v. 3, n. 21, 2023

Scientific Journal of Applied Social and Clinical Science

TIME SERIES ANALYSIS FOR THE QUARTERLY GROSS DOMESTIC PRODUCT OF AMAZONAS

Casemiro Rodrigues de Souza
Department of Statistics and Geoprocessing, SEDECTI
Manaus – Amazonas
https://lattes.cnpq.br/0040882418539239

Natacha Porto de Sousa
Department of Statistics and Geoprocessing, SEDECTI
Manaus – Amazonas
https://lattes.cnpq.br/7569346209650187

Josenete Cavalcante Costa
Department of Statistics and Geoprocessing, SEDECTI
Manaus – Amazonas
https://lattes.cnpq.br/5162903372537980

Meilyn Leiene Machado Barbosa
Universidade Federal do Amazonas
Manaus – Amazonas
https://lattes.cnpq.br/3985989468414525

Alcides Saggioro Neto
Department of Statistics and Geoprocessing, SEDECTI
Manaus – Amazonas
https://lattes.cnpq.br/9717735951352217

Victória Juliana Moda Ferreira
Universidade Federal do Amazonas
Manaus – Amazonas
http://lattes.cnpq.br/403125839124086

Abstract: In this work, estimates were made for the GDP of Amazonas, in order to help managers in decision-making, seeing that currently the GDP has a two-year lag. The source of information for the data that make up the GDP is from the Brazilian Institute of Statistics and Geography (IBGE) in partnership with the Secretariat for Development, Science, Technology and Innovation (SEDECTI) using a historical series from the first quarter of 2010 to the fourth quarter de 2020. For data analysis and later making estimates, the sectors were broken down by their activities, Agriculture, Industry and Services. Three models were used for the estimates: ARIMA, SARIMA and ARFIMA, being necessary to verify the most adequate model for each variable. First, the stationarity of the data was verified through graphical analysis and the Said & Dickey (1984) – ADF and Kwiatkowski (1992) – KPSS test. These tests generated four results, providing guidance on which model is appropriate. After the application of the ADF and KPSS tests, the d parameter was estimated, which is the order of the differences necessary to remove the tendency of the series, using the GPH and Reisen methods. The most appropriate model was chosen using the Akaike Information Criterion (AIC). Finally, an estimate was generated for the years 2021 to 2023 and verified whether the selected models adequately describe the dynamics of the data. The result generated three models, namely: ARIMA, SARIMA AND ARFIMA and the estimates had a satisfactory dynamic in relation to the data set, that is, how close the predicted data are in relation to the values later observed. Through GDP estimates, managers can make better decisions, since GDP is the main indicator for the Government to make its budget planning for investments in health, security, education and infrastructure.
Keywords: GDP, Econometrics, Time Series, Macroeconomics, Statistics.

INTRODUCTION

The Gross Domestic Product - GDP is the indicator most used in macroeconomics to measure the total of goods and services produced by resident production units intended for final consumption, being equivalent to the sum of values added by the various economic activities plus taxes net of subsidies, on products. This indicator includes all goods and services consumed by people, companies and the government. The methodology for calculating the GDP is coordinated by the Brazilian Institute of Geography and Statistics – IBGE and is subdivided into four sections. The first refers to the construction of added value in agriculture and livestock. The second describes the procedures for obtaining the added value of Industry, and the third presents the estimation method for the Services sector. The fourth section describes the criteria for sharing taxes on products. The Value Added (AV) is calculated using the Gross Value of Production (GVP), where the GPV is the monetary expression (Price x Quantity) of the production carried out over a certain period of time minus the Intermediate Consumption (CI) which is the Value of Goods and Services (inputs) purchased and used in production, as follows:

$$VA = VBP - IC$$

The source of information for the data that make up the GDP first comes from existing data from other surveys, such as the Annual Survey of the Construction Industry - PAIC, Survey of Household Budgets - POF, Municipal Agricultural Production - PAM, among other surveys prepared by IBGE. Energy, communication and tax data are provided by the responsible bodies in each state, such as Amazonas Energia and SEFAZ (State Secretariat of Finance). In some cases where there is no information, IBGE makes an estimate taking into consideration, previous periods.

After collecting the information and obtaining the Added Value for the sessions, everything is added up to obtain the Nominal GDP, where:

$$VA\ Agriculture + VA\ Industry + VA\ Service + Taxes\ on\ products = Nominal\ GDP$$

Nominal GDP is calculated at current prices, considering prices recorded in the period in which the product was sold and real GDP, excluding the effects of inflation.

Given the scenario, this study aims to perform a time series analysis to identify the most appropriate model that models the 10-year quarterly series (period from the first quarter of 2010 to the fourth quarter of 2020) and obtain estimates to make forecasts for the GDP. For the analyzes and later to make the estimates, the activities of the sectors (Agriculture, Industry and Services) plus taxes were used, divided into Agriculture: Agriculture, Livestock, Forest production, Fishing and aquaculture. Industry: Extractive industries, SIUP (Public Utility Industry Services) and Construction. Services: Public administration, Food and lodging, Arts and culture, Financial activities, Real estate activities, Professional activities, Retail, Private education and health, Information and communication, Domestic services and Transportation. In which they were estimated for the quarters of 2021, 2022 and 2023.

METHODOLOGY FOR TIME SERIES MODELS

Stationary Processes: A time series is stationary when it develops randomly, in time, around a constant mean, reflecting some form of stable equilibrium, (Box, GEP & Jenkins, GW, 1994). However, most of the series found in practice show some form of non-stationarity, requiring some transformation. One way to analyze this non-stationary model is to incorporate a difference process ($\Delta^d Z t$) in the ARMA model. This is the model known as ARIMA (autoregressive integrated moving average model), where d is the order of differences necessary to remove the trend from the series. In this work, it was verified the need to adjust three models, the ARIMA, SARIMA and the ARFIMA and the stationarity of the time series was verified through the tests Said & Dickey (1984) – ADF and Kwiatkowski (1992) – KPSS.

