

# PROBABILISTIC AND TEMPORAL MODELING OF THE COFFEE MARKET IN BRAZIL

## MODELAGEM PROBABILÍSTICA E TEMPORAL DO MERCADO DE CAFÉ NO BRASIL

Article received on: 12/10/2025

Article accepted on: 3/13/2026

### Otávio Magno Teixeira\*

\*Universidade Federal de São Carlos (UFSCAR), Araras, São Paulo, Brasil

Orcid: <https://orcid.org/0000-0001-6565-6017>

[otaviomagno@estudante.ufscar.br](mailto:otaviomagno@estudante.ufscar.br)

### Gilberto Rodrigues Liska\*

\*Universidade Federal de São Carlos (UFSCAR), Araras, São Paulo, Brasil

Orcid: <https://orcid.org/0000-0002-5108-377X>

[gilbertoliska@ufscar.br](mailto:gilbertoliska@ufscar.br)

### Jeronimo Alves dos Santos\*

\*Universidade Federal de São Carlos (UFSCAR), Araras, São Paulo, Brasil

Orcid: <https://orcid.org/0000-0002-4793-4973>

[jeronimo@ufscar.br](mailto:jeronimo@ufscar.br)

### Marta Cristina Marjotta-Maistro\*

\*Universidade Federal de São Carlos (UFSCAR), Araras, São Paulo, Brasil

Orcid: <https://orcid.org/0000-0003-2548-6214>

[marjotta@ufscar.br](mailto:marjotta@ufscar.br)

### Daniel Augusto de Melo Pedro\*\*

\*\*Universidade Federal de Alfenas (UNIFAL-MG), Alfenas, Minas Gerais, Brasil

Orcid: <https://orcid.org/0009-0007-2000-9124>

[danielmelo0416@gmail.com](mailto:danielmelo0416@gmail.com)

The authors declare that there is no conflict of interest

### Abstract

Coffee is one of the most valuable commodities in the world, and in this scenario, Brazil stands as the largest producer and exporter globally. In this sense, this work aims to propose the best probabilistic model for the monthly analysis of prices, calculate probabilities of average prices per coffee sack according to levels of practical interest, and investigate the occurrence of trends and seasonality in the coffee market price series through time series models. For this purpose, historical data provided by COOXUPÉ for the period from January 1981 to December 2022 were used. The results indicate that the Gamma distribution performed better in most months, except for January, February, and March, where the log-Normal distribution stood out. February, January, and March, respectively, showed the highest probabilities of prices exceeding 200 and 400 dollars. On the other hand, among all the months, October has the lowest probability of values exceeding 100 and/or 200 dollars.

### Resumo

O café é uma das mais valiosas commodities do mundo e, nesse cenário, o Brasil se apresenta como o maior produtor e exportador mundial. Nesse sentido, o trabalho tem por objetivo propor o melhor modelo probabilístico para a análise mensal de preços, calcular probabilidades de ocorrências de preços médios de sacas de café de acordo com níveis de interesse prático e investigar a ocorrência de tendência e sazonalidade nas séries de preço do mercado de café via modelos de séries temporais. Para tal, utilizou dados históricos fornecidos pela COOXUPÉ correspondente ao período de janeiro de 1981 a dezembro de 2022. Os resultados apontam que a distribuição Gama obteve melhor desempenho na maioria dos meses, com exceção dos meses de janeiro, fevereiro e março em que a distribuição log-Normal se destacou. Os meses de fevereiro, janeiro e março, respectivamente, apresentaram as maiores probabilidades para a ocorrência de



Regarding the temporal analysis, the  $SARIMA(3,1,3)(2,1,1)_{12}$  model showed the smallest prediction errors and can be recommended for making forecasts.

**Keywords:** Time Series. R Programming Language. Commodity. Probability. Coffee Prices.

*preços que ultrapassam 200 e 400 dólares. Em contrapartida, entre todos os meses, o mês de outubro possui a menor probabilidade de valores superiores a 100 e/ou 200 dólares. Com relação à análise temporal, o modelo  $SARIMA(3,1,3)(2,1,1)_{12}$  foi o que apresentou os menores erros de previsão e pode ser recomendado para efetuar previsões.*

**Palavras-chave:** Séries Temporais. Linguagem de Programação R. Commodity. Probabilidade. Preços do Café.

## 1 INTRODUCTION

Coffee is one of the most valuable commodities in the world, and approximately 70% of the global production is traded internationally (ICO 2020; Paiva *et al.* 2018). According to the International Coffee Organization (2021), this is a growing market, with global production volume having increased by over 60% since the 1990s and now holding an estimated value of over USD 200 billion annually. The coffee's notable participation in the global economy reflects the influence of the commodity on the economies of various emerging producing countries. In countries with minimal development, exports can contribute to foreign exchange, highlighting the impact on economic and political structure, as well as the risks of collapse if a permanent price decline occurs (Sá Barreto; Zugaib, 2016).

Brazil is the world's largest producer and exporter of coffee, and the coffee activity has a significant role in the country's economy and national agricultural production (Carrasco-Gutierrez, Almeida 2013; CONAB 2022; Mehta, Chavas 2008). According to the MAPA (2022), coffee moved USD 6.4 billion in 2021, accounting for 8.28% of the country's exports. However, the high market volatility causes insecurity for coffee growers, as fluctuations impact the price of the final product and the economic returns of the activity (Arêdes; Pereira; Teixeira, 2009; Sá Barreto; Zugaib, 2016).

Being a commodity, coffee is one of the agricultural products that is primarily traded in its raw or semi-processed form. Globally, its price is standardized with quotations and negotiations conducted through commodity exchanges (Albieri; Terra,

2022). According to Sá Barreto & Zugaib (2016), price fluctuations are primarily linked to supply and demand movements, as experienced throughout history. An excess of supply leads to price decreases, while scarcity results in price increases. Nevertheless, there are other non-systemic and market-related factors that influence this determination and shape production cycles. The authors mention among market agents the actions of investment funds, exchange rates, market speculation, interest rates, economic crises, and the existence of possible others. For non-systemic agents, climate changes were highlighted, a factor closely related to production.

