure in that heteroscedasticity observed over different periods may be autocorrelated (Gujarati, 2003). Thus, a study by Engle (1982) proposed to model time-varying conditional variance through applying the Autoregressive Conditional Heteroscedasticity process (ARCH), which is expected to mainly capture the dynamic behavior of conditional variance using lagged disturbance (AL-Najjar, 2016).

In many studies, ARCH methods seem to outperform the other methods applied; nevertheless, it depends on the market and the characteristics of the time series to which the model is being applied. A study made to forecast the returns of the Dhaka Stock Exchange achieved good results with this method. The results of the in-sample statistical performance show that both the ARCH family models are selected as the best performing model jointly for DSE20 index returns, where for DSE general index return series, the ARCH model outperforms other models (Masukujjaman, 2013).

In opposition to the ARIMA models, ARCH models were developed based on classical regression assumptions.

Given the model introduced by Engel, Bollerslev introduced the Generalized ARCH model called the GARCH model. In this model, the conditional variance depends not only on the disturbance term of the previous period but also on the conditional variance of the previous period (Abounoori & Tour, 2019). GARCH (1,1) is the most widely used model. Akgiray (1989) is the first researcher who used the GARCH model to forecast volatility, and he demonstrated that GARCH produces better forecasts than most of the other forecasting methods such as Random Walk (RW), Historical Mean (HM), Moving Average (MA), and exponential smoothing (ES) when applied to monthly US stock market data (Miles & Huberman, 1994); (Chtourou, 2015).

The GARCH model family is a series of models designed to explain the regularity of fluctuations in time series, and they prove to be of really good ability when describing the volatility of financial data, adding a wide range of theoretical and practical value (Lin, 2018). In this particular paper, the model was used to predict the stock market in China. It achieved the best performance among several models. The author also points to the fact that those models are usually more effective when used to predict financial time series in developed countries, nevertheless the methodology worked perfectly, even given the features of the Chinese market and the characteristics of the time series in this case.

3. THEORETICAL FRAMEWORK

Initially, it is important to define time series, to understand the concept, and by consequence, the further application of the models, better. A time series can be defined as a collection of random variables indexed according to the order they are obtained in time. The primary objective of time series analysis is to develop mathematical models that provide plausible descriptions for sample data (Shumway & Stoffer, 2011). In the context of time series analysis, several techniques might be used to understand how the data behaved over time and to make predictions. In this work, a few models will be tested and studied in further detail, particularly the ARIMA and ARCH/GARCH models, with a special emphasis on the ARIMA models.

When dealing with time series analysis, some important steps must be followed. One of the first and most critical aspects, predominantly, is to plot the data on a graph at the outset. In this way, some features regarding the time series can be observed, and the researcher can understand more about the kind of data under deliberation. From the plot of the data, the existence of some kind of trend in the data can be detected, which means that the data must be observed, and downward or upward movements over time must be discerned. Another important aspect is the presence of the seasonality in the data, which means some repeating patterns over time. This facet is particularly important for a variety of forecasting methods since some of them are not able to deal with seasonality, and it must be removed before the forecasting procedure. Outliers are other issues that must be analyzed in a time series. The occurrence of outliers is particularly important in regression models and can even alter the efficiency of the model. The variance must also be analyzed to understand if the data varies over time, or if the variance is constant, which might sometimes be hard to identify when only considering the plot of the data.

The normality of the data must be tested before modeling, since this is not necessarily possible by just looking at the histogram, another way of testing, is to apply the Jarque-Bera test, which is a normality test. This test first computes the skewness and kurtosis measures of the OLS residuals and uses the following test statistic:

$$JB = n\,\frac{S^2}{6} + \frac{(K-3)^2}{24}$$

Where *n* is equal to the sample size, *S* the skewness coefficient, and *K* the kurtosis coefficient. For a normally distributed variable, S = 0 and K = 3. Therefore, the JB test of normality is a test if the joint hypothesis that S and K are 0 and 3, respectively. In that case, the value of the JB statistic is expected to be 0. Under the null hypothesis that residuals again are normally distributed, Jarque and Bera showed the JB test asymptotically follows the chi-square distribution with two df in the given equation. Applications of the test will be seen in the data analysis section.

3.1. ACF AND PACF

Another important concept before defining the models themselves is to make a definition of the Autocorrelation Function (ACF) and the Partial Autocorrelation Function (PACF).

The ACF is defined as:

$$\rho(s,t) = \frac{\gamma(s,t)}{\sqrt{\gamma(s,s)\gamma(t,t)}}$$

The ACF measures the linear predictability of the series at time t, say x_t, using only the value x_s (Shumway & Stoffer, 2011).

The partial autocorrelation function gives the partial correlation of a stationary time series with its own lagged values, regressed the values of the time series at all shorter legs. It contrasts with the autocorrelation function, which does not control for other lags.

In general, a partial correlation is a conditional correlation. It is a correlation between two variables under the assumption that we know and consider the values of some other set of variables if considering a regression where, x_1, x_2, and x_3 are predictor variables. In this scenario, the partial correlation between y and x_3 is the correlation between variables determined considering how both are related to x_1 and x_2.

It can be mathematically denoted as:

$$\frac{Covariance\ (y, x_3|x_1,x_2)}{\sqrt{Variance(y|x_1,x_2)Variance(x_3|x_1,x_2)}}$$

To determine the order of the models being considered, the ACF and the PACF must be analyzed together.

3.2. STATIONARITY

Stationarity is another concept that is crucial while modeling data. A stationary time series must have a constant mean and variance over time. Broadly speaking, a stochastic process is said to be stationary if its mean and variance are constant over time and the value of the covariance between the two periods only depends on the lag between the two time periods and not the actual time at which the covariance is computed (Gujarati, 2003). Broadly speaking, a time series is said to be stationary if there is no systematic change in mean (no trend) over time, if there is no systematic change in variance and if period variations have been removed (Aljandali & Tatahi, 2018).

One important point to be noted is that before executing any time series analysis the time series data is supposed to be stationary. There are different test for stationary diagnosis, like Augmented Dickey Fuller test. If data is stationary then we can proceed, on the other hand if it is not, then we difference it and make stationary. ARMA model of the differenced series is called ARIMA model (Prasad & Choubey, 2018).

In order to determine stationarity in a time series, a few procedures can be undertaken. One of them is a test denominated as the Augmented Dickey-Fuller ADF test. The test consists in estimating the following equation:

$$\Delta Y_t = \delta Y_{t-1} + \mu_t$$

The number of lagged difference terms to include is often determined empirically. The idea is to include enough terms so that the error term in the equation above is serially uncorrelated. Details in the application and the assumption for the test can be seen when they are applied in the time series.

3.3. ARIMA MODELS

In the context of this research, the ARIMA models will be the ones most analyzed. It is, therefore, imperative to understand them well. These models, also known as the Box-Jenkins methodology, encompass autoregressive and integrate with moving average methods. Thus, the origin of the term ARIMA (Autoregressive Integrated Moving Average).

The models were introduced in 1976 and are based on the estimation of the dependent variable y as a function of the lag of the y variable itself, indicated by *p* autoregressive terms, *q* terms representing the past errors. As most economic/financial time series are not stationary, the application of the ARIMA models (p,d,q) demands the transformation of the variables. The ACF and the PACF must be analyzed in conjunction to determine the orders of the model, where the goal is to minimize the AIC, which is a method of measuring the quality of the model.

We must first introduce the AR (Autoregressive) and MA (Moving Average) terms of the equation to understand the models better.

3.3.1. AR Models

Autoregressive models are based on the idea that the current value of the series, X_t, can be explained as a function of the past values. This can be defined by the equation:

$$X_t = \Phi_1 X_{t-1} + \Phi_2 X_{t-2} + \cdots + \Phi_p X_{t-p} + \varepsilon_t$$

Where X_t is the observation in time t, Φ_p is the parameter of the model AR, p is the order of the model, and ε represents the error that cannot be explained by the model.

3.3.2. MA Models

As an alternative to the autoregressive representation in which the X_t on the left-hand side of the equation is assumed to be combined linearly, the moving average model of order *q,* abbreviated as MA(*q*), assumes that white noise ω_t on the right-hand side of the defining equation is combined linearly to form the observed data (Shumway & Stoffer, 2011). This can be represented as:

$$X_t = \omega_t + \theta_1 \omega_{t-1} + \theta_2 \omega_{t-2} + \cdots + \theta_q \omega_{t-q}$$

Where there are *q* lags in the MA, and θ_1, θ_2, … , θ_q are the parameters.

3.3.3. ARMA Models

Together, the two previous models form the ARMA model. They can be mathematically be represented as:

$$X_t = \Phi_1 X_{t-1} + \Phi_p X_{t-p} + \alpha_t - \theta_1 \alpha_{t-1} - \cdots - \theta_q \alpha_{t-q}$$

It is important to mention that these models are not suitable to model non-linear relationships. Nevertheless, they are very important to the understanding of stationary time series and autoregressive models.