ADF AND KPSS UNIT ROOTS TESTS

In this article we will use two tests to verify the existence of a unit root. We will apply the tests Said & Dickey (1984) – ADF and Kwiatkowski (1992) – KPSS. The ADF test examines the null hypothesis of the existence of a unit root, $H_0 = I\ (1)$, while the KPSS test examines the null hypothesis of the absence of a unit root, $H_0 = I\ (0)$. Using the ADF and KPSS tests together can generate four results.

a) The rejection of the null hypothesis of the ADF test and the non-rejection of the null hypothesis of the KPSS test, indicates strong evidence of a stationary process;

b) Non-rejection of the null hypothesis of the ADF test and rejection of the null hypothesis of the KPSS test, indicating that the series has a unit root;

c) The non-rejection of the null hypotheses in both the ADF and KPSS tests, which generates uncertainty about the data generating process;

d) The rejection of the null hypothesis in both the ADF and KPSS tests, suggesting that the generating process is neither I (0) nor I (1), indicating a probable fractional integration.

ARIMA MODEL

Many series found in practice are not stationary, but we can model these series making them stationary by differentiating this series. Let Z_t be a non-stationary time series. We take $W_t = \Delta Z_t = Z_t - Z_t - 1$, being differentiated once from Z_t. We denote by $W^d = \Delta^d Z_t$ the differentiated series d times of Z_t.

We can represent Wt by an ARMA model, as Wt is a difference of Z_t. Then Z_t is an integral of W_t, so we say that Z_t follows an integrated autoregressive model of moving averages, that is, an ARIMA model of order (p,d,q) and we write ARIMA (p,d,q), where p is the order of the autoregressive component, d is the number of differences taken in the series, and q is the order of the moving average component.

Therefore, we can describe all models using the ARIMA nomenclature:

1) ARIMA (p,0,0) = AR (p)
2) ARIMA (0,0,p) = MA (q)
3) ARIMA (p,0,q) = ARMA (p,q)

SARIMA MODEL

The SARIMA (Seasonal AutoRegressive Integrated Moving Average) model is a statistical model used to model time series that exhibit seasonality, trend and long-term dependence. It is an extension of the ARIMA (AutoRegressive Integrated Moving Average) model that allows dealing with time series that exhibit seasonality. A SARIMA model is a statistical model that uses an autoregressive (AR) process, a moving average (MA) process, and an integration (I) process to model a time series. The SARIMA model is characterized by four main parameters:

a) The AR(p) parameter: the number of autoregressive terms used in the model. Each autoregressive term represents the influence of past values of the series on its future evolution.

b) The MA(q) parameter: the number of moving average terms used in the model. Each moving average term represents the influence of past errors on the future evolution of the series.

c) The integration parameter (d): the degree of integration used in the model. The d parameter allows the time series to have a deterministic trend.

d) The seasonality parameter (P, D, Q): the number of autoregressive, integrated, and seasonal moving average terms used in the model. These parameters allow the SARIMA model to capture seasonal patterns in the time series.

The combination of these parameters allows the SARIMA model to capture the statistical properties of the time series, including its autocorrelation, partial autocorrelation, seasonality and other properties. The SARIMA (p, d, q)(P, D, Q) model is described:

$$\emptyset_p(B)\varphi_p(B^s)\Delta^d\Delta_s^D Z_t = \theta_q(B)\Theta_Q(B^S)\alpha_t$$

where $\Delta_S^D Z_t = (1 - B)^D Z_t$, where D is the order of seasonal differentiation, $\varphi_p(B^s)$is the seasonal operator MA(q) by $\Theta_Q(B^s) = 1 - \Theta_1(B^s) - \Theta_2(B2^s) - \dots - \Theta_Q(BQ^s)$ and $\varphi_1 \dots \varphi_p$ are parameters of the seasonal model AR(p) and are parameters of $\Theta_1 \dots \Theta_Q$ the seasonal model MA(q).

ARFIMA MODEL

The ARFIMA model is built from a time series and is used to predict future values of the series. The model is fitted to the data using parameter estimation techniques such as maximum likelihood. We say that Z_t it is an integrated fractional autoregressive process of moving averages, or ARFIMA (p,d,q) with d $\in$ (-1/2, 1/2) if it is stationary and satisfies the equation:

$$\phi(B)(1-B)^d Z_t = \theta(B)a_t$$

Where a_t is the white noise and $\phi(B)$ e $\theta(B)$ are the polynomials in B of degrees p and q, respectively. The reason for choosing this family of processes, for purposes of modeling series with long memory behavior, in which the effect of parameter d on distant observations decays hyperbolically as the distance increases, while the effects of parameters ϕ e θ decays exponentially. So, d must be chosen in order to explain the high-order correlation structure of the series, while the parameters ϕ and θ explain the low-order correlation structure.

MODEL ADJUSTMENT

One approach used in the analysis of parametric models of time series is the Box-Jenkins model, which is based on model fitting, following a scheme cycle:

1) Identification: based on the analysis of autocorrelations, partial autocorrelations and other criteria;
2) Estimation: where the identified model parameters are estimated;
3) Diagnosis: checks whether the selected model adequately describes the dynamics of the data;
4) Forecast: if the cycle above is successful, we continue with the forecast.

RESULTS

The data presented come from the Regional Accounts of the Brazilian Institute of Geography and Statistics (IBGE) and the Secretariat for Development, Science, Technology and Innovation (SEDECTI). Activity by activity of each sector was analyzed, after their respective estimates, these activities were added together with taxes to obtain the nominal GDP. The chosen series starts in the first quarter of 2010 and ends in the fourth quarter of 2020. As the GDP has a two-year lag, due to the data collection process, for the most recent years it is decided to make estimates.