Furthermore, the coffee market is characterized by oligopsony, where the supply comes from many producers, but the demand is concentrated among a small number of buyers (Ghoshray; Mohan, 2021; Sá Barreto; Zugaib, 2016). This market concentration enhances the influence of coffee buyers from importing countries on prices, while diminishing such capability in producing countries (Ghoshray; Mohan, 2021). Another important factor for price dynamics is that the activity was impacted by the Covid-19 pandemic; the uncertainties of the period affected both supply and demand in the sector while causing disruptions throughout the production chain (ICO, 2020). Currently, coffee production is being influenced by the military conflict involving Russia and Ukraine. In 2020, these countries accounted for 3.85% of global coffee consumption. Russia supplies 20% of the ammonia transported by sea in the global market, and the tension caused an increase in oil prices (ICO, 2022).

Understanding the dynamics of the coffee market enables better risk management, allowing producers to make informed decisions in cultivation and position their product in the market with optimal pricing and profitability (Carrasco-Gutierrez, Almeida 2013). This understanding can also support the formulation of political and economic strategies by governments of exporting and importing countries, cooperatives, futures market participants, roasters, and others (Carrasco-Gutierrez, Almeida 2013; Ghoshray; Mohan, 2021; Sá Barreto; Zugaib, 2016). The application of statistical and econometric models in studying the dynamics of the sector is widely spread. The employed model will vary depending on the intended objective. For example, if the purpose is to understand the causal structure of the market, it is possible to estimate and test structural equations. However, if the goal is to test hypotheses among variable correlations, forecasting, or even formulating

policies, it is possible to perform partial or complete sector modeling (Carrasco-Gutierrez, Almeida 2013).

Seeking to comprehend the dynamics of the international coffee market and the determinants influencing price formation, Sá Barreto & Zugaib (2016) utilized a Three-Stage Least Squares simultaneous equations model and chi-squared tests to measure the acceptability of the results. Mehta & Chavas (2008) based themselves on reduced equations of price determination in vector autoregressive (VAR) to evaluate the evolution of commodity prices. Bressan (2004), Carrasco-Gutierrez & Almeida (2013) and many others explored time series forecasting models in search of their results for coffee. During the formulation of this work, it can be observed the ease in finding works that used the ARMA and ARIMA models.

The use of time series models or dynamic models in the study of coffee prices can be widely explored. A time series can be modeled by different methods, differentiated by complexity. Its application starts from the presence of a persistent or systematic pattern in the behavior of the variable. Among the most employed models, we can highlight those of Box & Jenkins, Neural Networks, exponential smoothing, moving average, Bayesian, and structural (Bressan, 2004).

Given the current market scenario, its significance, and the disruptions caused by the Covid-19 pandemic and the conflicts between Ukraine and Russia, understanding the dynamics of coffee prices is of substantial importance. In the present study, this understanding will be achieved through probabilistic models, enabling the development of strategies by all market agents involved, within the context of price volatility and its related consequences. To achieve the proposed objective, the following steps were taken: i) the best probabilistic models were fitted to each monthly price series according to their occurrence frequency; and ii) the probabilities of occurrence for average coffee bean prices were calculated according to levels of practical interest. iii) Investigated the occurrence of seasonality in coffee market price series; iv) Determined the trend in the temporal price series of the activity; and v) Proposed the best temporal model for short-term forecasting.

## 2 MATERIALS AND METHODS

### 2.1 Data

The series of historical data obtained by COOXUPÉ for 41 years was analyzed, covering the period from January 1981 to December 2022, consisting of records of coffee bag prices at the national level was followed. The time series consist of monthly information about the average price of coffee bags traded at COOXUPÉ in dollars. It is worth noting that COOXUPÉ's arabica coffee production is exported to 49 countries across five continents. In the coffee export ranking, the cooperative has been the top exporter among operating companies in Brazil for seven consecutive years. Exports account for more than 60% of COOXUPÉ's activities (COOXUPÉ, 2022).

In accordance with the proposed objectives, the data series will be organized in two ways:

a) The historical data series was divided into monthly subseries, totaling twelve subseries, for probabilistic modeling. Descriptive statistics were calculated for the data, encompassing measures such as maximum value, minimum value, mean, median, standard deviation, coefficient of variation, and skewness coefficient.

b) The historical data series was analyzed as it is originally, maintaining its chronological character, for temporal modeling.

### 2.2 Probabilistic modeling of the price of coffee bags

The probability distributions used in this study will be presented below. Let  $x$  be the random variable associated with the price of a coffee sack (in dollars), the Gamma distribution has a probability density function (Eq. 1) given by:

$$f(x) = \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} e^{-\frac{x}{\beta}} \quad (1)$$

defined in  $0 < x < \infty$ ,  $\beta$  is the scale parameter,  $\alpha$  is the shape parameter, and the gamma function is defined by  $\Gamma(\alpha) = \int_0^{\infty} x^{\alpha-1} e^{-x} dx$ . The probability density function of the Log-Normal distribution (Eq. 2) is given by:

$$f(x) = \frac{1}{\sigma x \sqrt{2\pi}} e^{-\frac{[\ln(x)-\mu]^2}{2\sigma^2}} \quad (2)$$

where  $x > 0$ ,  $\ln$ ,  $\ln$  represents the neperian logarithm,  $\mu$  is the mean of the logarithms of variable  $x$ , and  $\sigma$  is the standard deviation of the logarithms of variable  $x$ . The Normal distribution has a probability density function (Eq. 3) given by:

$$f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2} \quad (3)$$

where

$\mu$  is the mean and  $\sigma$  is the standard deviation of the random variable (Devore, 2018).

The probabilistic models in equations 1 and 2 have asymmetric nature, and the one in equation 3 has symmetric nature. The parameters of these probabilistic models are initially unknown. To determine them, some estimation method is required. For this purpose, the maximum likelihood method was employed, which involves determining the values of the parameters  $\alpha$ ,  $\beta$ ,  $\mu$  and  $\sigma$  that maximize the probability of the sample having occurred, under the condition of independence. Further details about the method can be found in Liska *et al.* (2020).

