Diversitas Journal
ISSN 2525-5215
Volume 8, Número 4 (out./dez. 2023) p. 2857 - 2876
https://diversitasjournal.com.br/diversitas_journal

Applying Box-Jenkins methodology to forecast the monthly demand of customers in a branch of *Caixa Econômica Federal*

Uso da metodologia Box-Jenkins para previsão da demanda mensal de clientes em uma agência da Caixa Econômica Federal

SILVA, Jonas Ferreira da[1]; NEPOMUCENO, Thyago Celso Cavalcante[2]; RODRIGUES, Naialy Patricia[3]

[1] 0000-0002-3009-8876; Federal University of Pernambuco (*Universidade Federal de Pernambuco*). Caruaru, Pernambuco (PE), Brazil. Email: jonas.ferreirasilva@ufpe.br

[2] 0000-0001-8327-6472; Federal University of Pernambuco (*Universidade Federal de Pernambuco*). Caruaru, Pernambuco (PE), Brazil. Email: thyago.nepomuceno@ufpe.br

[2] 0000-0002-5954-9458; Federal University of Pernambuco (*Universidade Federal de Pernambuco*). Caruaru, Pernambuco (PE), Brazil. Email: naialy.rodrigues@ufpe.br

A B S T R A C T

The management of queues in the banking service is a current and recurring challenge that service units face daily in modern times, and especially in *Caixa Econômica Federal* branches, due to its profile and social mission, such a problem is very relevant in affecting the quality of life of customers and employees, leading to effects on a scale where simple demands can result in hours of waiting. This situation is even more critical considering the current pandemic and post-pandemic context. The present study aims to estimate the demand for service in a typical branch of *Caixa Econômica Federal* (CEF) for strategic planning and distribution of resources over time and sectors of the bank. Data were collected on the times and quantities of service in the years 2019 and 2020. After data collection, the data were processed and the Box-Jenkins methodology was applied to find the best prediction models for the data reported by the establishment. It was found that from the use of Box-Jenkins models, varying their parameters, it is possible to find important results when it comes to demand forecasts. The method proved to be very effective, even with the limited amount of data.

R E S U M O

A gestão de filas no serviço bancário é um desafio atual e recorrente que unidades de atendimento enfrentam diariamente em tempos modernos, e em especial nas agências da Caixa Econômica Federais, devido seu perfil e missão social, tal problemática é bastante relevante ao afetar a qualidade de vida de clientes e empregados, levando a efeitos em escala onde demandas simples podem resultar em horas de espera, situação ainda mais crítica levando em consideração o contexto pandêmico e pós-pandêmico atual. O presente trabalho objetiva estimar a demanda de atendimento em uma agência típica da Caixa Econômica Federal (CEF) para o planejamento estratégico e distribuição de recursos ao longo do tempo e setores do banco. Foram levantados dados sobre os tempos e quantidades de atendimento nos anos de 2019 e 2020. Após a coleta, os dados foram tratados e foi aplicado a metodologia Box-Jenkins para encontrar os melhores modelos de previsão para os dados informados pelo estabelecimento. Contatou-se que a partir da utilização de modelos Box-Jenkins, variando seus parâmetros, é possível encontrar resultados importantes, quando se trata de previsões de demanda. O método se mostrou muito eficaz, mesmo com a quantidade limitada de dados.

INFORMAÇÕES DO ARTIGO

Histórico do Artigo:
Submetido: 09/11/2022
Aprovado: 02/12/2022
Publicação: 01/12/2023

Keywords:
Data analysis, Times series econometrics, ARIMA, Prediction models, Bank, Caixa Econômica Federal Bank.

***Palavras-Chave*:**
Análise de dados, Econometria de séries temporais, ARIMA, Modelos de previsão, Banco, Caixa Econômica Federal.

DOI: 10.48017/dj.v8i4.2517

Introduction

Queues are always present in society, whether to get a ticket to a movie theater or to solve problems in service units such as banks and public offices. *Caixa Econômica Federal*, also known as Caixa or CEF, is a Brazilian financial institution in the form of a private law agency, with its own assets and administrative autonomy, with outstanding importance in the formation, consolidation and development of the Brazilian State.

The services offered by the institution have always been aligned with the objectives of sustainability, social inclusion and support for populations in situations of social vulnerability, with emphasis on the Brazil Aid, PIS, FGTS, among other social programs that are part of the daily life of the working class, in addition to the monopoly of pledge operations and management of federal lotteries, one of the main sources of funding for the federal government (Silveira et al. 2013).

In this way, topics such as production capacity management are of great relevance, as the planning of operations according to the process or concept of the service can ensure that the established quality goals are met so that the customer receives what he is expecting.

In recent years, Caixa Econômica branches have had a significant increase in the number of queues in all states of Brazil for the withdrawal of FGTS and other social programs, making it impossible for the lead time, that is, the time between arrival and departure from the banking establishment, to be short due to the fact that most of the time the institution is not ready to meet the demand. In this way, the number of people waiting for care becomes relevant strategic information for the institution, enabling agile solutions so that there are no large crowds in queues (Monteiro et al. 2017).

Therefore, the objective of this proposal is to build a forecasting model based on time series capable of presenting the monthly amount of demand for certain types of service in Caixa Econômica Federal branches, allowing finding ways to meet demand more efficiently. This approach has been widely used in studies on demand forecasting in various sectors, such as the study by Apostolopoulos et al. (2020), which applied the Box-Jenkins methodology to predict electricity demand in Rural areas of Greece, obtaining promising results. Other studies, such as the one by Liu et al. (2019) also used the Box-Jenkins methodology to predict the Standardized Precipitation Index in a region of China, highlighting its effectiveness in forecasting short-term demand. In addition, the study by Sánchez-Torres et al. (2018) proposed a demand forecasting model for the tourism sector in Colombia, also based on time series, obtaining satisfactory results.

In addition, another study conducted by Amano & Almeida et al. (2023) sought to analyze the influence of external variables on the real estate services conducted by Caixa Econômica Federal and other institutions. The paper reinforces the importance of demand forecasting for the institution's strategic planning, and De Carvalho et al. (2023) applied time-series approach to predict bitcoin prices.