The model already discussed is the integration of the AR and MA models proposed by Box and Jenkins.

The table below presents the order of the events when modeling using ARIMA models:

STEP	ACTION
1	Examine the series for stationary. Analyze ACP and PACF and perform the unit-root test
2	If it is not stationary, difference it one or more times to make it stationary
3	Identify if the series is AR, MA or ARIMA
4	Estimate the model
5	Analyze the residuals
6	Forecast using the model

Table 3.1 - Course of Action for ARIMA models

With regard to the points in the table, it is salient to mention that the ACF and the PACF must be analyzed to estimate the orders of the model. The table below shows the theoretical patterns for ACF and PACF:

THEORETICAL PATTERNS OF ACF AND PACF		
Type of Model	**Typical pattern of ACF**	**Typical pattern of the PACF**
AR (*p*)	Decays exponentially or with damped sine wave pattern or both	Significant spikes through lags *p*
MA (*q*)	Significant spikes through lags *q*	Declines exponentially
ARMA (*p*,*q*)	Exponential decay	Exponential decay

Table 3.2 - Theoretical Patterns of ACF and PACF

Another important detail involves the residuals of the model. The residuals from the tentative model are examined to find out if they are white noise. If they are, the tentative model is probably a good approximation to the underlying stochastic process. If they are not, the process is started all over again. Therefore, the Box-Jenkins method is iterative (Gujarati, 2003).

As mentioned previously, one of the methods to evaluate the model is the Akaike Information, which can be written mathematically as:

$$\ln AIC = \left(\frac{2K}{N}\right) + \ln\left(\frac{RSS}{n}\right)$$

Where *k* is the number of regressors, *n* is the number of observations and RSS the Residual Sum of Squares, which is a measure of the discrepancy between the data and the estimation model.

The intention when comparing two or more models is to consider the model with the lowest AIC. The model penalizes the error in the variance by a term proportional to the number of parameters.

Finally, analyzing the residual requires a specific test, which is called the Ljung-box test. In the context of the model, the residuals must be independent of each other. The p-value of the test will show whether the residuals are independent or not. The null hypothesis for this test is that the residuals are not correlated.

The Ljung-Box test statistic can be defined as:

$$Q' = T(T+2)\sum_{h=1}^{N} \frac{\rho^{\wedge 2}(h)}{T-h} \sim \chi_N^2$$

Where *T* is the sample size, $\rho^{\wedge 2}$ is the sample correlation at lag *h* and *N* is the number of lags being tested.

Considering that the out-of-sample forecast will be performed, other measures might be helpful to evaluate the predictive power of the model. In this scenario worth to present the Root Mean-square error (RMSE).

The RMSE is a frequently used measure of the difference between the observed and the predicted values, and it is highly efficient to measure the accuracy of the model. Although the threshold might be difficult to define, it is a measure that researchers must consider.

Can be defined as:

$$RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Observed - Predicted)^2}{N}}$$

Where *N* is the number of observations and predicted and actual are the values of the model.

3.4. ARCH/GARCH MODELS

The ARCH family modes emerged in the context of high volatility. It is normally used in the context of financial time series due to its characteristics, such as high variance and volatility. ARCH, in this case, stands for autoregressive conditional heteroscedasticity. As the name suggests, heteroscedasticity, or unequal variance, may have an autoregressive structure in that heteroscedasticity observed over different periods may be uncorrelated (Gujarati, 2003). In essence, the ARCH model is a model for volatility in a time series.

In order to make it possible to use the ARCH model, a test must be performed to make sure that the time series present the necessary features that justify the use of the heteroscedastic models. If the time series has mean 0, the ARCH model can be written as:

$$y_t = \sigma_t \varepsilon_t$$

Where σ_t:

$$\sigma_t = \sqrt{\alpha_0 + \alpha_1 y_{t-1}^2}$$

The ARCH(1,0) model for the variance of the model y_t is that conditional on y_{t-1}, the variance at the time *t* is:

$$Var(y_t | y_{t-1}) = \sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2$$

The analysis of the ACF and the PACF must be done to understand the orders of the model, and the model can be tested after this step.

ARCH models have had several variations since they were first developed. These variants include the GARCH models, which are the generalized autoregressive conditional heteroscedasticity models. The most popular version of the model is GARCH (1,1), and it can be mathematically written as:

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \beta_1 \sigma_{t-1}^2$$

GARCH models were only presented as variations of the ARCH. In the framework of this study, those models will not be used.

4. METHODOLOGY

As previously described, the main objective of the study is to undertake an in-depth probe into time series analysis, to understand the past, and predict the future. In works of this nature, the following steps are taken to achieve better results, in general:

- Problem Definition
- Information Gathering
- Exploratory Data Analysis
- Choice of the Best Model
- Evaluation of the results

Problem Definition

The problem definition is the first point, and it is essential to define all the steps of the study. The question for this study is relevant to the stock market and financial fields. It is also a conundrum that other researchers are trying to understand better. Furthermore, the problem definition also helps to delimitate the questions dealt with in the research, which should not exceed the limits proposed in the first place.

Information Gathering

Due to legal issues and transparency, all data involving financial markets have to be published. In this scenario, all information needed to develop this study can be found on the São Paulo Stock Exchange website, and in many other internet tools devoted to studying this field.

Exploratory Data Analysis

A solid exploratory data analysis must be done to make the right choices involving the model. In the framework of this study, some steps must be considered and followed, as detailed below:

- Plot the original series (returns)
- Perform histogram to try to detect some patterns in the data
- Perform ACF and PACF
- Perform all the necessary tests to make sure the modeling will present the best results

Regarding to the first step, it is necessary to plot the initial series to try to distinguish some important characteristics about the time series, such as trend, seasonality, and outliers. In the event of the identification of one of these aspects, we must treat them properly; otherwise, ARIMA models are not going to fit the data well. Other vital information we need to discern based on this first plot is the stationarity of the series. The latter being a requirement to develop some forecasting models.

In the case of the histogram, it is important to see how data is distributed and to assist in detecting the outliers in the data more easily.

The ACF and the ACF will allow us to understand if the data is stationary, and also give us some insights about the order of the models. This condition works for both homoscedastic and heteroscedastic models.

Choice of the best model

After performing the exploratory analysis, different models must be tested. In the context of this study, the ARIMA and ARCH family models will be tested. The best model for prediction will be chosen based on the conclusions of this scrutiny.

The criteria will be established based on the behavior of the time series since the presence of high volatility makes it indispensable to determine which model has more predictive power.

Evaluation of the results

The evaluation will be made considering the output for each model. ARIMA and heteroscedastic models must be tested when considering a financial time series. The efficiency of each type of model will be determined by the features of the time series, which reflect the way the market behaved in the period studied. In this scenario, both models are relevant for the study, and the results might lead to several implications that will help gain more meaningful insights into the phenomena under study.

5. RESULTS AND DISCUSSION

5.1 Data Analysis

The results of all the techniques used to forecast the data chosen for this work will be presented in this chapter. First of all, it is important to present a detailed exploratory data analysis. Later on, the results of the modeling using ARIMA and ARCH models will be presented, as well as the conclusion about which model better fits the data, and consequently, which one of them should be used to predict the assets with more accuracy.

The theoretical framework of all methods was presented in the methodology section, as well as in the chapter involving the whole conceptual background of all the techniques. The process will be conducted in the exact same way for all the companies.

5.2 Exploratory Data Analysis

In terms of the exploratory data analysis, the companies will be evaluated separately. Although the same process was used for each of the companies, not all graphics will be chosen, since the most important element is to understand the features of the data. The figure below shows the statistics for the company Ambev, and based on the graphic and also taking the statistics into consideration, the conclusion is that the dispersion of the data is not considered high since the standard deviation does not present a high value. The data possibly have some outliers, which can be seen easily, especially in the right tail of the distribution. Regarding the possibility of having a normal distribution, the conclusion is that the distribution is not normal; in order to reach this conclusion, the results of the Jarque-Bera test and also the probability must be evaluated. High values in the Jarque-Bera test combined with a low value of probability (in this scenario, the probability is to have a normal distribution), makes this conclusion very clear.