After the proposal of a probabilistic model and the estimation of its parameters, it is necessary to assess whether the proposed model fits the data. This is done through a

goodness-of-fit test, and there are several in the literature, such as the Chi-Square test, Kolmogorov-Smirnov test, Anderson-Darling test, among others. The Chi-Square test (Eq. 4) compares the observed and expected cumulative frequencies using the test statistic:

$$Q = \sum_{i=1}^k \frac{(Fo_i - Fe_i)^2}{Fe_i} \quad (4)$$

where  $k$  is the number of classes,  $Fo_i$  is the observed cumulative frequency of the  $i$ -th class, and  $Fe_i$  is the expected frequency of the  $i$ -th class according to the distribution being tested. The  $Q$  statistic is evaluated using the chi-square distribution with  $k-1$  degrees of freedom, and if  $Q > \chi^2_{\alpha}(k-1)$ , the null hypothesis of the test, which states that the proposed model fits the data, is rejected (Levin; Fox, 2004). Similarly, there is the Kolmogorov-Smirnov test (KS), which is based on analyzing the proximity or fit between the sample distribution function and the population distribution function under the null hypothesis (Reis; Beijo; Avelar, 2017). In the KS test, the hypotheses of interest are given by  $H_0$ : The distribution function from which the sample originates follows the assumed known distribution function. Then, the test statistic value should be compared with the critical value (using tables), for the test's significance level. In both tests, the null hypothesis is also rejected if the p-value is smaller than the adopted significance level.

Regarding the verification of the assumption of independence and randomness of observations, Ljung-Box tests were employed due to the maximum likelihood estimators. The test statistic is compared with the quantile of the chi-square distribution with one degree of freedom at the adopted significance level. The Mann-Kendall test was also used (Salviano; Groppo; Pellegrino, 2016). The Ljung-Box test assumes the null hypothesis of independence of observations, while the Mann-Kendall test assumes the absence of trend. In both tests, if the p-value is smaller than the adopted significance level, the null hypothesis is rejected. For all tests, a significance level  $\alpha$  of 1% will be used.

After confirming the fit of a probabilistic model, one can proceed with the calculation of probabilities and/or quantiles of practical interest. The probability of a coffee sack's average price occurring up to a hypothetical value  $y$  is calculated using the

cumulative distribution function ( $F(y)$ ) (Eq. 5), which means that:

$$F(y) = P[X \leq y] = \int_{-\infty}^y f(x)dx, \quad (5)$$

in other words, the probability of an average price occurring up to a quantity  $Y$  is given by the integral solution of any of the probability density functions presented earlier. Therefore, the probability of an average coffee sack price exceeding a value  $Y$  is given by (Eq. 6):

$$P[X > y] = 1 - F(y). \quad (6)$$

The calculation of quantiles of practical interest is obtained by taking the inverse of the cumulative distribution function given the probability. Further details can be found in Casella & Berger (2001).

### 2.3 Temporal modeling of the price of coffee bags

Regarding time modeling, it should be considered that the random variable average coffee sack price, or time series denoted by  $Z_t$ , with  $t = 1, \dots, n$ , can be decomposed into the sum  $Z_t = T_t + S_t + a_t$ , where: the component  $T_t$  explains the trend of the series, i.e., the increase or decrease in observed values according to a behavior that can be polynomial or exponential; the component  $S_t$  explains the seasonality of the series, i.e., phenomena that occur regularly (daily, weekly, monthly, etc.); and the random component or error  $a_t$ . The usual assumption is that  $a_t$  is a purely random series or independent white noise, with zero mean and constant variance (Metcalf; Cowpertwait, 2009).

The existence of a trend was tested using the sign test, proposed by Cox and Stuart (1955); seasonality, particularly periodicity, was tested using the Fisher test, initially described by Fisher (1929), and extended by Whittle (1952).

According to the recommendations of Morettin & Tolo (2006), it is advisable to remove the seasonal component of the series,  $S_t$ , before testing for the existence of a

trend, and eliminate the trend component of the series,  $T_t$ , before testing for the presence of seasonality. Morettin & Toloï (2006) describe some useful procedures for removing trend and seasonality from a series, such as differencing and polynomial fitting.

The verification of whether the random component  $a_t$  is white noise will be done using the test proposed by Ljung & Box (1978). In this work, models from the SARIMA class (Seasonal ARIMA) as described by Box, Jenkins & Reinsel (2008) will be fitted, and the assumption of white noise for the component  $a_t$  was verified through the Ljung-Box test, as presented earlier.

The  $SARIMA(p, d, q)(P, D, Q)_s$  model considers the trend and seasonality of the series (Eq. 7) and can be represented by

$$\phi_p(B)\Phi_P(B^s)(1-B)^d(1-B^s)^D Z_t = \theta_q(B)\Theta_Q(B^s)a_t \quad (7)$$

where  $B$  is the lag operator, defined as  $B^j Z_t = Z_{t-j}$ ;  $\phi_p(B) = 1 - \phi_1 B - \dots - \phi_p B^p$  is the autoregressive polynomial of order  $p$ ;  $\theta_q(B) = 1 - \theta_1 B - \dots - \theta_q B^q$  is the moving average polynomial of order  $q$ ;  $\Phi_P(B) = 1 - \Phi_1 B^s - \dots - \Phi_P B^{Ps}$  is the seasonal autoregressive polynomial of order  $P$ ;  $\Theta_Q(B^s) = 1 - \Theta_1 B^s - \dots - \Theta_Q B^{Qs}$  is the seasonal moving average polynomial of order  $Q$ ;  $d$  is the number of differencing operations needed to remove the trend from the series;  $D$  is the number of lag differences needed to remove stochastic seasonality from the series.

The fitted models will be compared using the mean squared prediction error (MSE), mean absolute percentage error (MAPE), and Akaike Information Criterion (AIC). These will be calculated based on the last year of the series, referred to as the test series. Thus, the models were fitted using 40 years of data and tested on the last year, 2022. For these criteria, the most suitable time series model is the one that yields the lowest values (Liska *et al.*, 2013).

Concluding the methodology, the statistical computational system R (R Core Team, 2021) and package TSA was used for conducting the analyses.