In this context, demand forecasting is an important tool for *Caixa Econômica Federal*, allowing the identification of patterns and trends in the demand for banking services and the planning of adequate production capacity to meet this demand. The use of the Box-Jenkins methodology is a viable and effective option to make these forecasts, since it allows the analysis of time series and the identification of seasonal patterns and trends. With this information in hand, the institution can make better decisions and ensure customer satisfaction by reducing waiting times in queues and streamlining service.

Some studies have used this methodology to predict demand in queuing systems in different contexts Sampaio et al. (2019). In a study conducted in a hospital in Taiwan conducted by Chen et al. (2011), for example, predicted the number of patients in an emergency department. In the context of Caixa Econômica Federal, a recent study used by Nepomuceno et al (2023) the econometric methods to assess the service in a branch over time. The results showed that the methodology showed good accuracy in forecasting strategies and can be used as a tool to support the management of production capacity in Caixa branches.

Efficient queue management is a challenge for public and private institutions around the world. Caixa Econômica Federal, as one of Brazil's leading financial institutions, faces this challenge on a daily basis and can benefit from applying forecasting techniques to improve demand management in its branches. The construction of a forecasting model based on time series is a promising initiative in this regard, which can bring benefits to the institution and its clients.

Development

Theoretical Framework

A Time Series can be defined as a sequence of values observed over time, at equal intervals. One of the main objectives of the study of Time Series is to create models that demonstrate the behavior of the phenomenon studied and, from there, generate predictions. Forecasting models are applied in various areas of Engineering, Economics, Medical Sciences, among others, serving as a basis for planning, allowing the evaluation of demand in advance, projecting capacity and need for resources, in addition to other activities.

The first work on ARIMA models was published in 1970, Box and Gwilym Kenking popularized ARIMA (autoregressive model integrated with moving averages) in the textbook, time series analysis: Forecasting and control (Box and Jenkins, 1970).

ARIMA models initially generated a lot of excitement in the academic community, largely due to their theoretical foundations. If their assumptions are met, they typically provide great predictions, this means that the model errors do not contain information that could improve the predictions. Methodologists call this phenomenon white noise. This does not apply, however, as ARIMA models are necessarily superior to other forecasting options with and without regular distributions (Morettin 2017)

A case study in an electronics industry used the Box-Jenkins methodology to develop forecasting models for exports and imports of paper and paper products in Turkey (Ersen et al. 2019). The process involved data collection and organization, validation and formulation of models, evaluation and generation of predictions, and analysis and comparison of results. Models were compared, using prediction error criteria (Root mean Error Square, Mean Absolute Error and Mean Absolute Percentage Error).

The forecast errors considered were the MAE (mean absolute deviation), which measures the accuracy of the forecast in relation to the mean absolute errors; the MAEP (mean percentage absolute error), which indicates the proportion of errors in relation to the current values of the series.

The ARIMA model is especially well-suited for well-behaved data due to its ability to capture time-series patterns, trends, and seasonality. Its flexibility and adaptability allow it to adjust to different forecasting scenarios, making it a reliable choice when data behaves steadily.

Another article by Pereira, S. L. A., & de Carvalho Lima, J. E. (2018) with the application of the Box-Jenkins model in forecasting the production of automobiles. First, the data on car production in the period between January/2007 and June/2018 were collected and then it was plotted in the form of time series, then the analysis of the correlograms began to find the best parameters for the ARMA model, and after this analysis it was tested if the model is stationary, which in fact happened, i.e. it is not necessary to use the ARIMA model.

The authors used a test known as Kruskal-Wallis to understand if there is seasonality in the data, and the test proved that there is. Soon it was possible to find all the elements to model the problem, after finding the ideal model it was tested and passed the autocorrelation test and also the normality of the residuals, and finally its predictions were plotted. And then it was concluded that in fact the ARIMA model showed a great option for forecasting production, since it did not show a trend in the data, that is, they were random, which in fact happens in real life, since the demand in the purchasing sector, especially automobiles, it varies a lot as the months go by.

These articles have shown that if used correctly, respecting the assumptions and diagnosing the data in a coherent way, it is possible to find optimal predictions with the Box-Jenkins method.

Methodological Procedure

The *locus* of the research took place in a CEF agency, located in *Jaboatão dos Guararapes*, a municipality in the state of Pernambuco, a city that, according to the estimate of the Brazilian Institute of Geography and Statistics, is the second municipality in the state with the highest number in the Human Development Index (HDI) with 0.717, and an average income of two minimum wages for formal workers with an estimated population of 711,330 inhabitants and having 5 CEF branches.

Despite a good HDI compared to other municipalities in the same state, people who need to use the services both for the Emergency Aid and for the withdrawal of the FGTS needed to fit into some economic requirements of which they are not contributors to a high HDI. Another point is the location of this agency, which made it possible to move people from various communities and from nearby surrounding cities with no agencies.

The choice of the specific agency was due to the fact that one of the authors was a former employee, so he was able to access the system and thus, it was possible to obtain the data, referring to the years 2019 and 2020, as they were the periods available at the time of collection, the information was separated monthly, with a total of twenty-four observations and were separated into some categories, for this present work, three of them will be discussed, Caixa, FGTS and Expresso, as they are the ones that represent the largest volume of service. The objective of this article is, based on the Box-Jenkins methodology, to obtain the number of services that are performed by these three sectors and the fund to allocate service capacity in the best way to meet this demand. Table 1 report the descriptive statistics.

Table 1.

Descriptive statistics of the data.

