Revista do TCU | v. 157 | Continuous flow | 2026 | e2026-2186 | DOI: https://doi.org/10.69518/2186.
Original article | Public accounting and data science

SIMULTANEOUS ANALYSIS OF REVENUES FROM MULTIPLE GOVERNMENT AGENCIES

Otoniel Arruda Costa | |
Court of Accounts of the State of São Paulo (TCE/SP), São Paulo (SP), Brazil

Henrique Raymundo Góia | |
University of São Paulo (USP), São Paulo (SP), Brazil

ABSTRACT

Budget planning is legally mandated, yet many administrators still face challenges in revenue forecasting. Despite advances in data science, gaps remain in the application of predictive models for this purpose. This study proposes a computational model in R capable of simultaneously analyzing time series of budget revenues from multiple collecting agencies, automatically comparing different methods – Autoregressive Integrated Moving Average (ARIMA) and Exponential Smoothing. The developed algorithm enables analysts to review outputs from all methods using a single block of code, eliminating the need for repetitive function calls. The algorithm was developed using data from the portals of the São Paulo State Court of Accounts (TCE-SP) and the State System Foundation for Data Analysis (SEADE). Two cities with similar economic characteristics as of December 31, 2020 – Sorocaba and Sao Jose dos Campos – were selected, and time series forecasting models were generated for their revenues for the 2022 and 2023 fiscal years. The methodology involved comparing models using the Akaike Information Criterion (AIC), the Corrected Akaike Information Criterion (AICc), and the Bayesian Information Criterion (BIC), alongside the Mean Absolute Percentage Error (MAPE) statistic. The ARIMA method yielded the best information criterion results for both cities and the lowest MAPE for Sao Jose dos Campos (9.75%). With the Exponential Smoothing method, Sorocaba showed the lowest MAPE (12.31%). The results demonstrate the viability of the proposed model in facilitating comparative analyses between municipalities and supporting decision-making by public administrators, auditors, and other stakeholders interested in government economic data.

Keywords: business analytics; COVID-19; Fable; budget forecasting; X13-ARIMA-SEATS.

1 INTRODUCTION

One of the key factors for the balance between public revenues and expenditures, i.e., fiscal balance, is the adoption of a proper budget planning process. This is a conclusion that can be drawn from the provisions of Articles 11 and 12 of Brazilian Supplementary Law No. 101 of May 4, 2000 – Brazilian Fiscal Responsibility Law (Lei de Responsabilidade Fiscal, LRF).

To achieve this objective, Brazilian Federative Entities are required to prepare medium-term government planning through the Pluriannual Plan (Plano Plurianual, PPA) - established by ordinary law and effective for four years. Consistent with this planning instrument, they are required to annually issue the Budget Guidelines Law (Lei de Diretrizes Orçamentárias, LDO) and the Annual Draft Budget Law (Lei Orçamentária Anual, LOA), which guide the preparation and execution of public budgets.

Recent studies have shown that there are still difficulties in budget forecasting. Oliveira Junior and Levino (2024) reported that the Federal Government provided financial assistance to municipalities during the COVID-19 pandemic without robust technical analysis, highlighting the need for more accurate predictive models to support fiscal decision-making in uncertain scenarios.

In another study, Nascimento and Boente (2022, p. 11), after examining all the 5,569 Brazilian municipalities between 2015 and 2018, demonstrated that forecasting errors from previous fiscal years (19.7% on average) affect forecasting errors in subsequent years (17.5% on average), plus or minus. Moreover, Toledo Júnior (2010, p. 36), asserted that budget revenues are often adjusted to match the "desired level of expenditure".

On the other hand, fortunately, data sciences have been providing increasingly powerful analytical tools, including those for the public sector. Fávero and Belfiore (2021, p. 14) emphasized that professional applications "in addition to offering substantial data-processing capacity, are capable of developing a wide variety of appropriate and robust tests and models". A recent example is the study conducted by Faveret *et al*. (2021), which investigated similarities in the evolution of COVID-19 across Brazilian states. The authors employed data clustering techniques and time series analysis to examine disease occurrence patterns. In another application, the government of the State of Mato Grosso (Sá *et al*., 2020) recommended the use of neural networks as a predictive methodology.

Despite the advances, data analysis remains underused by public managers. Araújo, Behr, and Schiavi (2023) emphasized that many tools used by auditors do not incorporate modern data analysis techniques, thereby limiting forecasting accuracy. Furthermore, the authors noted that the encapsulation of available data analysis tools prevents professionals from understanding the underlying algorithms. Similar conclusions were reached by Batóg and Batóg (2021) and Liu, Liu, and Chen (2023), who advocated expanding research on predictive methods, techniques, and algorithms.

Given this scenario, the objective of the research is to propose a model that facilitates the analysis of multiple time series from revenue-collecting agencies through a computational algorithm in R programming language. The purpose of the model is to enable users of government economic information to save time when identifying the most appropriate forecasting technique for different time series. The study also intends to contribute to greater comparability between institutions and to achieve economies of scale in the volume of data analyzed.

Consider the hypothetical case of a budget analyst who compares the time series of budget revenues from the municipality where he works with those of another municipality with similar economic characteristics. In such a situation, the forecasts generated by the algorithm could serve as a starting point for investigating whether the other municipality is benefiting from specific legislation or public policies that might also be applicable to their own jurisdiction, particularly if significant differences were identified in the evolution of budget revenues over the same period.