### 3 RESULTS AND DISCUSSION

The results shown in Table 1 present some descriptive measures related to the data rearranged in monthly series, as mentioned in "a" of the data subsection. It is possible to observe that the highest values recorded for the maximum of average prices per sack occurred in the months of April, May, and January, respectively, while the lowest values recorded for the minimums were in the months of October, June, and July.

Table 1. Descriptive measures of the monthly average prices of coffee sacks (in USD) traded at COOXUPÉ from January 1981 to December 2021

Month	Minimum	Maximum	Median	Mean	Standard Deviation	Coefficient of Variation	Coefficient of Asymmetry
January	43.22	309.97	119.57	122.1680	57.9569	0.4744	1.2716
February	43.08	284.03	112.90	122.7024	57.1826	0.4660	1.1322
March	47.59	297.90	113.78	121.6237	55.0155	0.4523	1.1061
April	48.52	319.37	114.37	118.7402	54.7299	0.4609	1.2856
May	39.34	310.14	115.01	117.5605	54.2539	0.4615	1.1460
June	35.21	305.53	117.46	114.4441	52.6483	0.4600	1.1270
July	35.63	284.72	101.41	113.5027	54.6804	0.4818	0.8997
August	35.78	277.05	105.52	113.3934	54.8216	0.4835	0.7559
September	36.06	261.90	101.87	111.9978	54.1227	0.4832	0.7650
October	35.04	259.27	105.35	110.8412	52.6736	0.4752	0.7611
November	35.74	267.47	107.79	113.7805	54.1629	0.4760	0.7965
December	39.03	259.41	112.51	118.6302	55.9886	0.4720	0.8561

The highest standard deviations of the commodity were observed in the months of January, February, and December, in that order. This measure of data dispersion is traditionally used in determining price volatility and considered a risk measure by some market speculators (Capitani; Mattos, 2017).

Among the monthly subseries, the months of February, January, and March, respectively, had the highest averages, while the lowest averages occurred for the months of October, September, and August. These values may be correlated with the phenological and physiological cycle of the coffee plant, combined with market reactions. According to Sá Barreto and Zugaib (2016), one of the main factors contributing to price instability is production and supply situation.

Camargo and Camargo (2001) outlined the phenology of coffee plants (*Coffea arabica* L.) for Brazilian tropical conditions and noted in their work that during the

flowering period, which occurs from September to December of the first year and in months preceding the rise in average prices, the incidence of high temperatures and an intense water deficit at its beginning leads to flower abortion. Subsequently to fertilization, there are the stages of "chumbinhos" (small green berries) and expansion. During these stages, a severe drought can hinder fruit growth and result in a low-quality harvest. In the "granação" stage, which occurs from January to March of the second year, corresponding to months with higher average prices, water stress leads to the production of poorly formed fruits and "chochamento" (poor development) of beans (Camargo; Camargo, 2001). Coffee harvesting takes place during the dry months, typically between March and April to September, a period when average prices tend to decline.

What is more, Silveira, Mattos, and Saes (2016) discuss the importance of crop reports and their impacts on the market and future price volatility. This is a significant factor considering that coffee trade mainly occurs in commodity exchanges and futures markets. Reports released by the International Coffee Organization, especially during flowering periods, have a significant impact on prices. Similar effects, albeit on a smaller scale, are observed in studies published by CONAB in September and December.

From Table 2, it can be observed that the Gamma and log-Normal distributions fitted the coffee sack price data for all months according to the Kolmogorov-Smirnov and  $\chi^2$  tests at a 1% significance level. Given the aforementioned and considering that both distributions are asymmetric, a similar result was found for the price of "boi gordo" (cattle) in the state of São Paulo by Lucca Filho *et al.* (2022).

Table 2. Parameter estimates for the three different tested: Gamma, Normal, and log-Normal, along with the corresponding results of the Kolmogorov-Smirnov (KS) and Chi-Square ( $\chi^2$ ) tests, and the coefficient of determination ( $R^2$ ) values for the monthly subseries of coffee sack average prices in dollars

Month	Distributions	Parameter				KS	$\chi^2$	$R^2$
		$\alpha$	$\beta$	$\mu$	$\sigma$			
January	Gamma	5.1025	0.0418	-	-	0.7621	0.4809	0.9930
	Normal	-	-	122.1680	57.2457	0.2182	0.0141	0.9786
	log-Normal	-	-	4.7042	0.4503	0.5859	0.6971	0.9958
February	Gamma	5.1966	0.0424	-	-	0.9727	0.3269	0.9952
	Normal	-	-	122.7024	56.4809	0.4490	0.0142	0.9827
	log-Normal	-	-	4.7105	0.4468	0.9431	0.4958	0.9968
	Gamma	5.4656	0.0449	-	-	0.9842	0.7929	0.9983

March	Normal	-	-	121.6237	54.3405	0.6956	0.0616	0.9897
	log-Normal	-	-	4.7067	0.4355	0.9855	0.9280	0.9989
April	Gamma	5.3158	0.0448	-	-	0.9842	0.3646	0.9986
	Normal	-	-	118.7402	54.0584	0.8002	<0.0001	0.9941
	log-Normal	-	-	4.6799	0.4424	0.9744	0.6621	0.9979
May	Gamma	5.1290	0.0436	-	-	0.9693	0.3277	0.9959
	Normal	-	-	117.5605	53.5882	0.7078	0.0002	0.9952
	log-Normal	-	-	4.6663	0.4545	0.9194	0.3869	0.9931
June	Gamma	5.0585	0.0442	-	-	0.8259	0.2330	0.9920
	Normal	-	-	114.4441	52.0023	0.9539	<0.0001	0.9982
	log-Normal	-	-	4.6380	0.4611	0.5389	0.2180	0.9862
July	Gamma	4.6162	0.0407	-	-	0.9163	0.6274	0.9982
	Normal	-	-	113.5027	54.0095	0.5368	0.1535	0.9952
	log-Normal	-	-	4.6196	0.4800	0.9103	0.5354	0.9959
August	Gamma	4.4124	0.0389	-	-	0.8566	0.6443	0.9953
	Normal	-	-	113.3934	54.1489	0.4356	0.3332	0.9943
	log-Normal	-	-	4.6133	0.4960	0.9411	0.4780	0.9916
September	Gamma	4.5023	0.0402	-	-	0.8183	0.9018	0.9970
	Normal	-	-	111.9978	53.4586	0.4237	0.4796	0.9909
	log-Normal	-	-	4.6033	0.4878	0.9810	0.8889	0.9959
October	Gamma	4.6731	0.0422	-	-	0.8393	0.9972	0.9997
	Normal	-	-	110.8412	52.0273	0.5251	0.7504	0.9954
	log-Normal	-	-	4.5973	0.4777	0.9195	0.9778	0.9976
November	Gamma	4.6657	0.0410	-	-	0.8642	0.7423	0.9960
	Normal	-	-	113.7805	53.4983	0.4487	0.6746	0.9977
	log-Normal	-	-	4.6233	0.4781	0.6491	0.5056	0.9909
December	Gamma	4.7281	0.0399	-	-	0.9545	0.2459	0.9963
	Normal	-	-	118.6302	55.3016	0.7660	0.0356	0.9934
	log-Normal	-	-	4.6665	0.4773	0.6990	0.2662	0.9928