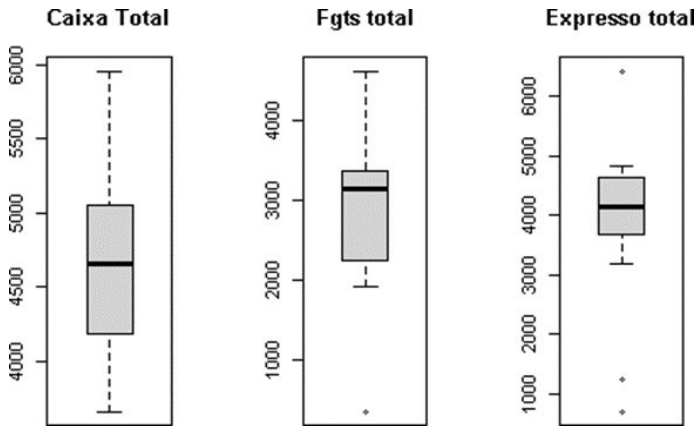
Year	Group	Min.	1st Qu.	Median	Average	3rd Qu.	Max.	Sd
2019	Caixa	3658	4191	4661	4660	4692	5953	641,6
2019	Expresso	693	3688	4136	3867	4588	6421	1442,08
2019	FGTS	358	2237	3148	2810	3365	4606	1048,52
2020	Caixa	995	1151	1821	2325	3538	5314	1321,93
2020	Expresso	657	3160	4634	4611	6161	8128	2414,01
2020	FGTS	1	18,5	77	1194,7	2199	3850	1557,63

Source: Caixa Econômica Federal.

Based on the values presented in table 1, it is possible to perceive a discrepancy between the data for 2019 and 2020, this is since in 2020 there was the pandemic and, consequently, the lockdown, which made people unable to go to the branches, greatly increasing the uncertainties of demand this year. Graphs 1 and 2 illustrates the boxplots.

Graph 1.

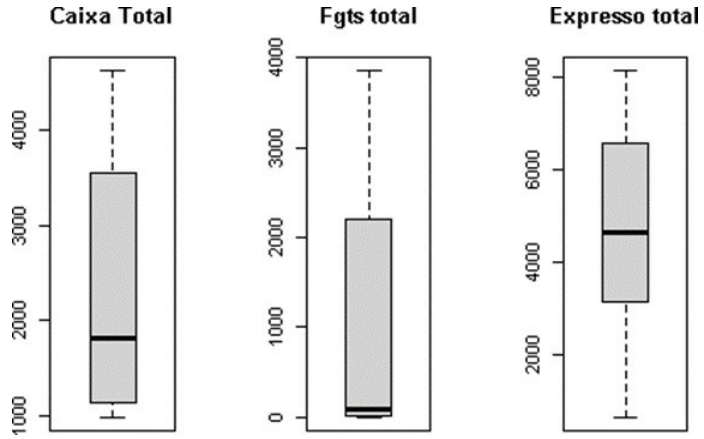
Data Box-Plot for the year 2019.



Source: Authors (based on data from Caixa).

Graph 2.

Data Box-Plot for the year 2020.



Source: Authors (based on data from Caixa).

For better visualization, a boxplot is presented in graphs 1 and 2, which is a type of statistical graph used to represent the distribution of data in a set of observations. Through it, it can be seen how the data varies, especially in 2020, due to the pandemic, where there were greater variations in service, which makes it difficult to create an assertive model for the scenario, but it can prepare the place for some future event that may vary the demand.

The work followed several stages. First, the data were processed to become time series. They were then evaluated using descriptive statistics to understand their behavior. The analysis of the correlograms helped to identify the models to be considered. Using the Akaike criterion (AIC), the best parameters were selected to explain the data, minimizing the AIC. However, the minimum value of AIC alone is not sufficient to affirm that the model is adequate. The presence of seasonality in the data was also investigated. Then, the diagnosis was performed to test whether the model actually explains the data. This involved the evaluation of the residuals, including the analysis of the correlations to identify correlations and the use of the qqplot to ascertain the normal distribution of the residuals. If the model went through all of these steps, with the lowest AIC, the best seasonality, and well-behaved residuals, it would be chosen to predict the future of distribution.

ARIMA Model

These models consider three processes, they are the autoregressive, moving averages and the amount of differences necessary for the data to be stationary, these differences are obtained as follows $yt = yt - yt{-}1$, And this will be done for all data until it is stationary.

If Yt is modeled as $(Yt - \delta) = \alpha 1(Yt{-}1 - \delta) + ut$ (I)

Where δ is the average of Y and ut is an uncorrelated random error with zero mean and constant variance $\sigma 2$ (it's a white noise), so Yt follows a first-order stochastic autoregressive process or AR(1). In this case, the value of Y in period t depends on its value in the previous period and on a random term; Y-values are expressed as deviations based on a mean value. In

other words, this model states that the predicted value of Y in period t is simply some proportion (=α1) plus a random shock or disturbance in period t; again, the values of Y are expressed around their mean values.

But if Y follows the following model $(Yt - \delta) = \alpha 1(Yt{-}1 - \delta) + ut + \alpha 2(Yt{-}2 - \delta) + u$ (II)
$(Yt - \delta) = \alpha 1(Yt{-}1 - \delta) + ut + \alpha 2(Yt{-}2 - \delta) + \cdots + \alpha p(Yt{-}p - \delta) + ut$ (III)

In that case Yt is an autoregressive process of order p-th, or AR(p). Moving Average Process (MA).

As mentioned earlier, the AR process is not the only mechanism that may have generated Y, assuming that Y is modeled following the equation $Yt = \mu + \beta 0ut + \beta 1ut{-}1$ (IV) Where u is a white noise stochastic error term. In this case, Y in the period t is equal to a constant plus a moving average of the current and past error terms. Therefore, it can be said that Y follows a first-order moving average process, or a MA(1). But if Y follows the expression $Yt = \mu + \beta 0ut + \beta 1ut{-}1 + \beta 2ut{-}2$ (V). So it's a process MA(2). More generally, $Yt = \mu + \beta 0ut + \beta 1ut{-}1 + \beta 2ut{-}2 + \cdots + \beta qut{-}q$ (VI) is a MA(q) process.

In short, a moving average process is just a Linear Combination of White Noise Error Terms in Autoregressive process of moving averages (ARMA). It may happen that Y has characteristics of both AR and MA and is, therefore, ARMA. Yt follows a process ARMA(1,1) and can be written as $Yt = \theta + \alpha 1Yt{-}1 + \beta 0ut + \beta 1ut{-}1$ (VII) So there's an autoregressive term and a moving average term. In the equation VII θ represents a constant term.