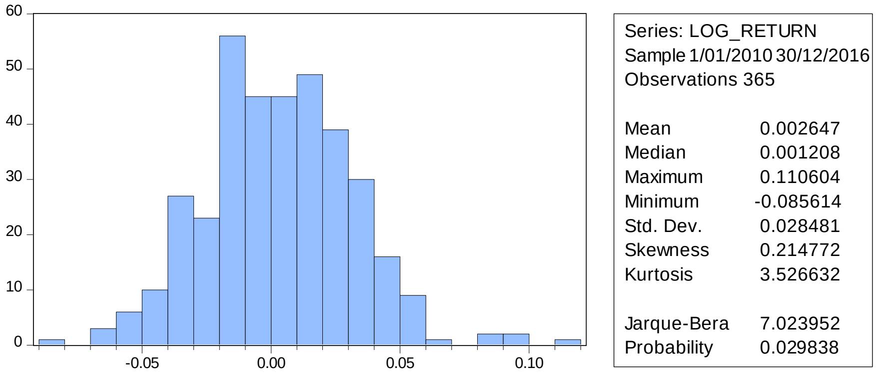

Figure 5.1 - Ambev Histogram

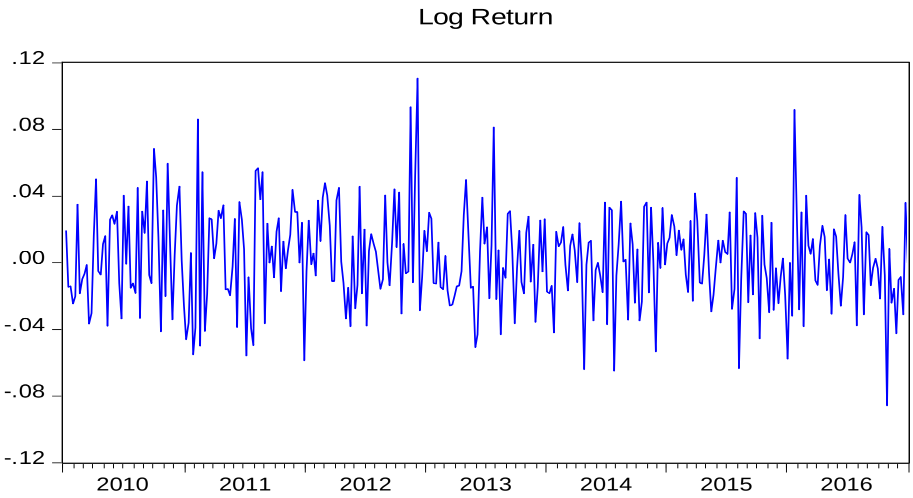

Figure 5.2 - Ambev Log Return

With respect to the log return plot, the conclusion is that the data has no trend and no seasonality. The augmented Dickey-Fuller test should be performed to determine stationarity. In this test, the null hypothesis is that there is a unit root, meaning that the series is not stationary. The alternative hypothesis is that there is no unit root, and in this case the series is stationary.

Augumented Dickey Fuller test

	t-statistic	Prob
Test Statistic	-19.66589	0.0000
Test Critical Values		
1%	-3.448111	
5%	-2.869263	
10%	-2.570952	

Figure 5.3 - Ambev Unit - Root test

Based on the outputs, we can conclude that the series is stationary since the value of the t-statistic is lower than the critical values. In this case, we reject the null hypothesis of having a unit root.

The table below shows the summary of the results for the other companies analyzed in the study:

	Companies			
	Petrobras	Itau S/A	Itau/Unibanco	Vale
Features				
Dispersion	Small Standard deviation	Small Standard deviation	Small Standard deviation	Small Standard deviation
Normality	Data is not normally distributed	Data is not normally distributed	Data is not normally distributed	Data is not normally distributed
Trend/Seasonality	No trend or seasonality	No trend or seasonality	No trend or seasonality	No trend or seasonality
ADF Test (Stationarity)	Stationary series	Stationary series	Stationary series	Stationary series

Table 5.1 - Summary of the results

5.3 ARIMA MODELS

One approach, advocated in the landmark work of Box and Jenkins, develops a systematic class of models called autoregressive integrated moving average (ARIMA) models to handle time-correlated modeling and forecasting (Shumway & Stoffer, 2011); considering these models, one of the first steps is to plot the time series and try to identify outliers, trend and seasonality. The returns were calculated in the original database in log return format to anticipate problems with stationarity, due to the high volatility of the data. The adoption of this method is an attempt to stabilize the variance and will not necessarily make the series stationary. Although the stationarity was already confirmed in each one of the cases by the augmented Dickey-Fuller test.

In pertinence to the figures, all the time series are stationary, which allows developing the ARIMA models. The models will be developed considering the optimization of the Akaike Information Criteria; in other words, the models of the lowest AIC will be chosen to make the forecasting.

The first step, proposed by Box & Jenkins, is based on identification procedures. The process involves choosing the right orders of the ARIMA model (p,d,q), and this process is made considering the ACF and the PACF for each case. This analysis can help to define the orders of the model.

With regard to the ACF and PACF for the company Ambev:

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.0...	-0.0...	0.390...	0.53...
		2	-0.0...	-0.0...	0.708...	0.70...
		3	0.01...	0.01...	0.811...	0.84...
		4	-0.0...	-0.0...	3.004...	0.55...
		5	0.05...	0.05...	4.089...	0.53...
		6	0.03...	0.03...	4.682...	0.58...
		7	0.02...	0.02...	4.854...	0.67...
		8	-0.0...	-0.0...	4.971...	0.76...
		9	0.06...	0.07...	6.508...	0.68...
		10	-0.0...	-0.0...	6.542...	0.76...

Figure 5.4 - Ambev ACF and PACF

Defining the order of the models based only on the plots of the ACF and PACF might be tricky, that is why several models must be tested and compared, and based on the outputs of this analysis, the best model will be chosen. The table below shows a comparison between several models for the company under study in this step:

Model Selection Criteria Table
Dependent Variable: LOG_RETURN
Date: 22/11/17 Time: 17:13
Sample: 1/01/2010 30/12/2016
Included observations: 365

Model	LogL	AIC	BIC	HQ
(2,3)(0,0)	785.389062	-4.265146	-4.190353	-4.235422
(2,2)(0,0)	783.436900	-4.259928	-4.195820	-4.234451
(4,3)(0,0)	787.034241	-4.263201	-4.167039	-4.224985
(4,2)(0,0)	784.323837	-4.253829	-4.168352	-4.219859
(4,1)(0,0)	783.598447	-4.255334	-4.180541	-4.225610
(4,0)(0,0)	782.936103	-4.257184	-4.193076	-4.231707
(3,4)(0,0)	786.972686	-4.262864	-4.166702	-4.224648
(3,3)(0,0)	785.389207	-4.259667	-4.174190	-4.225697
(3,2)(0,0)	785.388923	-4.265145	-4.190352	-4.235421
(3,1)(0,0)	783.401738	-4.259736	-4.195628	-4.234258
(3,0)(0,0)	781.852206	-4.256724	-4.203301	-4.235493
(2,4)(0,0)	784.294407	-4.253668	-4.168191	-4.219698

Figure 5.5 - Ambev Model Comparison

Based on the AIC, the ARMA (2,3) model is the best one, since the AIC presents the lowest values for this model. Considering the ARMA (2,3) model:

Dependent Variable: LOG_RETURN
Method: ARMA Maximum Likelihood (BFGS)
Date: 22/11/17 Time: 16:10
Sample: 8/01/2010 30/12/2016
Included observations: 365
Convergence achieved after 102 iterations
Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.002650	0.001440	1.840332	0.0665
AR(1)	-1.935385	0.021360	-90.60847	0.0000
AR(2)	-0.941223	0.021937	-42.90485	0.0000
MA(1)	1.931799	0.904016	2.136909	0.0333
MA(2)	0.874134	0.782056	1.117738	0.2644
MA(3)	-0.063092	0.151017	-0.417780	0.6764
SIGMASQ	0.000784	0.000899	0.872011	0.3838

R-squared	0.031181	Mean dependent var	0.002647
Adjusted R-squared	0.014944	S.D. dependent var	0.028481
S.E. of regression	0.028267	Akaike info criterion	-4.265146
Sum squared resid	0.286055	Schwarz criterion	-4.190353
Log likelihood	785.3891	Hannan-Quinn criter.	-4.235422
F-statistic	1.920360	Durbin-Watson stat	1.998939
Prob(F-statistic)	0.076673		

Inverted AR Roots	-.97-.07i	-.97+.07i	
Inverted MA Roots	.06	-1.00-.07i	-1.00+.07i

Figure 5.6 - Ambev ARMA Model

Although it is considered the best model based on the AIC, after further analysis, we can see that not all coefficients are statistically different from 0, and this conclusion can be reached by calculating the ratio between each coefficient and the standard error.