When analyzing the calculated skewness coefficients for the monthly subseries, a positive skewness is observed in all months. Therefore, asymmetric distributions are the most suitable for monthly modeling. Comparatively, the Gamma distribution performed better in most months, except for January, February, and March, where the log-Normal distribution stood out (Figure 1). The predilection for a particular distribution in some months arises from a slightly better fit. Both goodness-of-fit tests have their limitations, and in some cases, the KS and  $\chi^2$  tests selected different distributions, making the coefficient of determination a third criterion for decision.

Figure 1. Histograms of the monthly average prices of coffee sacks (in USD) from January 1981 to December 2021, along with graphs of the probability density functions of the fitted Gamma, Normal, and log-Normal distributions



In Table 3, it is possible to observe the probabilities of commodity price occurrences and the quantiles of 1 and 5%. The months of February, January, and March, respectively, showed the highest probabilities for the occurrence of values surpassing 200 dollars. Among all months, these are also the months with the highest likelihood of prices exceeding 400 dollars. On the other hand, among all the months, October has the lowest probability of values exceeding 100 and/or 200 dollars. This trend is like what occurred in the years 2018 and 2019, driven in 2018 by the completion of the harvest and the confirmation of Brazil's record production of 59.4 million sacks, coupled with global production exceeding consumption (CONAB, 2018, 2019).

Table 3. Probabilities (in %) of the occurrence of average prices of coffee sacks (in USD) and quantiles of practical interest for all months of the year

Month	Distributions	\$50	\$100	\$200	\$400	q1%	q5%
January	Gamma	94.5246	61.3014	8.7764	0.0267	281.5918	222.5364
	Normal	89.6286	65.0712	8.6976	<0.0001	255.3415	216.3289
	log-Normal	96.0721	58.7041	9.3546	0.2129	314.7760	231.5876
February	Gamma	94.7970	61.8613	8.8067	0.0253	281.0805	222.5099
	Normal	90.0988	65.6139	8.5568	<0.0001	254.0967	215.6053
	log-Normal	96.3019	59.3149	9.4165	0.2074	314.1896	231.7066
March	Gamma	95.1471	61.7098	8.0074	0.0165	273.9450	217.8699
	Normal	90.6257	65.4659	7.4606	<0.0001	248.0385	211.0058
	log-Normal	96.5984	59.2151	8.7131	0.1587	304.8225	226.5483
April	Gamma	94.4098	59.4013	7.2877	0.0141	269.9406	214.1390
	Normal	89.8241	63.5579	6.6396	<0.0001	244.4988	207.6584
	log-Normal	95.8700	56.7105	8.1089	0.1516	301.6041	223.1005
May	Gamma	93.8122	58.2030	7.2022	0.0156	270.4962	213.8708
	Normal	89.6298	62.8428	6.1977	<0.0001	242.2253	205.7053
	log-Normal	95.1485	55.3501	8.2203	0.1777	306.0496	224.5225
June	Gamma	93.0573	55.8175	6.3576	0.0118	264.5657	208.9120
	Normal	89.2375	60.9402	4.9961	<0.0001	235.4195	199.9803
	log-Normal	94.2298	52.8378	7.6075	0.1667	302.0924	220.6284
July	Gamma	91.7096	54.3430	6.7890	0.0190	270.8028	211.9908
	Normal	88.0157	59.8709	5.4631	<0.0001	239.1475	202.3404
	log-Normal	92.9801	51.2009	7.8666	0.2130	309.8698	223.4233
August	Gamma	91.0826	53.8866	7.1112	0.0248	274.8801	214.2443
	Normal	87.9145	59.7679	5.4864	<0.0001	239.3627	202.4605
	log-Normal	92.1276	50.6528	8.3643	0.2732	319.6626	227.9697
September	Gamma	91.0298	53.0576	6.5359	0.0187	269.5674	210.5156
	Normal	87.6922	58.8789	4.9864	<0.0001	236.3611	199.9293
	log-Normal	92.1783	49.8498	7.7123	0.2216	310.4908	222.6750
October	Gamma	91.2549	52.4950	5.9210	0.0129	263.3248	206.3791
	Normal	87.8881	58.2532	4.3293	<0.0001	231.8748	196.4185
	log-Normal	92.4283	49.3431	7.1131	0.1759	301.4623	217.6922
November	Gamma	91.9100	54.6325	6.7890	0.0184	270.4559	211.9361
	Normal	88.3408	60.1637	5.3521	<0.0001	238.2361	201.7773
	log-Normal	93.1580	51.5120	7.8999	0.2108	309.6857	223.5701
	Gamma	93.0576	58.1271	8.2168	0.0289	280.6845	220.2307

December	Normal	89.2700	63.1898	7.0594	<0.0001	247.2810	209.5933
	log-Normal	94.3055	55.1165	9.2795	0.2751	322.7333	233.1250

The volatility among prices and monthly variations can be explained by the low short-term elasticity of demand and supply in the coffee market (Sá Barreto; Zugaib, 2016; Silveira; Mattos; Saes, 2016), in addition to other factors such as dependency on exchange rates (Carrasco-Gutierrez, Almeida 2013), macroeconomic, financial, climatic, and meteorological variables (Cuaresma; Hlouskova; Obersteiner, 2018).