In general, in a process ARMA(p,q), there will be autoregressive terms p and moving average terms q following the Autoregressive process of moving averages (ARMA). If it is necessary to differentiate a time series d times to make it stationary and apply the ARMA(p,q) model, it can be said that the time series is ARIMA(p,d,q), that is, it is an integrated autoregressive time series of moving averages, where p denotes the numbers of the autoregressive terms, d denotes the number of times the series must be differentiated before it becomes stationary and q the number of moving average terms. An ARIMA time series(2,1,2) must be differentiated once (d=1) before making it stationary, and the stationary time series (first difference) can be modeled as an ARMA process(2,2), since it has two AR and two MA terms. Of course, if d=0, a series is stationary for ARMA(p,q). An ARIMA(p,0,0) process, for example, means a purely stationary AR(p) process; an ARIMA(0,0,q) means a purely stationary MA(q) process. Given the values of q, d and q it is possible to tell which process is being modeled.

It was mentioned earlier that the term u is a white noise, in terms of ARIMA this characteristic can be described as ARIMA(0,0,0), that is, it is a stationary process and has no Autoregressive parameters or moving averages.

Results and Discussion

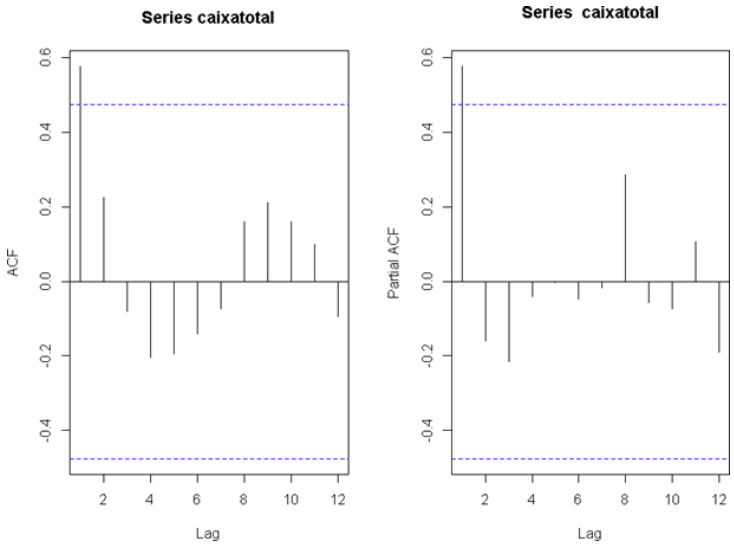
The following predictions and results are developed in the R language. The first group to be analyzed will be the one in the box, as already mentioned by the ARIMA methodology

that requires the dataset or its differences to be stationary, so this test will be done primarily.

In the Dickey-Fuller test two hypotheses are considered, the null hypothesis is that the data are not stationary and the alternative is that they are stationary, so for p- values greater than 0.05 (which was the maximum value defined by the author) there is not enough evidence to infer that the data are stationary, so the result generated by the test was 0.3869. the data is not stationary, so in ARIMA the first difference will be considered.

Graph 3.

Data Correlograms.

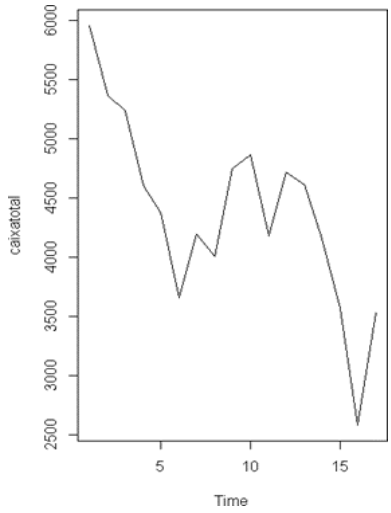


Source: The authors (2023).

Looking at graph 3, some models can be considered, they are (1,1,1), (1,1,0) and (0,1,1), but this is not accurate enough because the data depends on seasonality. Thus, the graph must be analyzed to find this parameter. This graph is represented in Graph 4.

Graph 4.

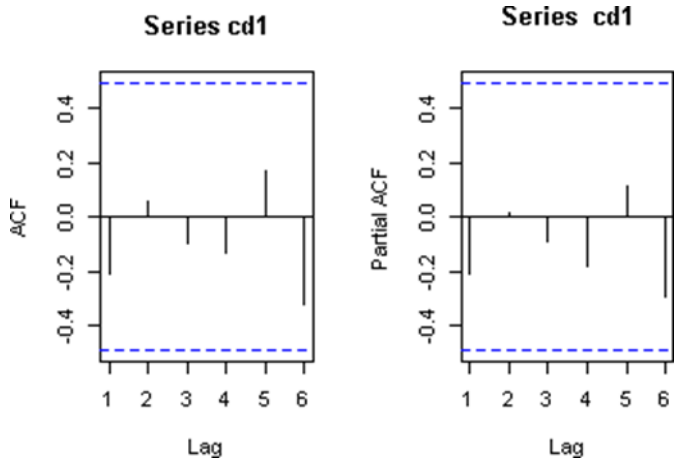
Data Plot.



Source: The authors (2023).

Graph 5.

Correlogram for seasonality.



Source: The authors (2023).

From this analysis, it is possible to define that the best frequency for this seasonality will be 6 periods. For the seasonal part, two models were tested, (1,1,1) and (0,1,0) based on Graph 5. The table of these models is presented below.

Table 2.

Selected models for prediction of data from Caixa.