In terms of the residuals, the table below shows that they are white noise, which is a positive point in the model since it shows that there is no information in the residuals that might be helpful for the model:

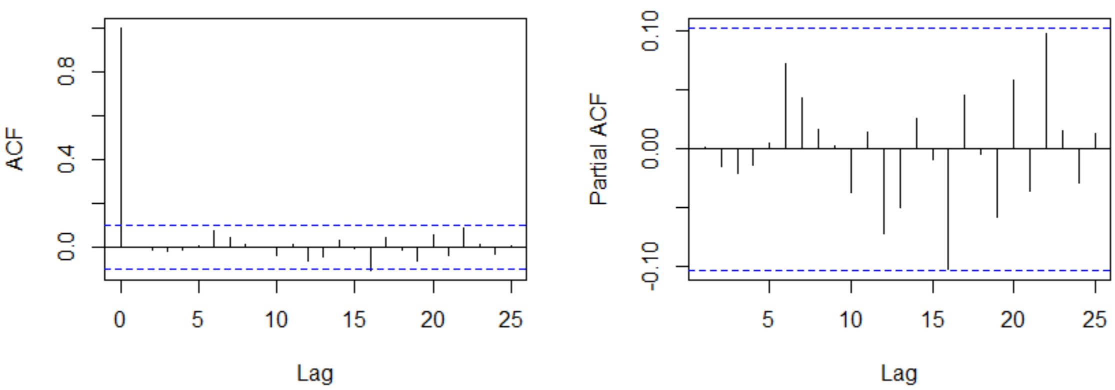

Figure 5.7 – Residuals Ambev ACF and PACF

Another important factor to be considered is the independence of the residuals, which is measured by the Ljung-Box test, and taking into account the high p-value for this test, the conclusion is that the residuals are independent of each other:

Ljung-Box Test	
p-value	0,5841

Table 5.2 - Ambev Ljung-Box test

Regarding the forecast for this model:

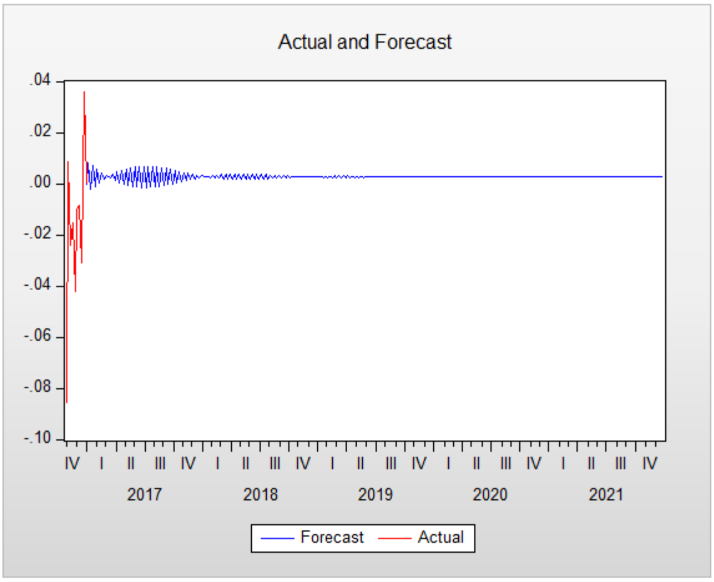

Figure 5.8 - Ambev ARMA Forecast

As expected, for the stationary model, the forecast will eventually converge to the mean and stay at this level for the remaining periods. In the plot above, before the forecast converges to the mean, some predictions are made.

The table below shows the summarized results for the other companies analyzed:

	Companies			
	Petrobras	Itau S/A	Itau/Unibanco	Vale
Features				
Order Best Model	ARMA (3,3)	ARMA (3,3)	ARMA (4,4)	ARMA (3,3)
AIC	-2,742414	-3,591795	-3,526002	-2,537487
Residuals	White noise	White noise	White noise	White noise
Ljung-Box Test	Independent Residuals	Independent Residuals	Independent Residuals	Independent Residuals
Forecasting	Effective in the short term, but eventually converges to the mean and stay at this point for all periods	Effective in the short term, but eventually converges to the mean and stay at this point for all periods	Effective in the short term, but eventually converges to the mean and stay at this point for all periods	Effective in the short term, but eventually converges to the mean and stay at this point for all periods

Table 5.3 - Summary of the results

5.3.1 Out-of-Sample Forecast

Considering the results presented so far, and to attest that the models described previously are the best to make predictions, the out-of-sample forecast will be performed. Comparing the fit out-of-sample allows us to evaluate features of model performance that are important for practical applications.

In the previous section, important conclusion about the model were reached, such as the ability of the ARMA models to make predictions in the short-term. Hence, the out-of-sample forecasting should be done considering the short term, to produce better results.

The period used to make the predictions involves the weekly data of the year of 2017. Regarding the company Ambev:

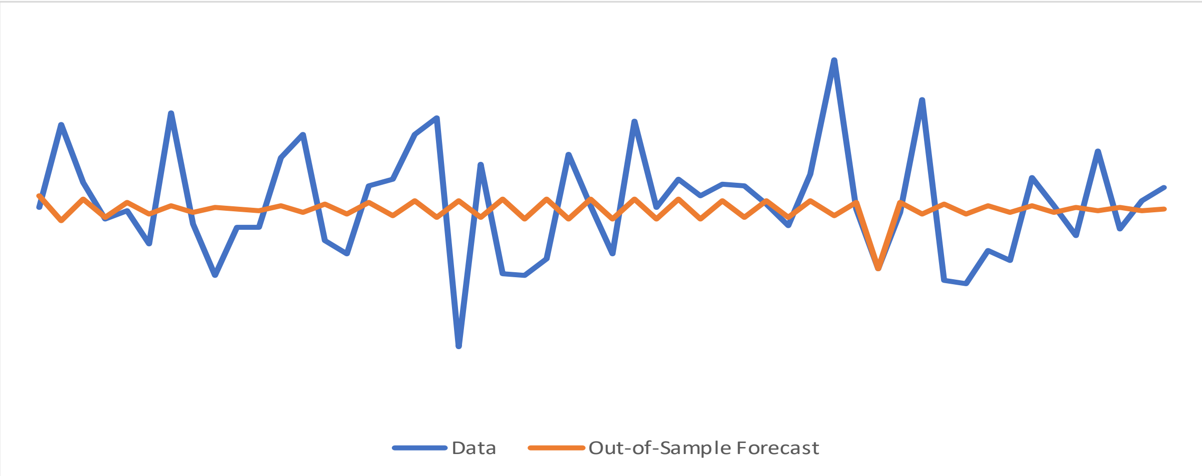

Figure 5.9 - Out-of-sample forecast Ambev

As the graph shows, although the data tends to stay around the mean, it has the ability of making predictions in the short-term. The modelling was made considering the best model for the company, which is the ARMA (2,3) based on the AIC. For the same model, the RMSE was calculated, as follows:

Company	Model	RMSE
Ambev	ARMA (2,3)	4,16745

Table 5.4 – RMSE Ambev model

The same procedure was adopted for all the other companies:

	Companies			
	Petrobras	Itau S/A	Itau/Unibanco	Vale
Features				
Model	ARMA (3,3)	ARMA (3,3)	ARMA (4,4)	ARMA (3,3)
RMSE	9,340332	5,595034	5,497613	10,21861
Forecasting	Make predictions in the short term but with time converges to the mean	Make predictions in the short term but with time converges to the mean	Make predictions in the short term but with time converges to the mean	Make predictions in the short term but with time converges to the mean

Table 5.5 - Summary of the results Out-of-sample forecast

In the table are summarized the values founded by company. Other models were compared and those of the table are the best based on the RMSE.

5.4 ARCH Models

The ARCH (Autoregressive Conditionally Heteroscedastic), is a model for a variance of a time series. The methodology was developed, especially to deal with the high volatility in problems involving econometrics and finance.

The goal of this work is to understand which model produces better results considering financial time series in the Brazilian stock market (São Paulo Stock Exchange), and although ARCH models are considered particularly attractive when dealing with financial time series, they might not be the best ones in every case.

The ARCH/GARCH models will be performed for each of the companies to compare the results, and after the full analysis, the results will be compared, considering the results of the ARIMA modeling as well.

As the first step, the model should be run. By doing this procedure, we will be able to analyze the residuals and identify if the ARCH family models might be suitable for the analysis. We need to confirm the presence of heteroscedasticity in the data series before processing to modeling, and this can be achieved by doing the ARCH effect test. Financial assets price of stock index series or another high-frequency data will often appear in the feature that a large fluctuation usually followed by another large fluctuation, and a small fluctuation usually followed by another even smaller fluctuation, this is called the ARCH effect (Lin, 2018).

Appertaining to the outputs of the analysis for the company Ambev:

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.002647	0.001491	1.775585	0.0766

R-squared	0.000000	Mean dependent var	0.002647
Adjusted R-squared	0.000000	S.D. dependent var	0.028481
S.E. of regression	0.028481	Akaike info criterion	-4.276435
Sum squared resid	0.295261	Schwarz criterion	-4.265751
Log likelihood	781.4495	Hannan-Quinn criter.	-4.272189
Durbin-Watson stat	2.064213		

Figure 5.10 - Ambev Regression Outputs

ARCH models are specifically designed to model and forecast conditional variances. The variance of the dependent variable is modeled as a function of past values of the dependent variable, in this case the dependent variable are the returns considered in the study.