Storing coffee beans and selling during the off-season is one of the possible strategies for coffee farmers and cooperatives to achieve higher profits and revenues. Administratively, product storage becomes viable when the difference between the expected price and the current price is greater than the sum of the interest rate and the storage cost. It becomes attractive when it can compensate for its associated risks (Arêdes *et al.*, 2008).

Agreement to Arêdes *et al* (2008), it is important to assess the profitability and risk of the operation. In their work, they used indicators such as Payback Period (PP), Net Present Value (NPV), Internal Rate of Return (IRR), and the Average Cost Indicator (CMe), calculated based on the cash flow of production, to make decisions regarding the investments in coffee storage adoption. The authors confirm the economic viability of storage and pointed out that the months of January to March, which have a higher probability of higher prices per sack, are better for commercialization. Conversely, the least favorable months for commercialization are from June to November, as they have the lowest probabilities for values exceeding 100 dollars.

Regarding time series modeling, in Figure 2, we can observe the behavior of the historical data series from January 1981 to December 2022. Upon analysis, it's evident that the series is non-stationary, exhibiting recurring peaks over the years and fluctuation of data around an inclined line. The conducted tests confirm both the trend and seasonality, as the null hypotheses of the Cox-Stuart sign test for trend and Fisher test for seasonality were rejected at a 1% significance level.

Figure 2. Graph of the monthly series of average prices for bags of coffee (in USD), from January 1981 to December 2022

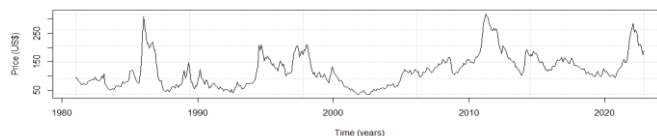


Table 4 presents fifteen different models adjusted for the historical series of average coffee bag prices. Considering the additivity of the model, the results of the Ljung-Box test, AIC, MSE, and MAPE (Table 4), as well as the sample autocorrelation functions, sample partial autocorrelations, and residuals, the model  $SARIMA(3,1,3)(2,1,1)_{12}$  is the best fit for the data. The estimated parameters along with their standard errors for the fitted models are provided in Table 5. Unfortunately, the preferred proposed model did not allow for the estimation of the standard error. Alternatively,  $SARIMA(2,1,2)(0,1,2)_{12}$  can be used in cases where the calculation of this measure of uncertainty is indispensable.

Table 4. Different viable SARIMA models for the historical series of average coffee bag prices (in USD), followed by their respective results for the Ljung-Box test, Akaike information criterion (AIC), mean squared error of prediction (MSE) and mean absolute percentage error (MAPE)

Models	Ljung-Box	AIC	MSE	MAPE
$SARIMA(1,1,3)(0,1,1)_1$	0.4426	3836.42	2253.9050	18.6269
$SARIMA(2,1,2)(0,1,2)_1$	0.8455	3831.79	1128.5710	13.2959
$SARIMA(2,1,2)(1,0,0)_1$	0.9040	3890.23	1274.6500	14.2880
$SARIMA(2,1,2)(1,1,1)_1$	0.8380	3831.88	1142.1110	13.3622
$SARIMA(2,1,2)(1,1,2)_1$	0.8750	3831.57	1257.8320	13.7722
$SARIMA(2,1,2)(1,1,3)_1$	0.8950	3834.93	1171.0870	13.4639
$SARIMA(2,1,2)(2,1,1)_1$	0.8855	3833.27	1153.5880	13.3771
$SARIMA(2,1,3)(1,1,2)_1$	0.8499	3832.93	1360.8840	14.1603
$SARIMA(2,1,4)(0,1,2)_1$	0.9567	3830.74	1750.9680	15.9546
$SARIMA(2,1,4)(0,1,3)_1$	0.9344	3831.26	1694.3620	15.6544
$SARIMA(3,1,2)(0,1,1)_1$	0.8258	3832.55	1591.5230	15.3001
$SARIMA(3,1,3)(2,1,1)_1$	0.9075	3836.65	1097.0360	13.1254
$SARIMA(3,1,5)(0,1,3)_1$	0.9389	3835.03	1671.8230	15.5614
$SARIMA(2,1,3)(0,2,3)_2$	0.6585	3736.67	1618.3960	15.6974
$SARIMA(2,1,3)(1,1,2)_2$	0.6355	3759.28	1415.5740	14.7789

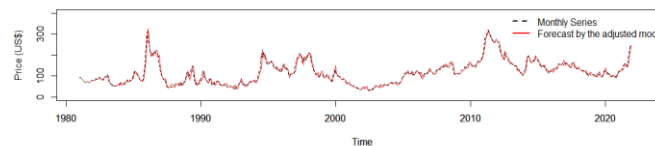
Table 5. Estimates of the parameters and their respective standard errors in parentheses for the models  $SARIMA(3,1,3)(2,1,1)_{12}$  and  $SAIMA(2,1,2)(0,1,2)_{12}$

Model	Parameter									
	Ar1	Ar2	Ar3	Ma1	Ma2	Ma3	Sar1	Sar 2	Sma1	Sma2
$SARIMA(3,1,3)(2,1,1)_{12}$	0.2313 NA	0.5398 NA	-0.6094 NA	0.0351 NA	-0.4433 NA	0.4295 NA	-0.0641 (0.0481)	-0.0433 (0.0367)	-0.9970 (0.0644)	- -
$SARIMA(2,1,2)(0,1,2)_{12}$	1.1922 (0.1722)	-0.6234 (0.1211)	- -	-0.9195 (0.1902)	0.4515 (0.1148)	- -	- -	- -	-1.0550 (0.0737)	0.0550 (0.0490)

NA: Not evaluated; Ar: Autoregressive parameter; Ma: Moving average parameter; Sar: Seasonal autoregressive parameter; Sma: Seasonal moving average paramete.