Model	ARIMA	Seasonality	AIC
M1	(1,1,1)	(0,1,0)	169.01
M2	(1,1,1)	(1,1,1)	169.41
M3	(0,1,1)	(0,1,0)	167.88
M4	(1,0,0)	(0,1,0)	167.88
M5	(0,1,0)	none	251.06

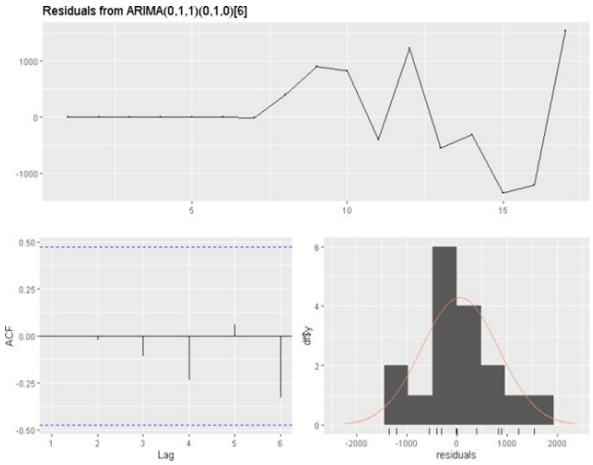
Source: The authors (2023).

Considering the values of the AIC to try to find the best choice it is not possible to simply analyze this parameter, since the values of M3 and M4 are the same, so another variable must be found which are the sum of the errors known as sigma squared, for M3 this value is 935767.4 and that of the M4 model is 936023.5 as M3 has the lowest value, it will be the chosen model.

Now that the best parameters have been found to explain the data, it is necessary to test if they are good enough and, for this, it is necessary to use the diagnostics stage, where an analysis will be carried out on the model residues, these plots are found in graph 6.

Graph 6.

Graphs for prediction model diagnostics for Caixa.



Source: The authors (2023).

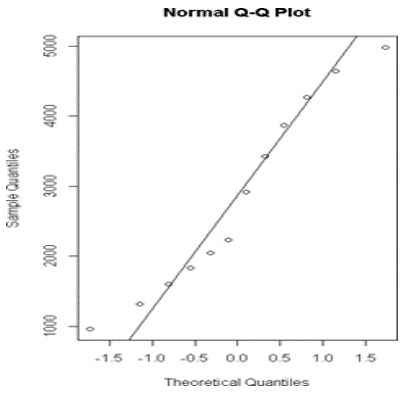
In graph 6 two things will be analyzed, the first is if all the data are within the blue range, which in fact happens meaning that they are not autocorrelated, this can be confirmed by means of the Ljung-box test.

In the Dickey-Fuller test presented above, the idea was to reject the null hypothesis, that is, to find the p-value low enough, for this test the objective is different, that is, the idea is to find a p-value high enough to test if data are self-correlated. In this case the result is 0.6399, supporting correlation.

The second parameter to be analyzed is the distribution of the residuals, which is shown in the histogram presented in graph 6. Using QQplot in Graph 7 we can investigate whether data follows a normal distribution.

Graph 7.

Data scatter plot.



Source: The authors (2023).

Analyzing this plot in graph 7 it seems in fact that each point is close enough to the line to confirm the hypothesis, To test this assumption we perform Shapiro-wilk test. The test has similar interpretation: the idea is to find a high enough p-value not to reject the null hypothesis. The result found was 0.3033, the residuals in fact follow a normal distribution.

After all this information, it can be concluded that in fact the chosen model explains the data well, and then it is possible to find the values that are the objective of this article, which are the predictions, which can be found in the table below.

Table 3.

Prediction data for the selected model.

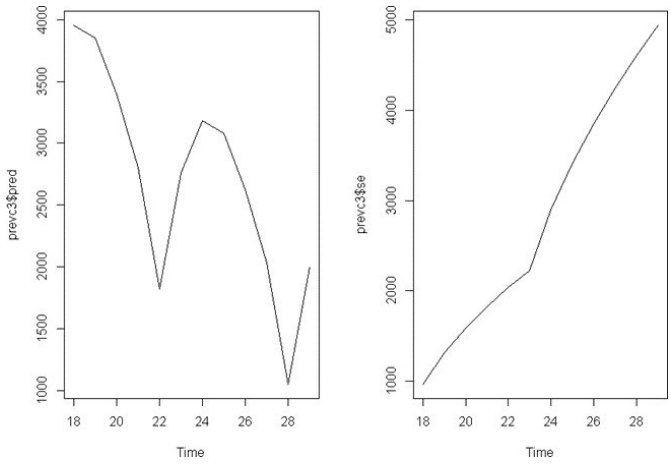
Months ahead	Forecast
1	3951
2	3847
3	3392
4	2803
5	1821
6	2761
7	3184
8	3080
9	2625
10	2036
11	1054
12	1994

Source: The authors (2023).

The graph of this forecast is plotted along with the graph of the residuals:

Graph 8.

Prediction and error plot.



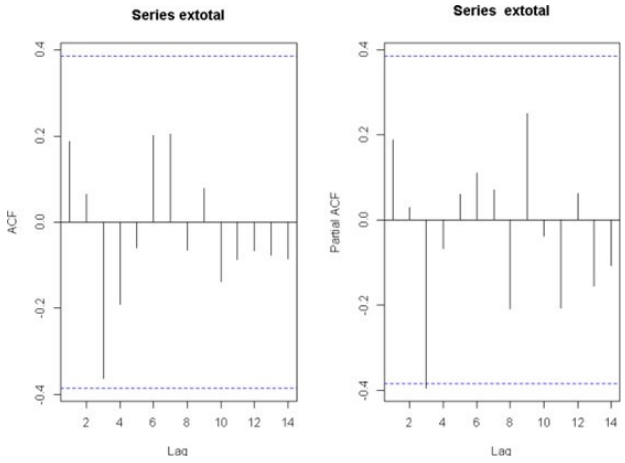
Source: The authors (2023).

Forecasts for Grupo Expresso

Again, the first step is to test whether the data is stationary, and as already shown, this can be done using the Dickey-Fuller test. Since the p-value is greater than 0.05 (0.07238), then the set is said to be non-tationary and the first difference will be considered for the ARIMA model. The next step is to analyze the correlation of the data.

Graph 9.

Expresso data correlogram.