The next step is to look at the residuals:

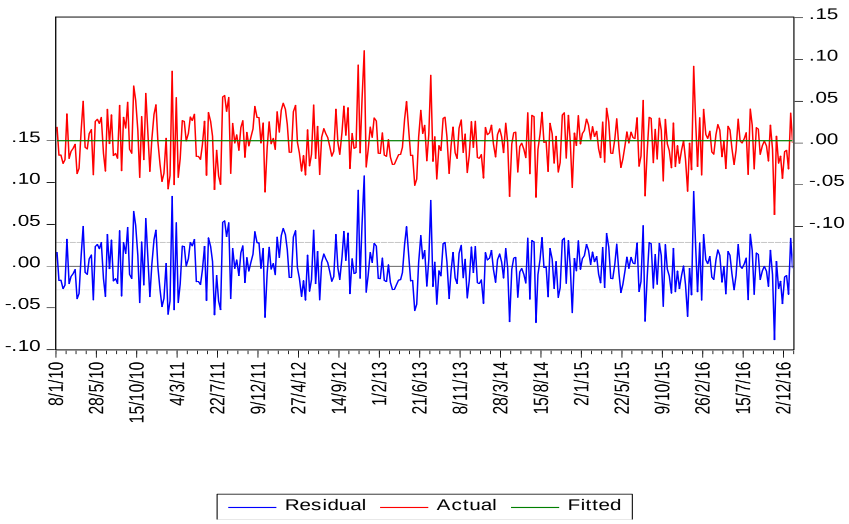

Figure 5.11 - Ambev Residuals

Based on the residuals, the conclusion is that the volatility is constant, and there are no periods without volatility. The ARCH test must be performed to validate the conclusions:

Heteroskedasticity Test: ARCH

F-statistic	1.273428	Prob. F(1,362)	0.2599
Obs*R-squared	1.275975	Prob. Chi-Square(1)	0.2586

Figure 5.12 - Ambev ARCH Test

With regard to the test, the null hypothesis is that there is no ARCH effect, and the alternative hypothesis is that there is an ARCH effect. In this case, the null hypothesis cannot be rejected (since the p-value is above 5%), so there is no ARCH effect. If performing the ARCH modeling and forecasting, in this case, the results will be poor:

Variable	Coefficient	Std. Error	z-Statistic	Prob.
C	0.002494	0.001479	1.685569	0.0919
Variance Equation				
C	0.000737	6.70E-05	10.99135	0.0000
RESID(-1)^2	0.091523	0.060293	1.517977	0.1290

R-squared	-0.000029	Mean dependent var	0.002647
Adjusted R-squared	-0.000029	S.D. dependent var	0.028481
S.E. of regression	0.028481	Akaike info criterion	-4.271231
Sum squared resid	0.295270	Schwarz criterion	-4.239177
Log likelihood	782.4996	Hannan-Quinn criter.	-4.258492
Durbin-Watson stat	2.064153		

Figure 5.13 - Ambev ARCH Outputs

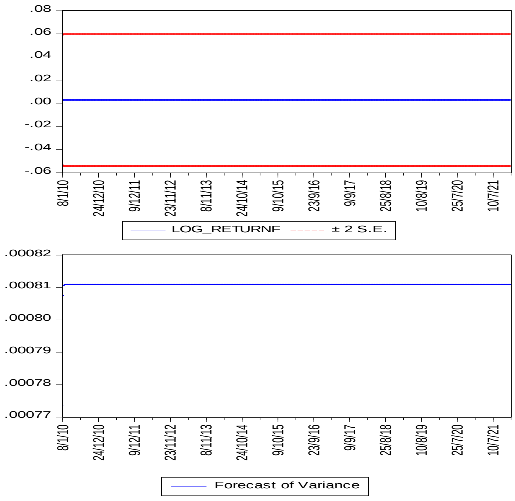

Figure 5.14 - Ambev ARCH Forecast

The results above were provided by the software considering an ARCH (1,0) model.

The table below shows the summary of the results for the other companies:

	Companies			
	Petrobras	Itau S/A	Itau/Unibanco	Vale
Test				
ARCH Effect	No	No	No	No

Table 5.6 - Summary of the results

Since the time series regarding the companies do not present the ARCH effect, this technique proved to be not effective in the case of the Brazilian market for the period under study.

5.5 Discussion

Dealing with stock market data is always challenging in the sense that several factors might influence how the market behaves. The stock market is constantly changing with uncertainties. Rapid dissemination of information and fast capital flow will lead to fluctuations in the stock price, and the undulating price and the undulating price will affect the market in return (Lin, 2018). Considering similar studies in other markets, the ARCH models are prevalent and in general, are used more often than other methods, primarily due to the characteristics of the data such as high volatility.

The literature supports the application of ARIMA and ARCH models to analyze and make predictions based on time series data. Specific characteristics of the data and even of the market being analyzed gives us some clue about which model might be more suitable. For financial markets in general, due to the characteristics of the time series, heteroscedastic models are very popular.

The conclusions about the results will be presented next, but some discussion concerning the findings must be made. The first factor is that even though heteroscedastic models perform well when trying to predict financial markets, in the specific market applied in this research, they were not as effective as they usually are, at least taking into consideration the literature on the topic.

Another important point is that even the ARIMA models efficacious for short-term predictions are not effective in the long-term. This aspect is another characteristic that the theory supports since these models tend to converge to the mean of the time series in the long-term and remain at that point for future periods.

In summary, the research reached the goal of understanding which methods perform better in respect of the Brazilian scenario for the time frame studied. The results confirmed what the theory states. Heteroscedastic models are not necessarily the best as they usually are in other markets. Besides, even though ARMA models do not provide the best predictions, are the ones under study who can provide better results, especially in the short-term.

6. CONCLUSIONS

The conclusion is that in the framework of the Brazilian stock market, ARIMA models perform better. Even given the problems with long-term forecasting, modeling the data with the ARIMA approach produces better results. The best way to use this approach granted the limitations would be to predict short-term with ARIMA and evaluate the quality of the predictions in each period.

The other approach that was considered to model the data was the ARCH approach. Nevertheless, the data does not present the features that are necessary to frame the analysis using this methodology. As the data analysis has shown, one of the key aspects to determine whether the ARCH model family might be suitable or not, namely, the ARCH effect is not present in any of the time series studied. Thus, in this case, this methodology will not produce good results considering the forecasting. Other time frames can be studied in the attempt to identify if the ARCH effect is present, and in this case, this mechanism can be used to produce forecasting for stock returns.

During the period that the study focused on, the Brazilian market behaved very well, without any degree of oscillations. This fact can explain why simpler models such as ARIMA were more effective in the attempt to forecast stock returns. Usually, concerning other markets, ARCH models are more effective during the high oscillations in the market due to crises and several other events that might influence this process. In the framework of the Brazilian stock market, higher fluctuations during the period would change the results entirely.

7. LIMITATIONS AND RECOMMENDATIONS FOR FUTURE WORKS

The study was conducted using the most popular techniques to make predictions in financial markets, specifically the ARIMA and ARCH/GARCH models. Nonetheless, several other techniques are also being used for this purpose nowadays. Even though financial time series is a distinctive case due to its features such as high volatility in some periods, or even constant depending on the market, other techniques might produce good results.

Recently, data mining techniques and artificial intelligence techniques like decision trees, rough set approach, and artificial neural networks have been applied to this area (Al-Radaideh, Assaf, & Alnagi, 2013), and presenting good results. A further study involving those time series might be to test a few AI techniques such as Neural Networks and compare them with traditional ones, to establish some comparisons on the outputs. In summation, more techniques can be analyzed and tested to produce even better results or to prove the traditional techniques as being the best to use in the financial field.

Regardless of the chosen technique, it is important to mention that during the period studied the financial time series presented a regular behavior, which means that the results would be different if during the period the market was facing instability due to crisis or any other event. Hence, the conclusions of the study could have been different in this scenario.

8. BIBLIOGRAPHY

AL-Najjar, D. M. (2016). Modelling and Estimation of Volatility Using ARCH/GARCH Models in Jordan's Stock Market. *Asian Journal of Finance & Accounting*, *8*(1), 152. https://doi.org/10.5296/ajfa.v8i1.9129

Benth, F. E. (2008). *Stochastic Modellling of electricity markets*.

Da, N., & Lewis, C. (2012). Applied Statistical Methods for Practitioners 2nd Edition. Retrieved from www.pefc.org

De Gooijer, J. G., & Hyndman, R. J. (2006). 25 Years of Time Series Forecasting. *International Journal of Forecasting*, *22*(3), 443–473. https://doi.org/10.1016/j.ijforecast.2006.01.001

Diebold, F. X. (2001). *Elements of Forecasting*. *The American Statistician* (Vol. 55). https://doi.org/10.1198/tas.2001.s126

Diebold, F. X. (2015). Forecasting in Economics, Business, Finance and Beyond.