The fitting of the  $SARIMA(3,1,3)(2,1,1)_{12}$  model can be observed graphically in Figure 3. According to the proposed model, there is a relationship between the current price of coffee bags and the values that occurred three months ago. In order to remove the trend from the series, differencing was applied, and to improve the model fit, a three-lag moving average was applied. Additionally, seasonality indicates that the current price is also influenced by the value that occurred two years ago. Seasonality of the series was removed using differencing and applying a one-year lag for the seasonal moving average.

Figure 3. Monthly series graph of average coffee bag prices (in USD), period from January 1981 to December 2021, and of the adjustment of the  $SARIMA(3,1,3)(2,1,1)_{12}$  model



The relationship between the current price and the value from two years ago demonstrates the congruence of the model with the biennial cycle of coffee production, inherent to coffee plants. Bienniality refers to the alternate production of years with high and low productivity, due to the simultaneous occurrence of vegetative and reproductive functions in the same year. The coffee plant cannot produce sufficient reserves for fruiting and growth simultaneously, so in one year, the reserves are used for fruiting, increasing productivity. However, this results in low branch growth, consequently leading to reduced production in the following year (Bacha, 1998).

According to Silva and Reis (2013), the biennial cycle has a significant impact on coffee supply, which can lead to price instability in the market, causing difficulties for the country's coffee policy and producers in general. Hence, planning for storage and carrying over from one harvest to another is required to maintain profitability. Upon analyzing historical production and the series of average prices per sack, it is evident that the years of low harvests had the highest price averages. If feasible, it would be beneficial for the COOXUPÉ cooperative to adopt this and other strategies for increased profitability.

Through the developed model, it was possible to forecast future values of the series, as shown in Table 6. According to the test for population correlation, at the 1% significance level, the null hypothesis that there is no population correlation between the predicted average prices and the observed prices for the year 2022 was rejected ( $p$ -value = 0.0049). The estimate for the correlation coefficient was 75.1%, indicating a good agreement between the predictions and the observed prices.

Table 6. Prediction of average coffee bag prices for the year 2022, return time of 1 year, via time series model  $SARIMA(3,1,3)(2,1,1)_{12}$

Month	Observed value	Predicted value	Difference
January	268.93	268.59	0.34
February	285.44	266.43	19.01
March	252.79	258.41	-5.62
April	263.19	249.41	13.78
May	254.56	243.84	10.72
June	248.68	240.30	8.38
July	208.65	241.55	-32.90
August	209.34	244.00	-34.66
September	210.58	242.89	-32.31
October	196.61	240.04	-43.43
November	174.62	240.24	-65.62
December	189.21	242.84	-53.63

In practical terms, when the difference between predicted and observed values is positive, the cooperative achieves higher profits than expected by the model. Conversely, in months where this difference is negative, there is a loss in profits compared to the predicted potential. The months from January to June show the

smallest differences between predicted and observed values, indicating that the prices set by COOXUPÉ are close to those projected by the model. From June to December, a larger difference can be observed compared to the first six months of the year, suggesting that the model overestimated the prices, and the cooperative could have made higher profits from the sale of the sacks.

#### 4 CONCLUSIONS

The Gamma and log-Normal distributions fitted the monthly data of average coffee sack prices. The Log-Normal distribution is preferred for the probabilistic study of the variable in the months of January, February, and March, while the Gamma distribution is more suitable for the remaining months. February, January, and March are the months with the highest probability of higher average values and are therefore the most recommended for coffee trading.

Regarding the time series model, the  $SARIMA(3,1,3)(2,1,1)_{12}$  obtained the best values for MSE and MAPE, making it the most suitable model for making predictions. The practical application of the models is efficient and can be a great asset in price negotiations and decision-making for market agents.

#### ACKNOWLEDGMENTS

Acknowledgments to the Federal University of São Carlos - UFSCar, Center for Agricultural Sciences, the National Council for Scientific and Technological Development - CNPq for granting the Undergraduate Research Scholarship and COOXUPÉ for providing the data.

#### REFERENCES

- ALBIERI, A. C. DE S.; TERRA, L. A. A. Estratégias de diferenciação da commodity de café. **Future Studies Research Journal: Trends and Strategies**, v. 14, n. 1, p. 1–27, 20 nov. 2022.
- ARÊDES, A. F. DE *et al.* Rentabilidade e Risco na Estocagem do Café pelos Produtores na Região de Viçosa-MG. **Revista de Economia e Agronegócio**, v. 6, n. 2, p. 235, 2008.

ARÊDES, A. F. DE; PEREIRA, M. W. G.; TEIXEIRA, E. C. **Previsão De Preço Do Quilo Do Café Arábica: Uma Aplicação Dos Modelos Arima E Garch**. [s.l.: s.n.].

BACHA, C. J. C. A cafeicultura brasileira nas décadas de 80 a 90 e suas perspectivas (1998). **Preços Agrícolas**, v. 12, n. 142, p. 14–22, ago. 1998.

BOX, G. E. P.; JENKINS, G. M.; REINSEL, G. C. Time series analysis : forecasting and control. p. 746, 2008.

BRESSAN, A. A. Tomada De Decisão Em Futuros Agropecuários Com Modelos De Previsão De Séries Temporais. **RAE-eletrônica**, jun. 2004.

CAMARGO, Â. P. DE; CAMARGO, M. B. P. DE. Definição e esquematização das fases fenológicas do cafeeiro arábica nas condições tropicais do Brasil. **Agrometeorologia**, v. 60, n. 1, p. 65–68, 2001.

CAPITANI, D. H. D.; MATTOS, F. Measurement of commodity price risk: An overview of Brazilian agricultural markets. **Revista de Economia e Sociologia Rural**, v. 55, n. 3, p. 515–532, 2017.

CARRASCO-GUTIERREZ, C. E.; ALMEIDA, F. M. DE M. Modelagem e Previsão do Preço do Café Brasileiro. **Revista de Economia**, p. 7–27, 2013.

CASELLA, G.; BERGER, R. R. **Statistical Inference**. 2. ed. [s.l.] Thomson Learning, 2001.