AGRICULTURE AND LIVESTOCK

The Agriculture Sector is divided into three activities: Agriculture, Livestock, Forest production, fishing and aquaculture, as shown in Figure 1.

To verify whether the series follow a stationary process, the Said & Dickey (1984) – ADF and Kwiatkowski (1992) – KPSS tests were used, as shown in Table 1.

Test / p-value	Agricul-ture	Lives-tock	Forest production, fishing and aquaculture
ADF	0.573	0.7244	0.01
KPSS	0.01084	0.01	0.02457

Table 1. Unit root test for Agriculture activities

Assuming a significance level of 5%, we found that in the Agriculture activity we accepted the ADF test and rejected the KPSS, meaning that the series has a unit root, indicating an ARIMA or SARIMA model. In Livestock, the same case occurred in Agriculture, we accepted the ADF and rejected the KPSS, also indicating an ARIMA model. In Forest Production, Fishing and Aquaculture, both tests were rejected, indicating a probable fractional integration. Using the GPH and Reisen methods, d was estimated to verify the order of the differences needed to remove the trend from the series, and the most appropriate model was chosen using the Akaike Information Criterion (AIC).

Activity / model	AIC	Model
Agriculture	1648.08	SARIMA (0, 1, 0)(2, 0, 0)
Livestock	1317.47	SARIMA (0, 1, 2)(1, 1, 0)
Forest production, fishing and aqua-culture	1717,723	ARFIMA (0, 0.49, 0)

Table 2. Models chosen for estimates

After choosing the best models, the activities from the 1st quarter of 2021 to the

Scientific Journal of Applied Social and Clinical Science ISSN 2764-2216 DOI 10.22533/at.ed.2163212330085

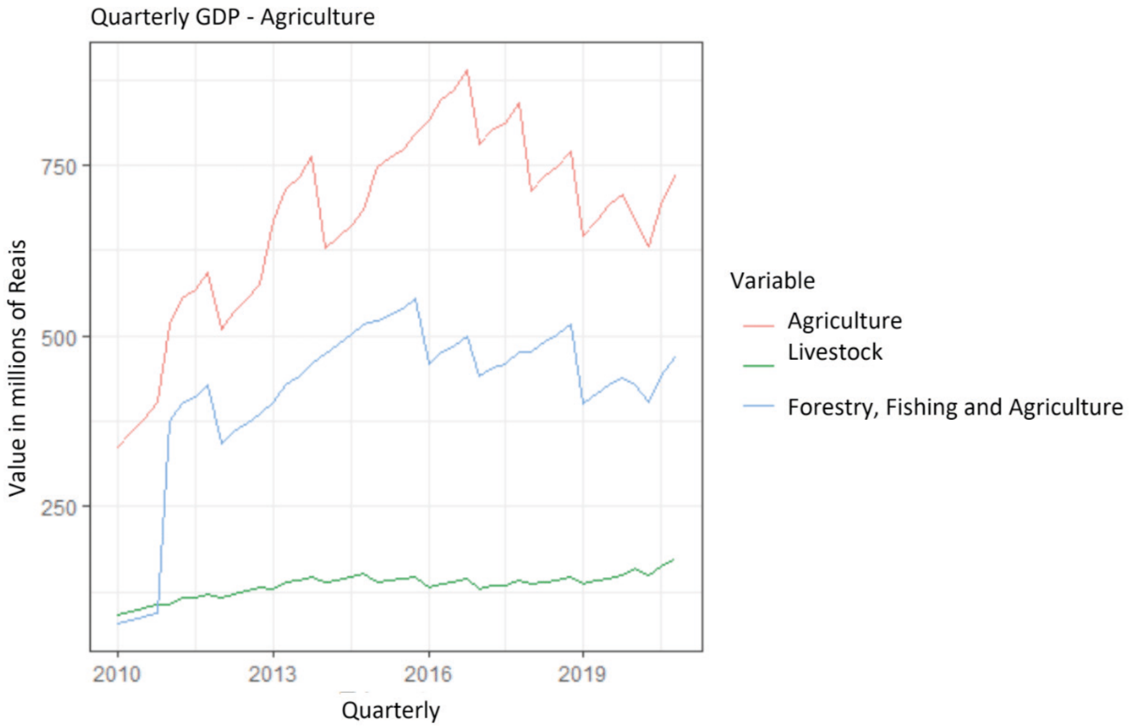

Figure 1. Historical series of Agricultural activities from the 1st quarter of 2010 to the 4th quarter of 2020.