Granger, C. W. J. (1992). Forecasting stock market prices: Lessons for forecasters. *International Journal of Forecasting*, *8*(1), 3–13. https://doi.org/10.1016/0169-2070(92)90003-R

Kholodnyi, V. A. (n.d.). *Quantitative Energy Finance*.

Abounoori, E., & Tour, M. (2019). Stock market interactions among Iran, USA, Turkey, and UAE. *Physica A: Statistical Mechanics and Its Applications*, *524*, 297–305. https://doi.org/10.1016/j.physa.2019.04.232

Afeef, Mustafa, Ihsan, Anjum, Zada, H. (2018). Forecasting Stock Prices through Univariate ARIMA Modeling, *13*, 130–143.

l-Radaideh, Q., Assaf, A., & Alnagi, E. (2013). Predicting stock prices using data mining techniques. *Acit2K.Org*, (April), 1–8. Retrieved from http://www.acit2k.org/ACIT/2013Proceedings/163.pdf

Aljandali, A., & Tatahi, M. (2018). *Economic and Financial Modelling with EViews: A Guide for Students and Professionals*. https://doi.org/https://doi.org/10.1007/978-3-319-92985-9

Büyükşahin, Ü. Ç., & Ertekin, Ş. (2019). Improving forecasting accuracy of time series data using a new ARIMA-ANN hybrid method and empirical mode decomposition. *Neurocomputing*, *361*, 151–163. https://doi.org/10.1016/j.neucom.2019.05.099

Chtourou, H. (2015). Modeling and forecasting the Dow Jones stock index with the EGARCH model. *International Journal of Economic Practices and Theories*, *5*(1), 51–61.

Engle, R. F. (n.d.). Autoregressive Conditional Heteroscedasticity with Estimate of the Variance of United Kingdom Inflation.pdf.

Gujarati, D. N. (2003). *Essetials of Econometrics*.

Lin, Z. (2018). Modelling and forecasting the stock market volatility of SSE Composite Index using GARCH models. *Future Generation Computer Systems*, *79*, 960–972. https://doi.org/10.1016/j.future.2017.08.033

Masukujjaman, M. (2013). Forecasting Volatility of Stock Indices with ARCH Model Forecasting Volatility of Stock Indices with ARCH Model, (April). https://doi.org/10.5430/ijfr.v4n2p126

Miles, M. B., & Huberman, M. a. (1994). Qualitative data analysis: An expanded sourcebook. *Evaluation and Program Planning*. https://doi.org/10.1016/0149-7189(96)88232-2

Prasad, S., & Choubey, M. (2018). Forecasting India ' s Total Exports : An Application of Univariate Arima Model, *9*(2).

Shumway, R. H., & Stoffer, D. S. (2011). *Time Series Analysis and Its Applications With R Examples*. *Book*. https://doi.org/10.1007/978-1-4419-7865-3

Yue, S., & Uc, X. (2012). Stock Price Forecasting Using Information from Yahoo Finance and Google Trend, 1–22. Retrieved from papers3://publication/uuid/B4CFEBCE-52C8-48E7-98C1-A943A759C6ED

McKenzie, J. (2011). Mean absolute percentage error and bias in economic forecasting. *Economics Letters*, *113*(3), 259–262. https://doi.org/10.1016/j.econlet.2011.08.010

Plummer, T. (2006). *Forecasting financial markets*. *Financial Times*.

9. ANNEXES

TOP 5 BRAZILIAN COMPANIES SÃO PAULO STOCK EXCHANGE				
Code	**Share**	**Type**	**Issued Shares**	**Participation (%)**
Total Issued Shares			**45.500.926.042**	**100**
ABEV3	AMBEV S/A	ON	4.342.975.742	7,435
PETR4	PETROBRAS	PN	4.124.795.738	5,752
ITSA4	ITAU/AS	PN N1	3.860.753.592	3,525
ITUB4	ITAU UNIBANCO	PN N1	3.119.093.161	11,153
VALE3	VALE	ON N1	3.001.189.620	7,999