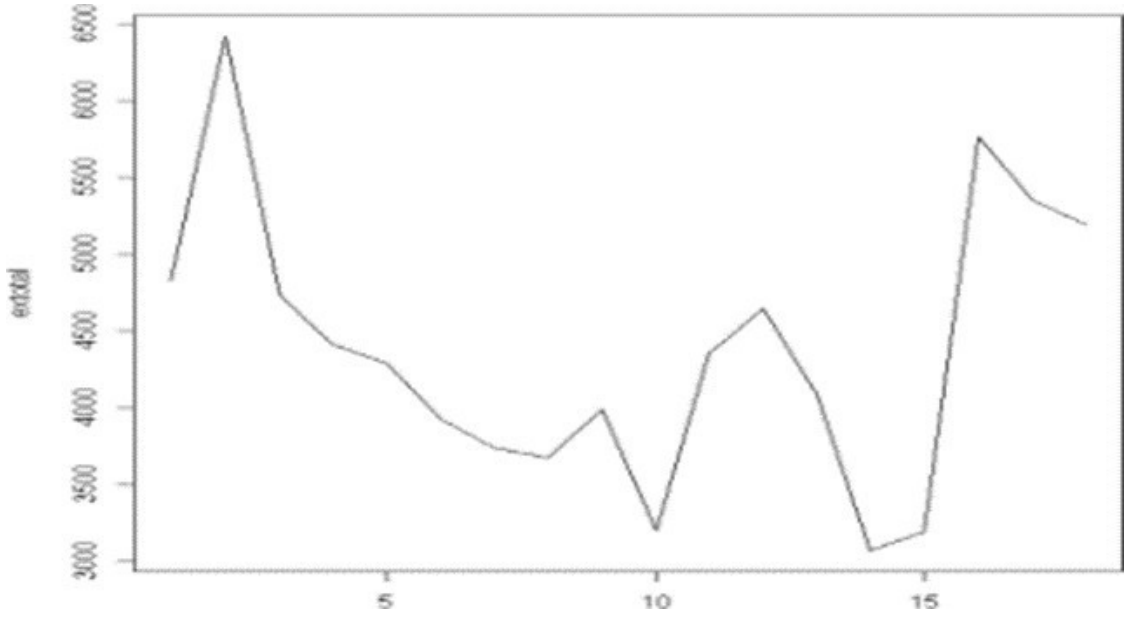


Source: The authors (2023).

For this case, as shown in Graph 9, several models can be considered, but the main ones are: (0,1,0), (0,1,3), (0,1,1) and (1,1,1). However, as already mentioned, this will not be enough to ensure efficient forecasting, so one should also analyze the seasonality of the data.

Graph 10.

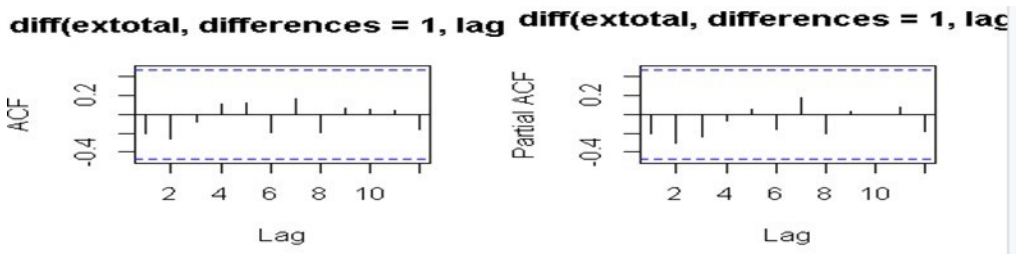
Expresso data plot.



Source: The authors (2023).

Graph 11.

Correlogram for seasonality.



Source: The authors (2023).

The seasonality period is 12, as for the parameters it is tested only (0.10), as it is a random component. Based on all this information, it is possible to select models to try to predict such data.

Table 4.

Selected models to predict expresso data.

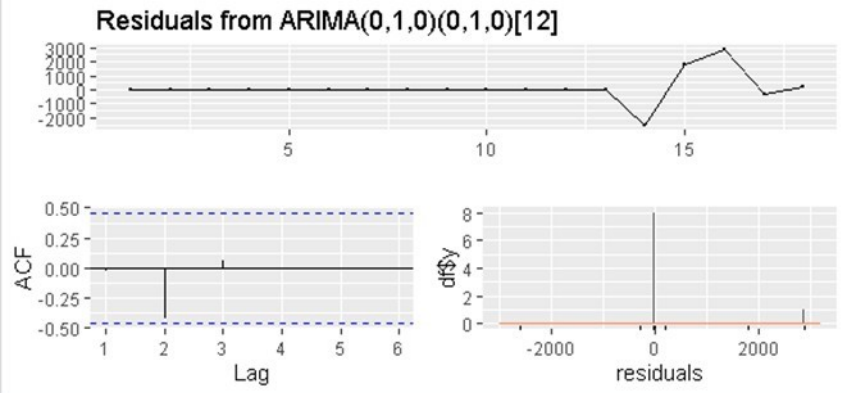
Model	*ARIMA*	*Seasonality*	*AIC*
M1	*(0,1,1)*	*(0,1,0)*	*91.78*
M2	*(0,1,0)*	*(0,1,0)*	*89.8*
M3	*(0,1,3)*	*(0,1,0)*	*93.65*
M4	*(1,1,1)*	*(0,1,0)*	*93.4*
Auto arima	*(0,0,1)*	*none*	*297.48*

Source: The authors (2023).

By the criterion of the lowest AIC, M2 presents the best parameters to explain the data. The next step is to diagnose the model, that is, as shown above, this part serves to show if in fact the chosen model explains the data well.

Graph 12.

Graphs for the diagnosis of the model chosen for expresso.

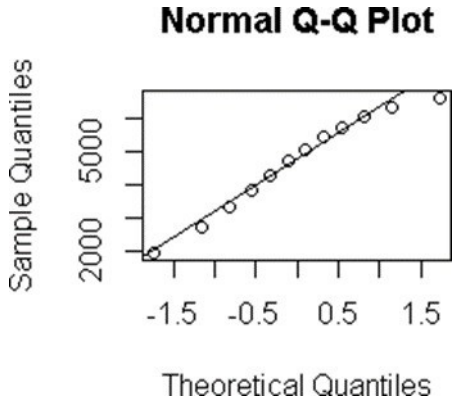


Source: The authors (2023).