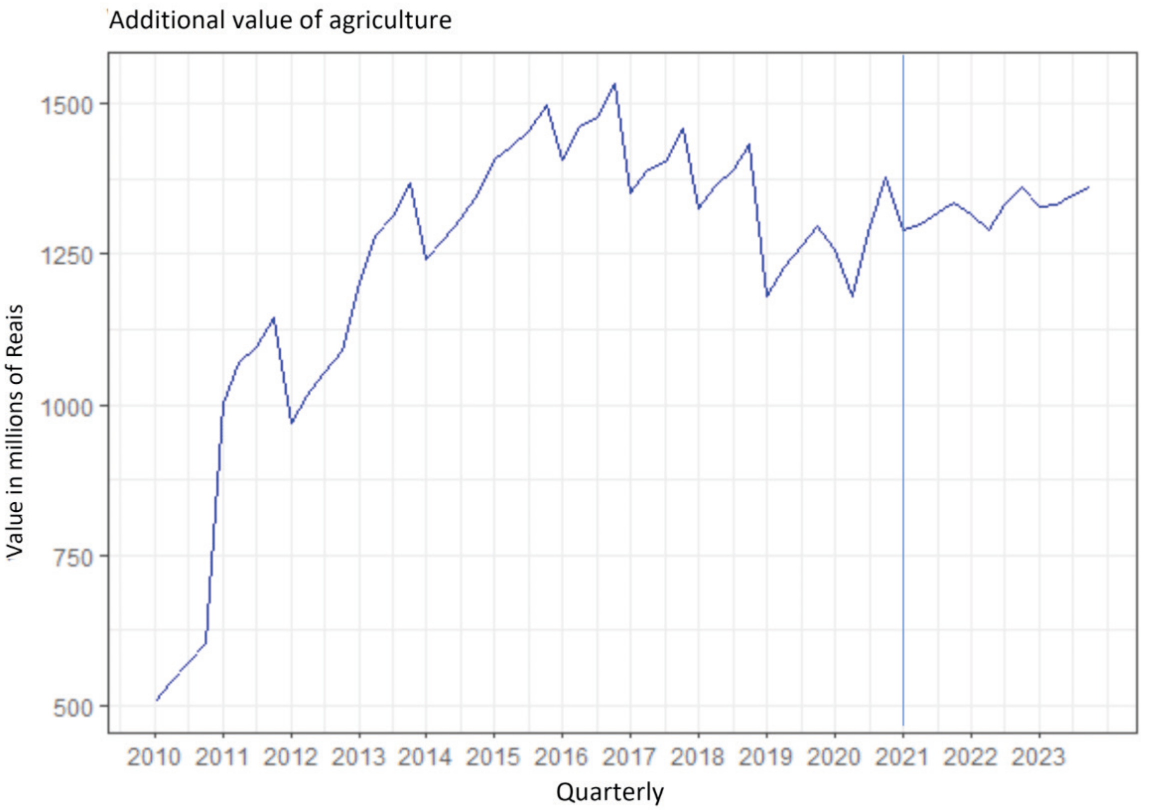

Figure 2. Historical series of Agricultural Added Value and estimates from the year 2021.

4th quarter of 2023 were estimated, in which the sum of the three activities has the VA of Agriculture, therefore, we have the estimate of Agriculture of Amazonas until the year 2023.

For the years 2021, 2022 and 2023, the VA estimates for Agriculture were R$5,248 million, R$5,307 million and R$5,377 million, respectively.

INDUSTRY

The Industry Sector is composed of four activities: Extractive Industries, Transformation Industries, SIUP (Public Utility Industry Services) and Construction, as shown in Figure 3.

To check whether the series follow a stationary process, the Said & Dickey (1984) – ADF and Kwiatkowski (1992) – KPSS tests were used, as shown in Table 3.

test / p-value	extractive industries	manufacturing industries	SIUP	Construction
ADF	0.4963	0.8683	0.935	0.944
KPSS	0.04598	0.01	0.01	0.1

Table 3. Unit root test for Industry activities

With a significance level of 5%, we verified that the activities of Extractive Industries, Transformation Industries and SIUP accepted the ADF test and rejected the KPSS, meaning that the series has a unit root, indicating an ARIMA or SARIMA model. In Construction both tests were accepted, which generates uncertainty about the data generation process, in this case we adjust the best model to an ARIMA or SARIMA.

activity / model	AIC	Model
extractive industries	1701.99	SARIMA (0, 1, 0)(2, 0, 1)
manufacturing industries	1800.92	SARIMA (0, 1, 0)(1, 0, 1)
SIUP	1667.99	ARIMA (0, 1, 0)
Construction	1719	SARIMA (1, 0, 2)(1, 0, 1)

Table 4. Models chosen for estimates

After choosing the best models, the activities from the 1st quarter of 2021 to the 4th quarter of 2023 were estimated, in which in the sum of the four activities we have the VA of Industry, as in Agriculture, we have the estimate of the Industry of Amazonas until the year from 2023.

For the years 2021, 2022 and 2023, the VA estimates for Industry were R$ 39,600 million, R$41,529 million and R$43,222 million, respectively.

SERVICES

The Services Sector comprises 11 activities: Public Administration, Housing and Food, Arts and Culture, Financial Activities, Real Estate Activities, Professional Activities, Commerce, Private Education and Health, Information and Communication, Domestic Services and Transport.

In tables 5, we have the unit root tests for the 11 activities of the Service Sector, with the tests by Said & Dickey (1984) – ADF and Kwiatkowski (1992) – KPSS.

Test / p-value	Public administration	accommodation and food	arts and culture	financial activities
ADF	0.9006	0.4438	0.08877	0.5938
KPSS	0.01	0.03944	0.01	0.01

Table 5.1. Unit root testing for Services activities

Test / p-value	real estate activities	Professional activities	Business	Private education and health
ADF	0.5476	0.4708	0.626	0.3338
KPSS	0.01	0.01	0.01	0.01

Table 5.2. Unit root testing for Services activities

test / p-value	information and communication	Domestic services	Transport
ADF	0.5895	0.3338	0.05195
KPSS	0.01	0.01	0.01