Table 9.1 - Top 5 Brazilian Companies São Paulo Stock Exchange

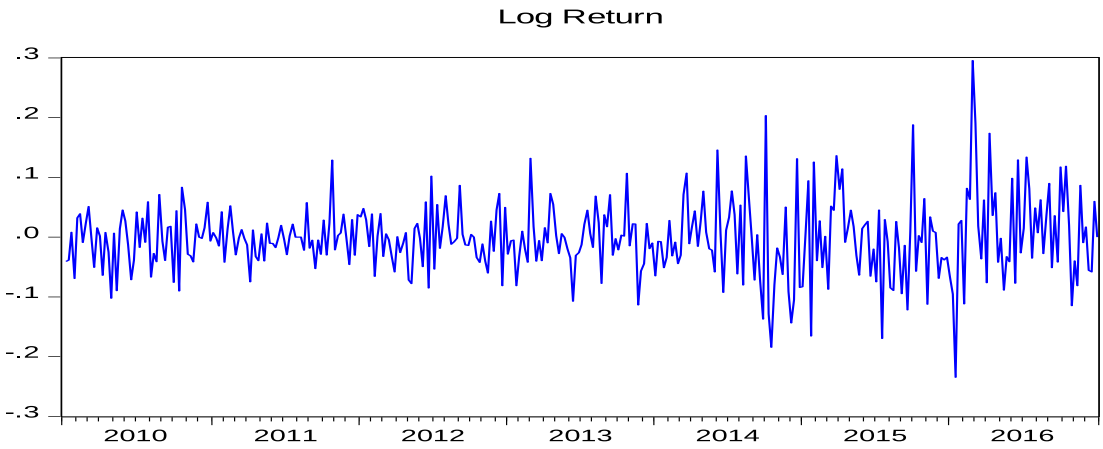

Figure 9.1 - Petrobras Log Return

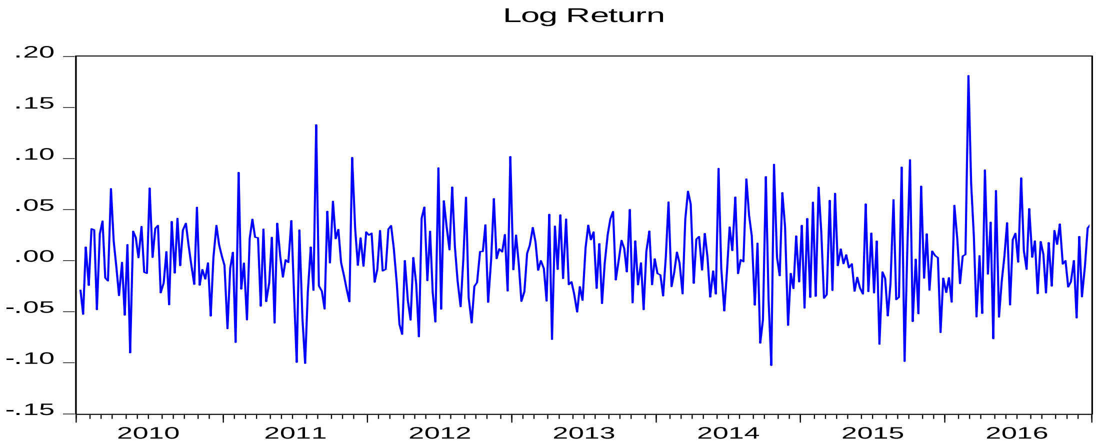

Figure 9.2 - Itau S/A Log Return

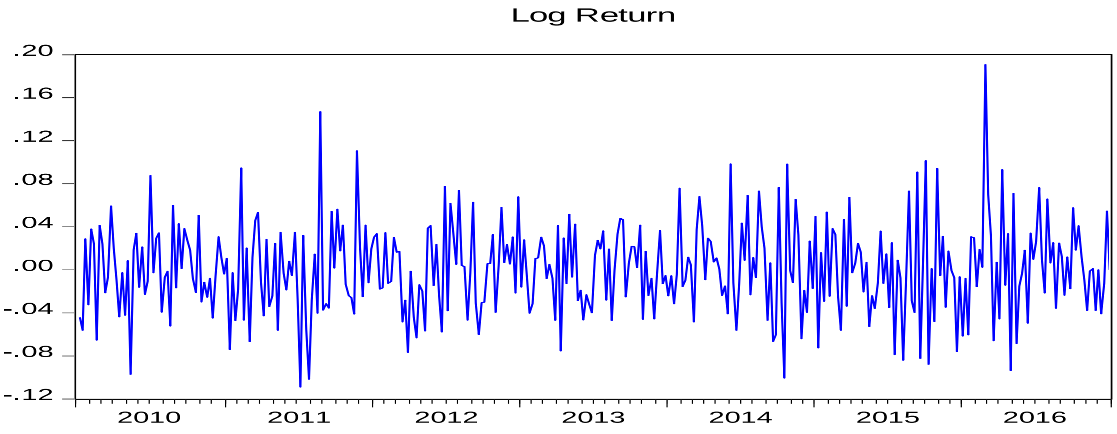

Figure 9.3 - Itau Unibanco Log Return

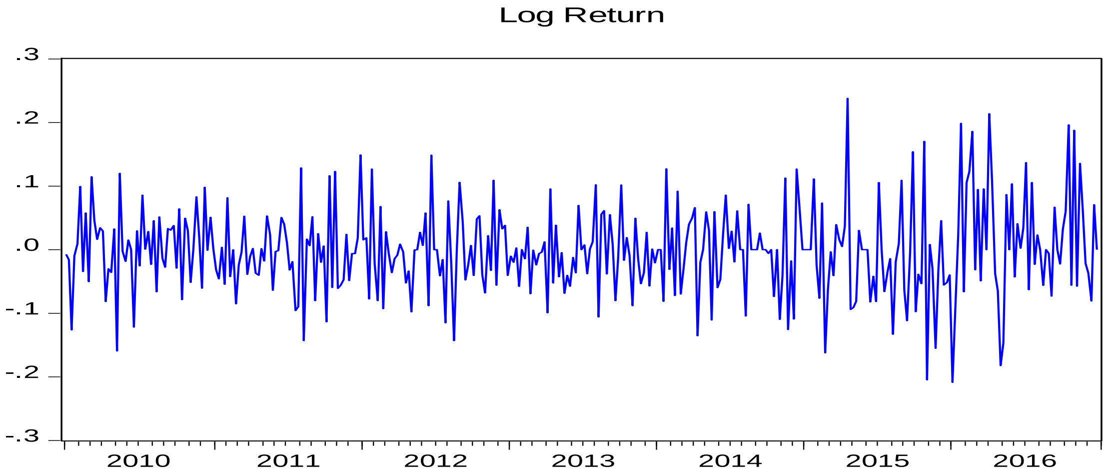

Figure 9.4 - Vale Log Return

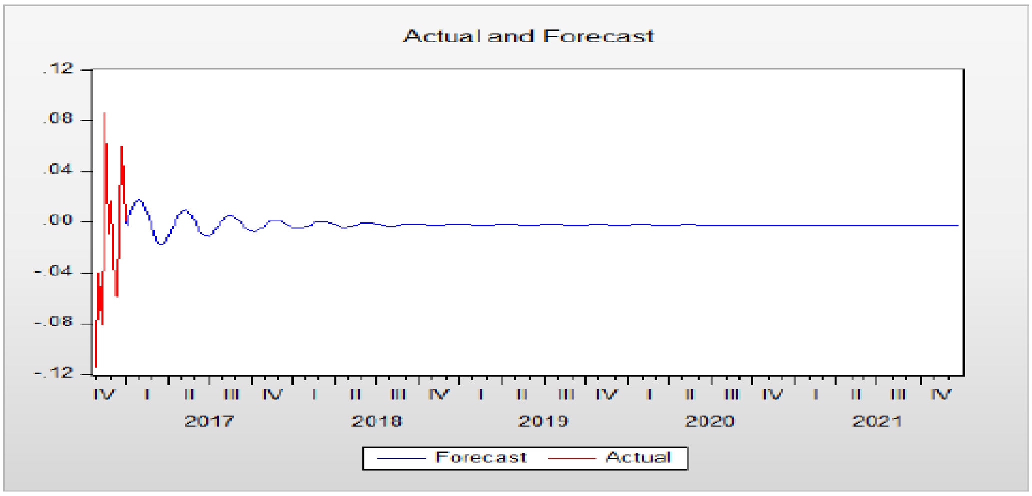

Figure 9.5 - Petrobras ARMA Forecast

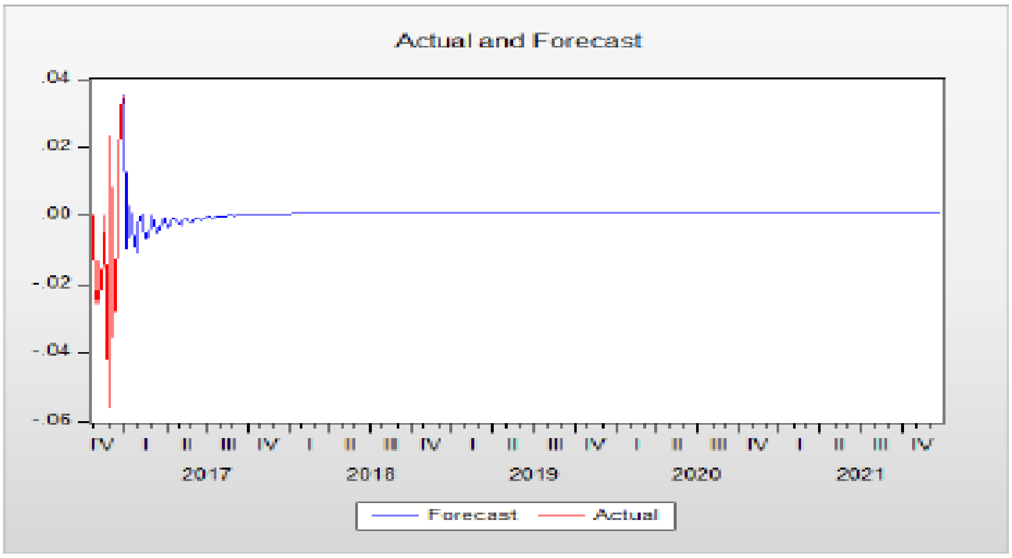

Figure 9.6 - Itau S/A ARMA Forecast