The residuals are not correlated, since in the correlogram they are all within the blue zone, in the line graph there are 4 points that are not behaved, but otherwise it is ok and to evaluate the assumption that the errors follow a normal it is necessary to plot the qq graph.

Graph 13.

Normality test.



Source: The authors (2023).

The points are close enough to the line, which is an indication that the errors follow a normal distribution also indicated by Shapiro test, whose result is 0.7729, confirming the hypothesis. Therefore, having the confirmation that the model is good enough to explain the data well, one can start with the prediction, whose table can be found below.

Table 5.

Data with the result of the forecast for the expresso.

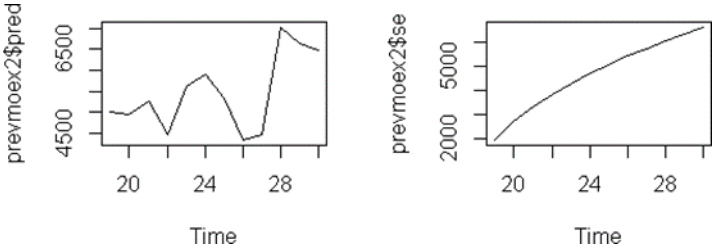
Months ahead	Forecast
1	5011
2	4943
3	5259
4	4470
5	5624
6	5918
7	5346

Source: The authors (2023).

Graph 14 shows the plot with the data of this prediction along with the error.

Graph 14.

Prediction and error data plot.



Source: The authors (2023).

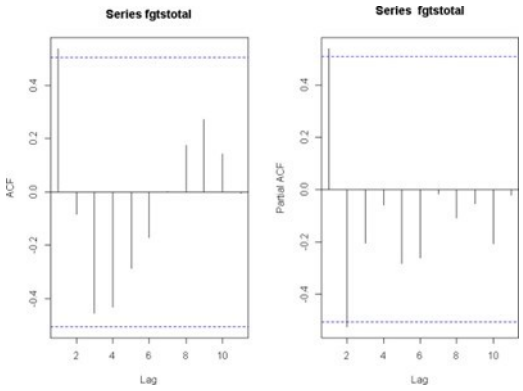
Forecast for the FGTS Group

Stationarity test

The set is not stationary because the p-value was 0.2314 so the first differences should be applied.

Graph 15.

Correlogram for FGTS data.

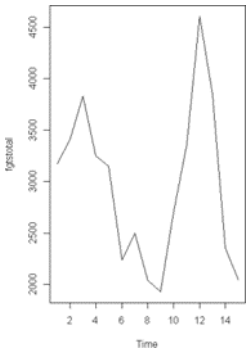


Source: The authors (2023).

In this case, the models considered are (0,1,0), (1,1,1), (0,1,2), (0,1,1) and (1,1,0).

Graph 16.

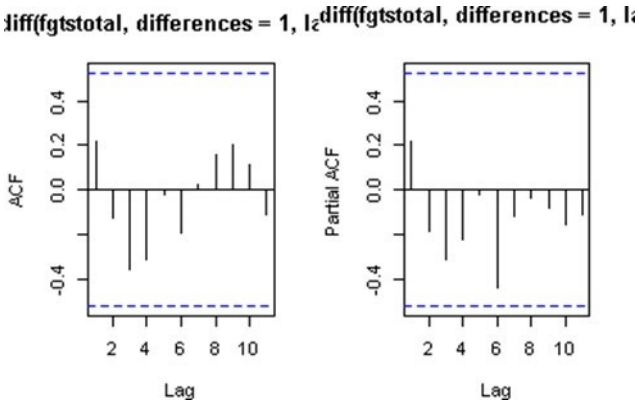
FGTS data plot.



Source: The authors (2023).

Graph 17.

Correlogram for seasonality.



Source: The authors (2023).

The seasonality period is 9 and given the correlogram only the set (0,1,0) is tested.

Table 6.

Selected models to predict FGTS data.

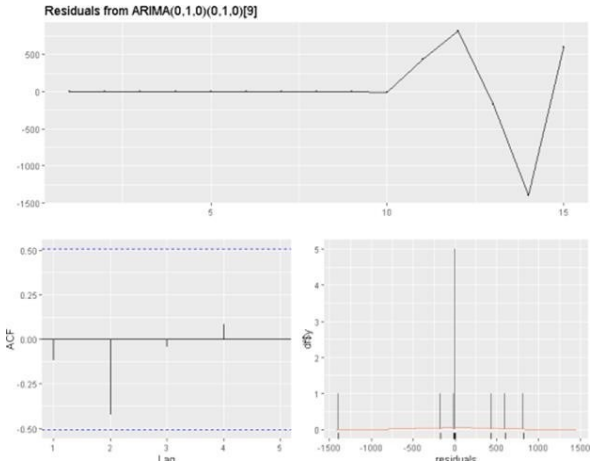
Model	ARIMA	Seasonality	AIC
M1	(0,1,0)	(0,1,0)	81.0252
M2	(1,1,1)	(0,1,0)	83.7655
M3	(1,1,0)	(0,1,0)	82.9552
M4	(0,1,1)	(0,1,0)	82.1
Auto.Arima	(0,0,1)	none	240.76

Source: The authors (2023).

Minor AIC belongs to the M1 model, so it will be chosen for the diagnostic part. The residuals plot is shown in graph 18.

Graph 18.

Diagnosis of the model chosen to predict FGTS.