Table 5.3. Unit root testing for Services activities

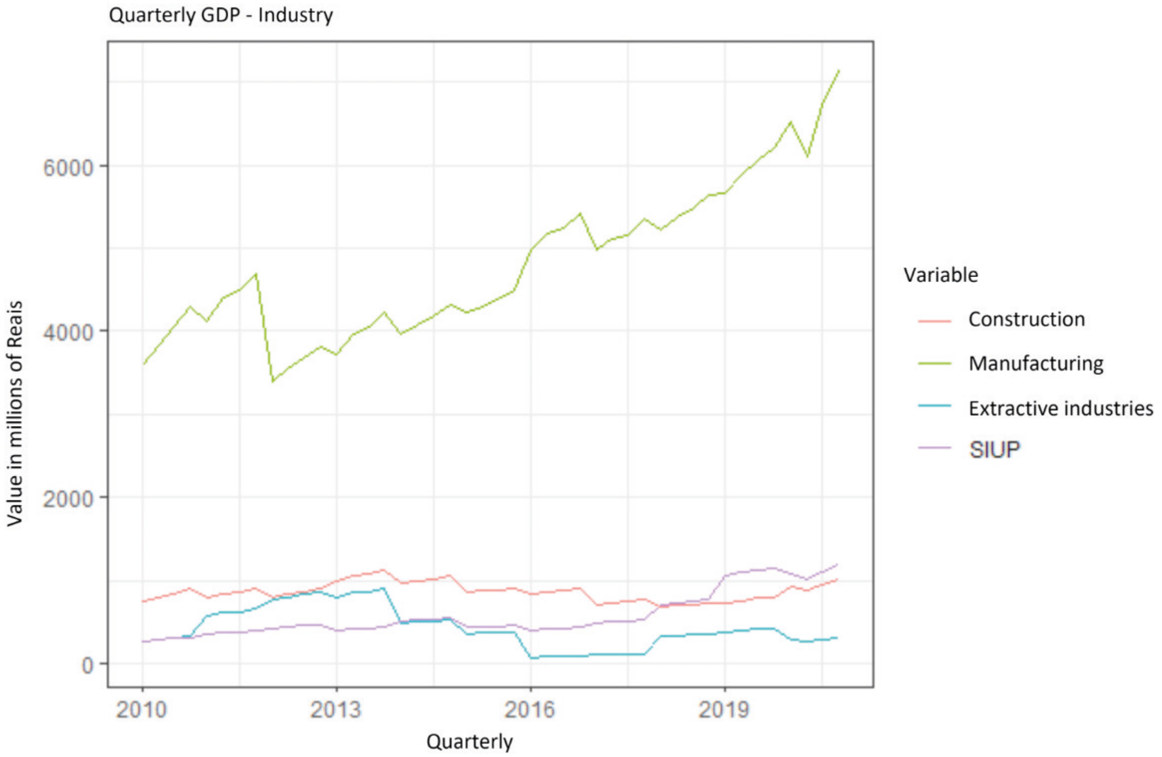

Figure 3. Historical series of Industry activities from the 1st quarter of 2010 to the 4th quarter of 2020.

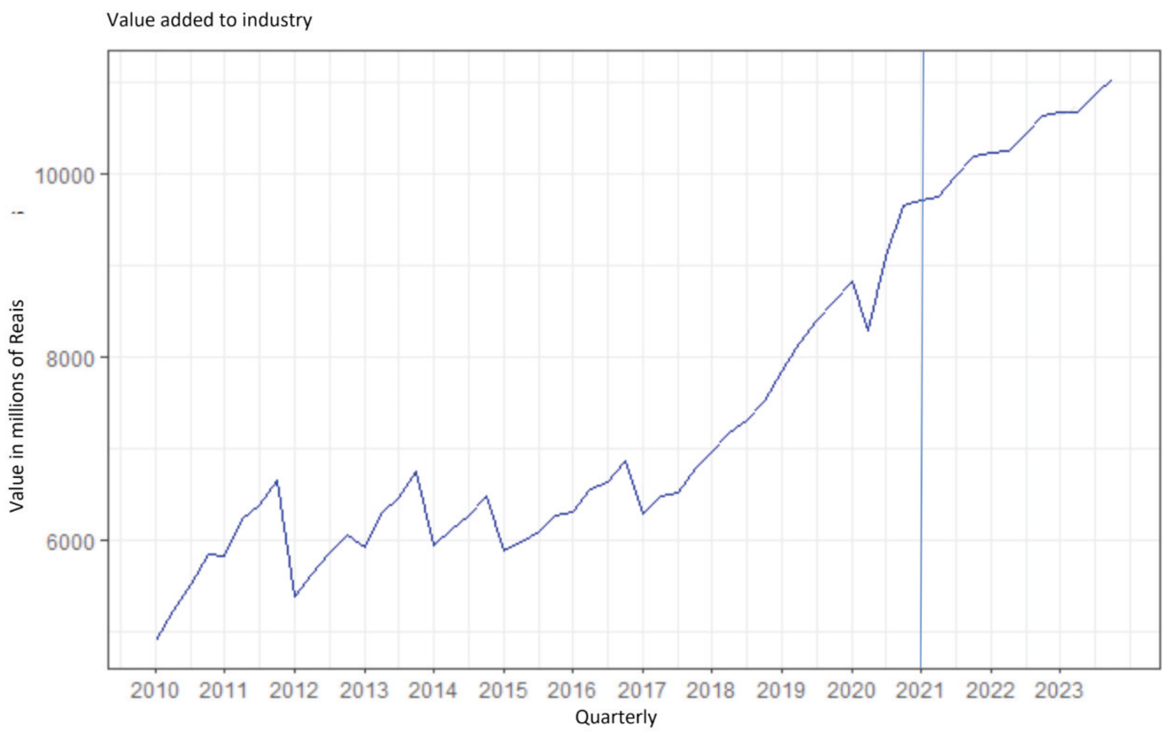

Figure 4. Industry Added Value historical series and estimates from the year 2021.