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Análise Mensal Café - Outubro de 2018**. Brasília: [s.n.].

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Análise Mensal Café - Outubro/Novembro de 2019**. Brasília: [s.n.].

CONAB - COMPANHIA NACIONAL DE ABASTECIMENTO. **Acompanhamento da Safra Brasileira - Safra 2022**. Brasília, DF: [s.n.]. Disponível em: <file:///C:/Users/Admin/Downloads/E-book\_BoletimZdeZSafrazcafZ-ZjaneiroZ22.pdf>. Acesso em: 27 abr. 2022.

COOXUPÉ. **Cooxupé Exportação**. Disponível em: <<https://www.cooxupe.com.br/exportacao/exportacao/>>. Acesso em: 6 jun. 2022.

COX, D. R.; STUART, A. Some Quick Sign Tests for Trend in Location and Dispersion. **Biometrika**, v. 42, n. 1/2, p. 80, jun. 1955.

CUARESMA, J. C.; HLOUSKOVA, J.; OBERSTEINER, M. Fundamentals, speculation or macroeconomic conditions? Modelling and forecasting Arabica coffee prices. **European Review of Agricultural Economics**, v. 45, n. 4, p. 583–615, 2018.

DEVORE, J. L. **Probabilidade e estatística para engenharia e ciências**. 9. ed. [s.l.] Cengage, 2018.

FISHER, R. A. Tests of significance in harmonic analysis. **Proceedings of the Royal Society of London. Series A, Containing Papers of a Mathematical and Physical Character**, v. 125, n. 796, p. 54–59, ago. 1929.

- GHOSHRAJ, A.; MOHAN, S. Coffee price dynamics: An analysis of the retail-international price margin. **European Review of Agricultural Economics**, v. 48, n. 4, p. 983–1006, 1 set. 2021.
- ICO - INTERNATIONAL COFFEE ORGANIZATION. Volatilidade Dos Preços Do Café: O Covid-19 E Os Fatores Fundamentais Do Mercado. maio 2020.
- ICO - INTERNATIONAL COFFEE ORGANIZATION. Visão Geral do Relatório sobre o Desenvolvimento do Café da OIC de 2020. fev. 2021.
- ICO - INTERNATIONAL COFFEE ORGANIZATION. Relatório sobre o mercado cafeeiro fevereiro 2022. p. 1–10, fev. 2022.
- LEVIN, J.; FOX, J. A. **Estatísticas para ciências humanas**. 9. ed. São Paulo: Prendice Hall, 2004.
- LISKA, G. R. *et al.* Estimativas de velocidade máxima de vento em Piracicaba – SP via séries temporais e teoria de valores extremos. **Revista Brasileira de Biometria**, v. 31, n. 2, p. 295–309, 2013.
- LISKA, G. R. *et al.* A simplex dispersion model for improving precision in the odds ratio confidence interval in mixture experiments. **Acta Scientiarum - Technology**, v. 42, p. 1–10, 2020.
- LJUNG, G. M.; BOX, G. E. P. On a Measure of Lack of Fit in Time Series Models. **Biometrika**, v. 65, n. 2, p. 297, ago. 1978.
- LUCCA FILHO, E. L. F. *et al.* Modelagem probabilística de preços máximos da commodity boi gordo para o estado de São Paulo. **Nativa**, v. 10, n. 1, p. 22–31, mar. 2022.
- MAPA - MINISTÉRIO DA AGRICULTURA PECUÁRIA E ABASTECIMENTO. Conheça a história do café no mundo e como o Brasil se tornou o maior produtor e exportador da bebida. 2022.
- MEHTA, A.; CHAVAS, J. P. Responding to the coffee crisis: What can we learn from price dynamics? **Journal of Development Economics**, v. 85, n. 1–2, p. 282–311, 1 fev. 2008.
- METCALFE, A. V.; COWPERTWAIT, P. S. P. **Introductory Time Series with R**. New York, NY: Springer New York, 2009.
- PAIVA, C. M. N. *et al.* Competitividade Do Café No Mercado Internacional: Uma Análise Econométrica. **Agroalimentaria - Universidad de los Andes**, v. 24, p. 43–58, 2018.
- REIS, C. J. DOS; BEIJO, L. A.; AVELAR, F. G. Temperatura mínima esperada para piracicaba-sp via distribuições de valores extremos. **Revista Brasileira de Agricultura Irrigada**, v. 11, n. 4, p. 1639–1650, 24 jul. 2017.
- SÁ BARRETO, R. C.; ZUGAIB, A. C. C. Dinâmica do mercado internacional de café e determinantes na formação de preços. **Economia & Região**, v. 4, n. 2, p. 7, 3 dez. 2016.

SALVIANO, M. F.; GROppo, J. D.; PELLEGRINO, G. Q. Análise de Tendências em Dados de Precipitação e Temperatura no Brasil Trends Analysis of Precipitation and Temperature Data in Brazil. **Revista Brasileira de Meteorologia**, v. 31, n. 1, p. 64–73, 2016.

SILVEIRA, R. L. F. DA; MATTOS, F. L.; SAES, M. S. M. The Reaction of Coffee Futures Price Volatility to Crop Reports. **Emerging Markets Finance and Trade**, v. 53, n. 10, p. 2361–2376, 2016.

TOLOI, C. M. C.; MORETTIN, P. A. **Análise de Séries Temporais**. 2. ed. [s.l.] Blucher, 2006.

WHITTLE, P. The simultaneous estimation of a time series harmonic components and covariance structure. **Trabajos de Estadística**, v. 3, n. 1–2, p. 43–57, fev. 1952.

### **Authors' Contribution**

All authors contributed equally to the development of this article.

### **Data availability**

All datasets relevant to this study's findings are fully available within the article.

### **How to cite this article (APA)**

Teixeira, O. M., Liska, G. R., Santos, J. A. dos, Marjotta-Maistro, M. C., & Pedro, D. A. de M. PROBABILISTIC AND TEMPORAL MODELING OF THE COFFEE MARKET IN BRAZIL. *Veredas Do Direito*, e235382. <https://doi.org/10.18623/rvd.v23.5382>