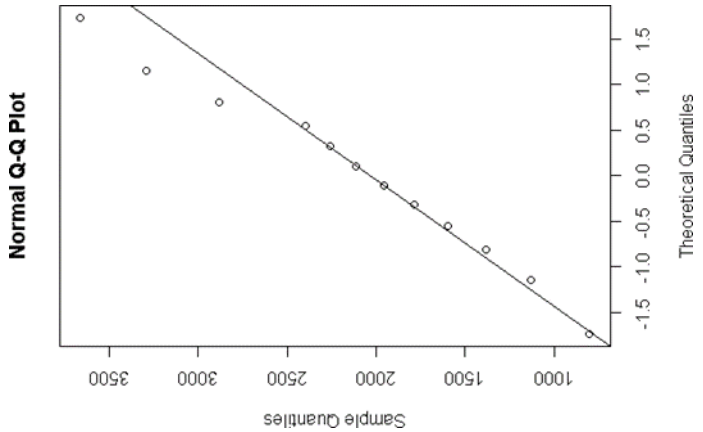


Source: The authors (2023).

All lags are within the statistically acceptable zone, so they are not autocorrelated, but once again it is not possible from the histogram to state that they follow a normal distribution, so qqplot (graph 17) and Shapiro (figure 7) will be used.

Graph 19.

Normality plot.



Source: The authors (2023).

As can be seen in chart 17, there are 3 leverage points, but they are in the same direction as the other points, so they are good leverage points and as found in Shapiro's test, the p-value is 0.9661, they do not interfere with the normality of the residuals.

Table 6.

Data with the forecast for FGTS.

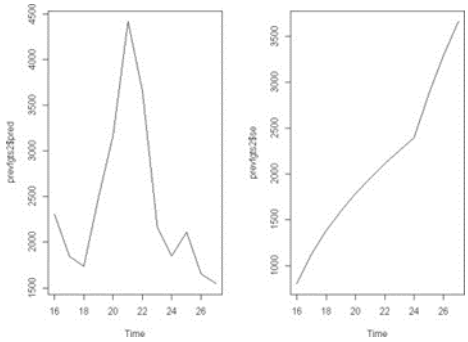
Months ahead	Forecast
1	2304
2	1845
3	1738
4	2500
5	3172
6	4413
7	3657
8	2161
9	1851
10	2111
11	1652
12	1545

Source: The authors (2023).

Therefore, the model explains the data well and its predictions can be found in the table below. Graph 20 shows the plot of the prediction and the error.

Graph 20.

Plot with prediction and error data.



Source: The authors (2023).

Conclusions

The case study presented showed several prediction models based on the Box-Jenkins methodology, where it was possible to find good results about the data provided by a branch of the federal savings bank, all the models followed the necessary methods so that there are no exorbitant errors when compared with the originals, that is, they went through the diagnostic stage and all proved to be good enough to explain the data.

That said, the objective of the article was achieved, that is, it was possible to show how

much the demand of customers will be for the institution so that they can look for ways to meet such demand.

The case study has limitations in the data, that is, there were not enough to understand the real behavior of customers well enough, since several elements about the year 2020 were removed due to the pandemic.

Finally, future research on the ARIMA methodology is recommended, since as it has been shown it is a model of easy understanding, easy application and due to its steps it is also very reliable, it is also recommended the use of exogenous variables, in other words they are variables that are outside the model, but that are related to it and, thus, it is possible to find better regressors for the final model, always using the AIC criterion to determine whether they are good variables or not for the model.

REFERENCES

Amano, F. H. F., & Almeida, R. P. (2023). Tributação e dinâmica imobiliária: uma análise comparativa para seis aglomerações brasileiras. Nova Economia, 33, 181-209.

Apostolopoulos, N., Chalvatzis, K. J., Liargovas, P. G., Newbery, R., & Rokou, E. (2020). The role of the expert knowledge broker in rural development: Renewable energy funding decisions in Greece. Journal of Rural Studies, 78, 96-106.

Box, G. E.; Jenkins, G. M.; Reinsel, G. C.; Ljung, G. M. (2015). Time series analysis:forecasting and control. John Wiley & Sons.

Chen, C. F., Ho, W. H., Chou, H. Y., Yang, S. M., Chen, I. T., & Shi, H. Y. (2011). Long-term prediction of emergency department revenue and visitor volume using autoregressive integrated moving average model. Computational and mathematical methods in medicine, 2011.

de Carvalho, L. M., Resende, G. P., & Takahashi, M. (2023). An Explanatory Model for the price of Bitcoin and the public interest in the topic. Socioeconomic Analytics, 1, 126-140.

Ersen, N., Akyüz, İ., & Bayram, B. Ç. (2019). The forecasting of the exports and imports of paper and paper products of Turkey using Box-Jenkins method. Eurasian Journal of Forest Science, 7(1), 54-65.

Liu, Q., Zhang, G., Ali, S., Wang, X., Wang, G., Pan, Z., & Zhang, J. (2019). SPI-based drought simulation and prediction using ARMA-GARCH model. Applied Mathematics and Computation, 355(C), 96-107.

Monteiro, E. B., Lopes, F. L., & Silva, M. R. (2017). Capacidade produtiva na gestão de serviços: revisão sistemática da literatura. Revista Brasileira de Gestão e Inovação, 5(2), 1-22.

Morettin, P. A. (2017). Econometria financeira: um curso em séries temporaisfinanceiras. Editora Blucher.

Nepomuceno, T. C. C., de Carvalho, V. D. H., Nepomuceno, K. T. C., & Costa, A. P. C. (2023). Exploring knowledge benchmarking using time-series directional distance functions and bibliometrics. Expert Systems, 40(1), e12967.

Pereira, S. L. A., & de Carvalho Lima, J. E. (2018). Aplicação do Modelo Box-Jenkins na Previsão da Produção de Automóveis. Id On Line: Revista Multidisciplinar e de Psicologia (12):42.

Sampaio, P., Cabral, J., & Costa, J. (2019). The impact of waiting lines on service qualityand customer satisfaction in healthcare. International Journal of Healthcare Management, 12(2), 89-94.

Sánchez-Torres, J. A., Sandoval, A. V., & Alzate, J. A. S. (2018). E-banking in Colombia: factors favouring its acceptance, online trust and government support. International Journal of Bank Marketing, 36(1), 170-183.

Silveira, F., Vieira, M. M. F., & De Castro, D. C. (2013). A Presença do estado no setor financeiro Brasileiro: o caso da Caixa Econômica Federal. GESTÃO. Org, 11(1), 132-159.