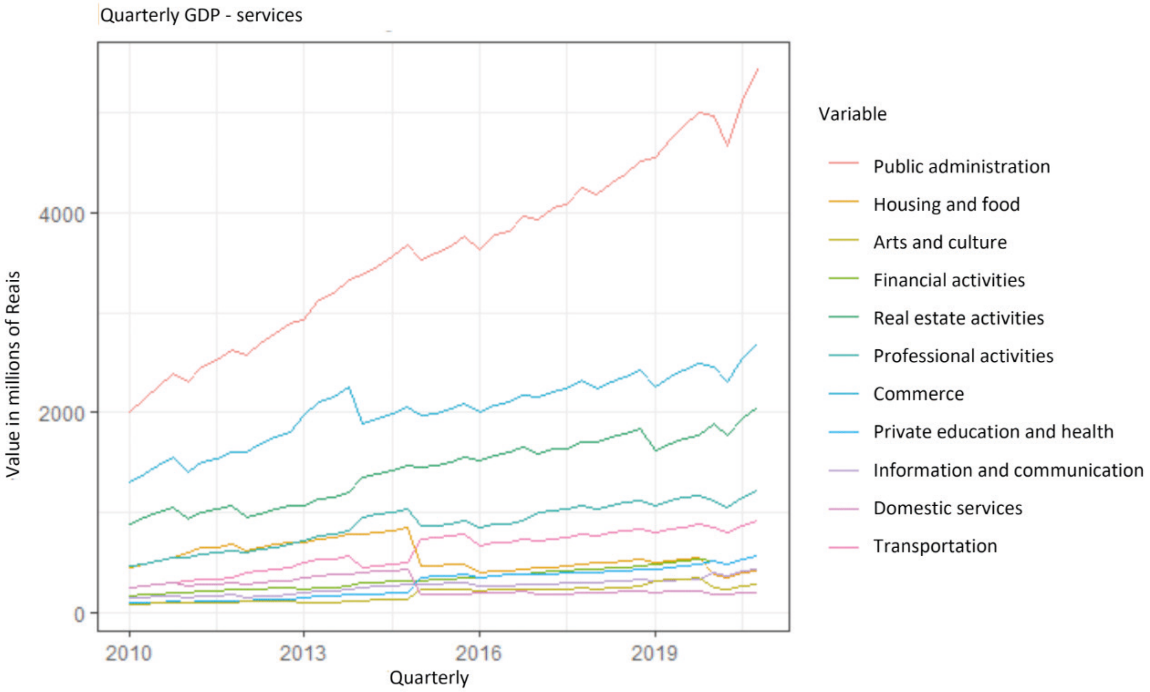

Figure 5. Historical series of Services activities from the 1st quarter of 2010 to the 4th quarter of 2020.

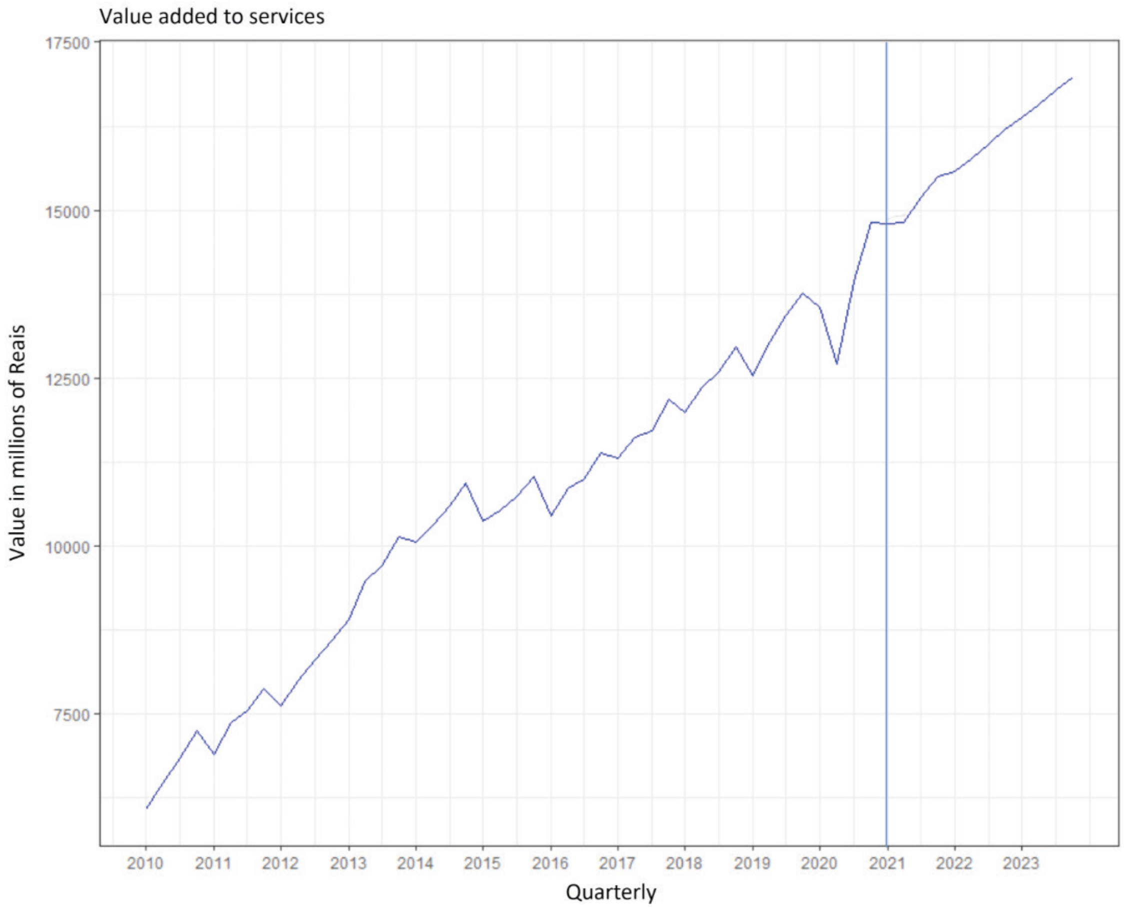

Figure 6. Historical series of Added Value of Services and estimates from the year 2021.

Figure 7. Historical series of Taxes from the 1st quarter of 2010 to the 4th quarter of 2020.