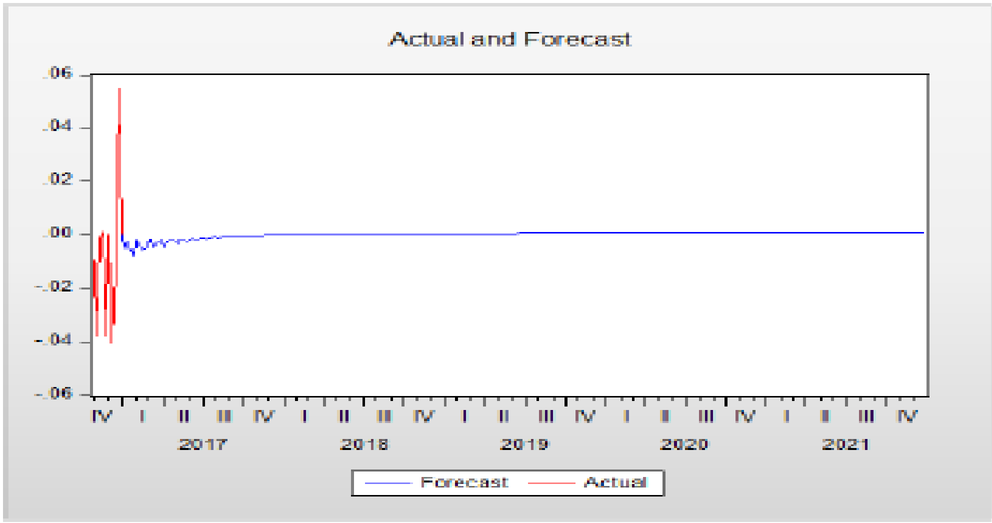

Figure 9.7 - Itau Unibanco ARMA Forecast

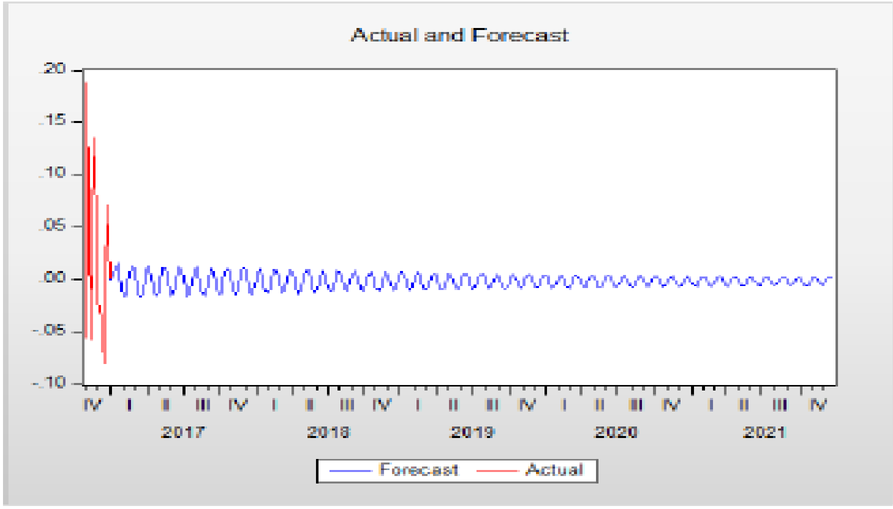

Figure 9.8 - Vale ARMA Forecast

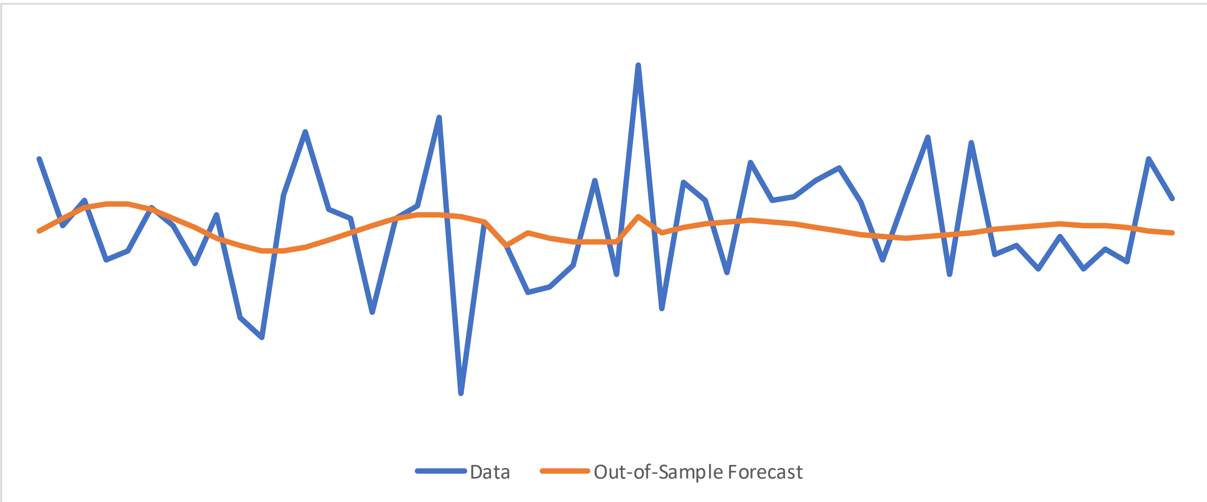

Figure 9.9 - Out-of-Sample Forecast Petrobras

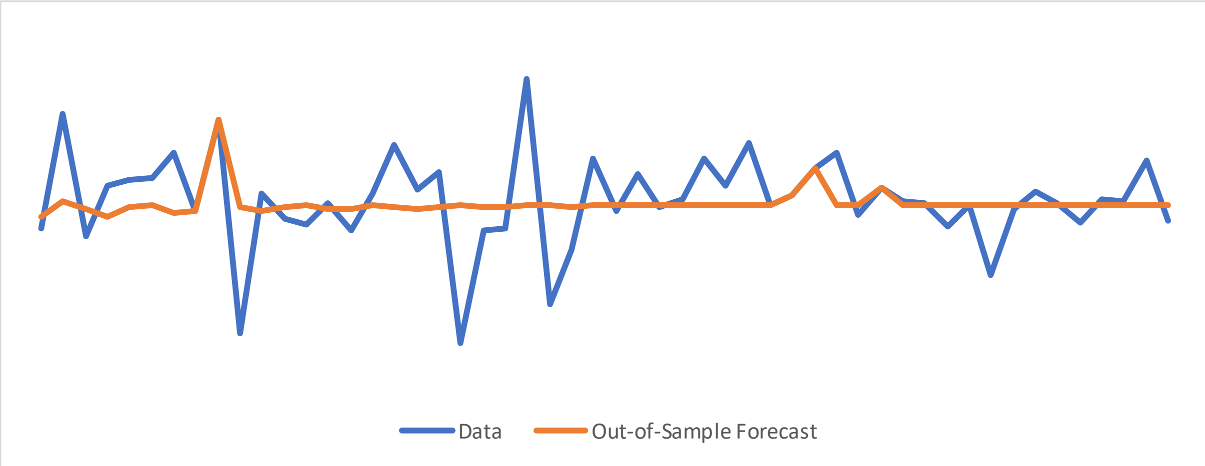

Figure 9.10 - Out-of-Sample Forecast Itau/SA

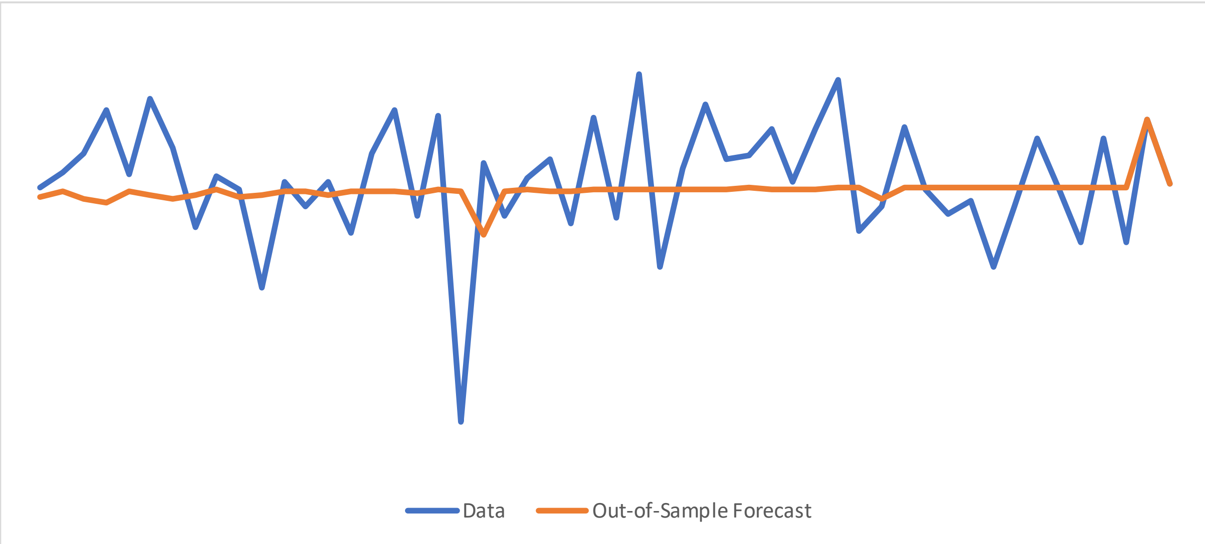

Figure 9.11 - Out-of-Sample Forecast Itau Unibanco

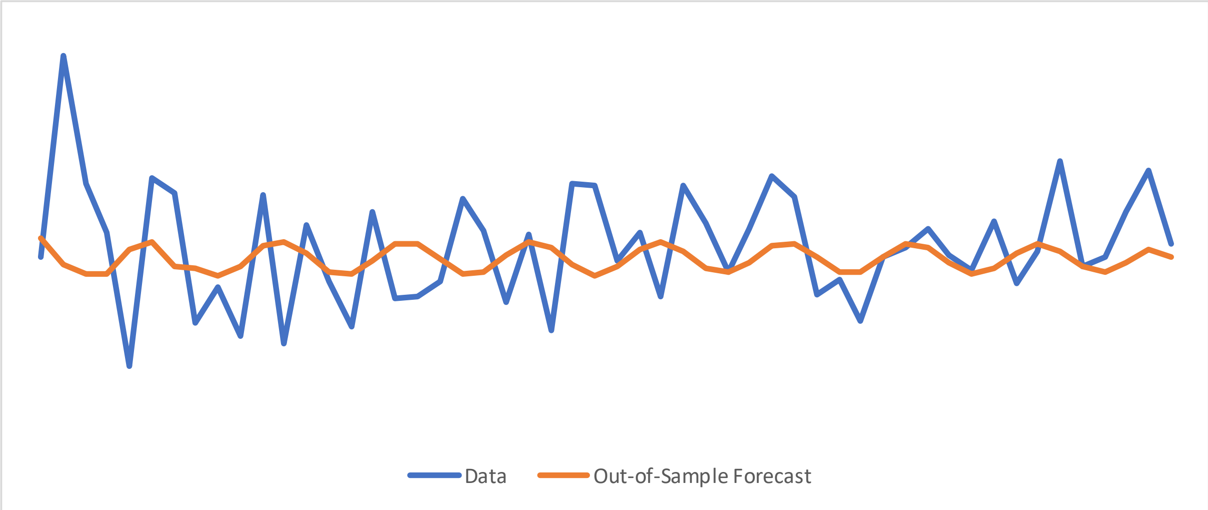

Figure 9.12 - Out-of-Sample Forecast Vale