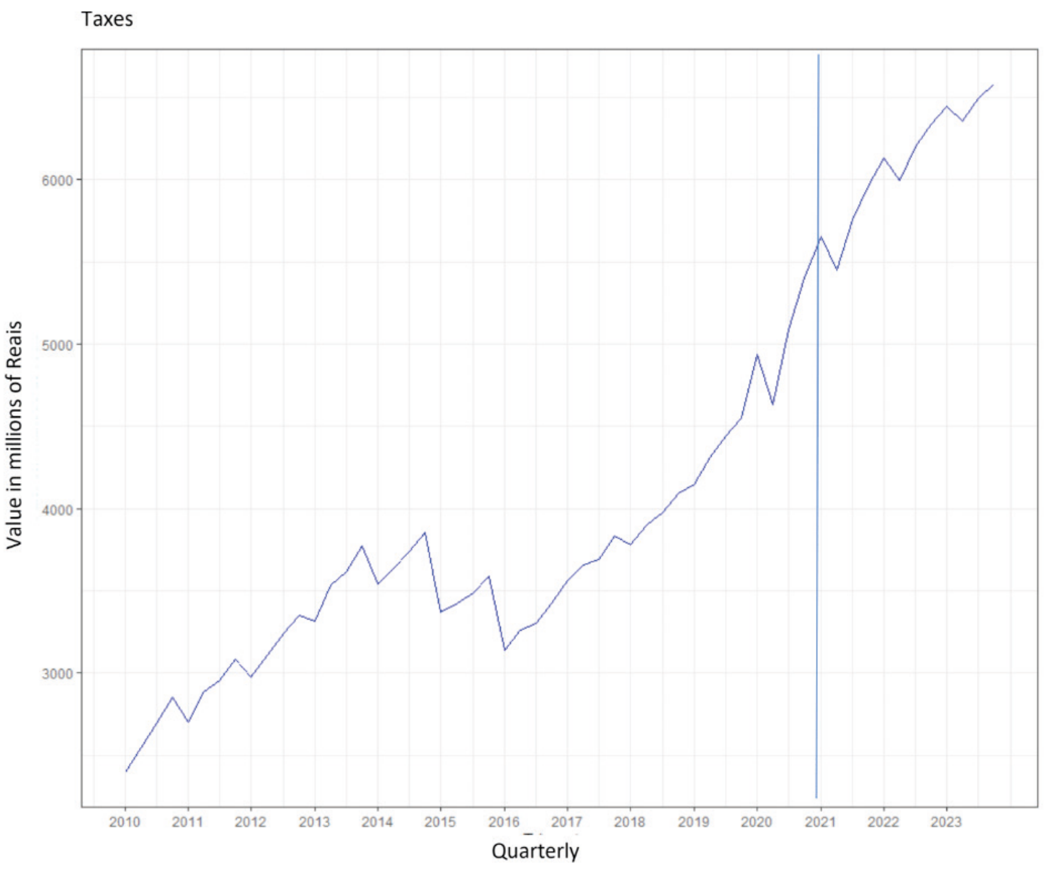

Figure 8. Historical series of Taxes and estimates from the year 2021.

With a significance level of 5%, we verified that all Services activities accepted the ADF test and rejected the KPSS, meaning that the series has a unit root, indicating an ARIMA or SARIMA model.

Activity / model	AIC	Model
Public administration	1719.69	SARIMA (0, 1, 1)(0, 0, 1)
Accommodation and food	1676.53	SARIMA (0, 1, 0)(0, 0, 1)
Arts and culture	1576.25	SARIMA (0, 1, 0)(2, 0, 0)
Financial activities	1525.95	ARIMA (0, 1, 2)
Real estate activities	1677.44	ARIMA (0, 1, 1)
Professional activities	1647.81	ARIMA(0, 1, 0)
Business	1710.94	ARIMA (0, 1, 0)
Private education and health	1592.79	ARIMA (0, 1, 0)
Information and communication	1549.8	ARIMA (0, 1, 0)(0, 0, 2)
Domestic services	1634.18	ARIMA (0, 1, 0)
Transport	1635.38	SARIMA (0, 1, 0)(0, 0, 2)

Table 6. Models chosen for estimates

After choosing the best models, the activities from the 1st quarter of 2021 to the 4th quarter of 2023 were estimated, in which in the sum of the 11 activities we have the VA of Services, thus, we have the estimate of the Services of Amazonas until the year 2023.

For the years 2021, 2022 and 2023, estimates of the Services VA were R$60,316 million, R$63,538 million and R$66,717 million, respectively.

TAXES

Taxes on GDP are those levied on production and net of subsidies. This indicator is the last one to be estimated and after that it is added to the other sectors to obtain the GDP.

In tables 7, we have the unit root tests for Taxes, with the Said & Dickey (1984) – ADF and Kwiatkowski (1992) – KPSS tests.

Test / p-value	Taxes
ADF	0.99
KPSS	0.01

Table 7. Unit Root Test for Taxes

With a significance level of 5%, we accept the ADF test and reject the KPSS, meaning that the series has a unit root, indicating an ARIMA or SARIMA model.

Activity / model	AIC	Model
Taxes	1749.04	SARIMA (0, 1, 0)(1, 0, 0)

Table 8. Models chosen for estimates

The best model for Taxes was SARIMA (0, 1, 0), (1, 0, 0). After choosing the best model, it was estimated until the year 2023.

The estimates for the years 2021 to 2023 were BRL 22,828 million, BRL 24,666 million and BRL 25,885 million, respectively.

AMAZONAS GROSS DOMESTIC PRODUCT

The Gross Domestic Product (GDP) is the sum of the Added Values of the sectors plus Taxes (Agriculture + Industry + Services + Taxes = GDP). For the best adjustment of the GDP, it was decided to estimate activity by activity, in which, following this methodology, the GDP value comes closer to the real, when this is released by the IBGE.

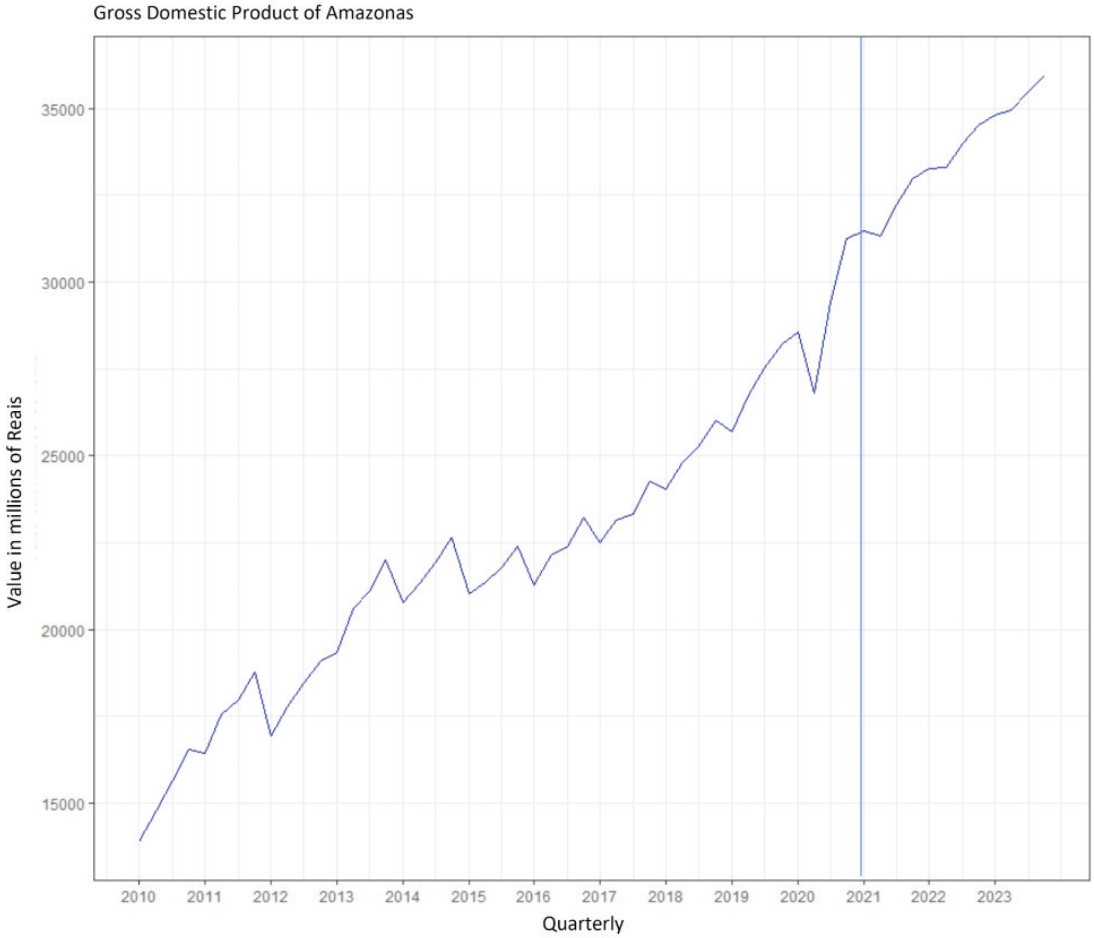

Figure 9. Historical series of the Gross Domestic Product and estimates from the year 2021.

Period	GDP (BRL million)
1st quarter 2021	31,470
2nd quarter 2021	31,318
3rd quarter 2021	32,232
4th quarter 2021	32,973
1st quarter 2022	33,248
2nd quarter 2022	33,302
3rd quarter 2022	33,964
4th quarter 2022	34,525
1st quarter 2023	34,818
2nd quarter 2023	34,946
3rd quarter 2023	35,481
4th quarter 2023	35,956

Table 9. Estimate by quarter of Amazonas' GDP

When we add the quarters, to obtain the annualized value, we estimate R$127,992 million for the year 2021, R$135,040 million for 2022 and R$141,202 million for 2023.

CONCLUSIONS

In this work, a time series analysis was performed for the quarterly Gross Domestic Product (GDP) of the state of Amazonas. GDP is the most used indicator in macroeconomics and measures the total of goods and services produced by resident production units intended for final consumption. The objective of the study was to estimate the GDP of Amazonas and help managers in decision-making, considering the two-year lag in official GDP data.

The analysis was based on a historical series covering the period from the first quarter of 2010 to the fourth quarter of 2020. Data were obtained through the Brazilian Institute of Geography and Statistics (IBGE) in partnership with the Secretariat for Development, Science, Technology and Innovation (SEDECTI). The analysis was carried out separating the sectors (Agriculture, Industry and Services) and taxes on products, and three time series models (ARIMA, SARIMA and ARFIMA) were applied to make estimates for the years 2021 to 2023.

To determine the most appropriate model, data stationarity tests were performed, such as the Said & Dickey (ADF) and Kwiatkowski (KPSS) tests. The order of differences needed to make the series stationary was estimated, and the Akaike Information Criterion (AIC) was used to select the best model. The results indicated that the ARIMA, SARIMA and ARFIMA models were adequate to describe the dynamics of GDP data.

The estimates generated by the models were considered satisfactory in relation to the observed data. Through these estimates, managers can make more assertive decisions, since GDP is an essential indicator for budget planning and investments in areas such as health, safety, education and infrastructure.

In short, the time series analysis for the quarterly GDP of Amazonas provided valuable information for managers, allowing a better understanding of the state's economic dynamics and assisting in making strategic decisions. The use of time series models proved to be an effective approach to estimate and forecast GDP, contributing to planning and regional economic development.

REFERENCES

BOX, G.E.P. & JENKINS, G.W. and REINSEL, G.C. **Time Series Analysis: Forecasting and Control**. Third Edition, 2015.

CRYER, Jonathan and CHAN, Kung-Sik, **Times Series Analysis with Applications in R**, Second Edition, 2008.

MONTGOMERY, D.C. & JONHSON, L.A. **Forecasting and Time Series Analysis**, New York, Mcgraw-Hill, 1976.

MORETTIN, P.A. & TOLOI, C.M. **Análise de Séries Temporais**. Editora Edgard Blücher, São Paulo, 2004.

SHUMWAY, Robert and STOFFER, David, **Times Series Analysis and Its Applications with R examples**, Third Edition, 2011.