Another possible application involves auditors comparing revenues from two agencies or governmental entities with similar characteristics to potential indicators of irregularities or fraud based on the detection of revenue collection patterns.

From this perspective, the proposed solution aims to provide the various stakeholders interested in economic and financial information with the means of proactively monitoring the future evolution of public revenues, using these forecasts as a supporting tool for management, oversight, and decision-making.

Potential users of the proposed solution include audit institutions, treasury, research organizations, public administrators, citizens, and other parties interested in monitoring fiscal management.

2 LITERATURE REVIEW

This section is subdivided into three subsections addressing the literature on public revenues, time series, and predictive processes and techniques.

2.1 Public Revenues

According to the Accounting Manual Applied to the Public Sector (Manual de Contabilidade Aplicado ao Setor Público – MCASP), budget planning is a relevant aspect of public management (Brazil, 2024). However, despite technological and methodological advances, this remains a challenging task for governments. Several countries face difficulties in this process and are looking for alternatives to improve the forecast of public revenues.

Batóg and Batóg (2021), when analyzing the risk factors related to regional revenue planning in Poland, concluded that successful forecasting processes should not rely solely on revenue structure, but also on the selection of an appropriate forecasting method. In their study, the authors proposed two models: one based on time series and the other based on revenue collection as a function of Gross Domestic Product (GDP).

Koniagina (2020) examined problems associated with the implementation of digital solutions for revenue forecasting processes in local governments of the Russian Federation. Among the main obstacles to the adoption of forecasting and planning algorithms, the author highlighted the “absence of a unified fiscal planning and forecasting tool integrated into the consolidated planning and forecasting system, at both regional and local levels” (Koniagina, 2020, p. 6).

In the United States of America, Levy, Chasalow, and Riley (2021) analyzed the use of algorithms in the decision-making process of public management from different perspectives, including computational, legal, social, and ethical dimensions. The authors observed that restrictions imposed by state governments on local government revenues may limit the financial capacity of these entities and compromise their performance. In this context, they argue that the adoption of technological tools represents a valuable alternative for supporting complex public policy decisions.

In Spain, Muñoz, Bolívar, and Arellano (2022) emphasized that the difficulties faced by regional and local governments in controlling public finances have favored the adoption of technological solutions aimed at improving fiscal management and decision-making processes.

In Brazil, the State of Mato Grosso incorporated the concept of neural networks into its Methodology Used Revenues and Expenditure Estimates (Sá *et al*., 2020), for public revenue estimation. According to the document, under this approach, the values of the time series serve as input data (independent or regression variables) to which weights are automatically assigned by the algorithm during the training process. With each new iteration, the model parameters are recalibrated based on the errors observed in previous simulations, allowing forecasting performance to improve over time.

In addition to the traditional challenges associated with the process of forecasting of governmental collecting agencies, the international public health emergency resulting from COVID-19 introduced additional difficulties. The most significant was the substantial variability observed in the data, particularly during 2020 and 2021, resulting from the partial interruption of economic activity worldwide. In this regard, Oliveira Junior and Levino (2024), using Spearman's Correlation test, showed that the transfer of financial assistance by the Federal Government to the municipalities affected by COVID-19 did exhibited no clear correlation with revenue losses experienced by those entities, thereby impairing the technical assessment underlying the process of decentralization of resources to subnational entities. Nascimento and Boente (2022, p. 11) asserted that municipalities should adopt more “realistic” forecasting techniques. They also recommended that future research should consider external factors, such as periods of economic crisis.

2.2 Time Series

The development of a forecasting model requires an understanding of the concept of time series. An appropriate definition is “any set of observations ordered through time.” According to Morettin and Toloi (2018, p. 1), time-series analysis offers broader and more interesting objectives than forecasting alone, including describing the behavior of revenues and expenditures and investigating the mechanisms responsible for generating such behavior.

One of the most widely used methods for time-series forecasting is Exponential Smoothing, commonly referred to as Error, Trend, and Seasonality (ETS). Its main characteristic is the reduction of the impact of random fluctuations present in the data by attributing greater weight to more recent observations. According to the authors (2018, p. 89), the fluctuations occurring at the extremes of a series represent sources of “randomness,” which helps explain the widespread use of the method, due to its “simplicity, computational efficiency, and reasonable accuracy.” ARIMA models, according to Sartoris (2003), are based on historical observations from the series itself, in other words, they operate as a regression of the series on its own past values.

Morettin and Toloi (2018) classify Exponential Smoothing into three categories: Simple Exponential Smoothing, recommended for series exhibiting neither trend nor seasonality; Holt's method, suitable for series with trend but no seasonality; and the Holt-Winters method, appropriate for series that simultaneously exhibit trend and seasonality.

Ferreira and Duca (2020, p. 75) define trend as the pattern of growth or decline typically present in a series, “not necessarily linear.” They further explain that seasonality refers to the repetition of values at annual intervals.

Another important concept is the stationarity of time series. Sartoris (2003, p. 342) defines a time series represented by Yt as stationary when its historical value tends towards zero, such that any eventual shock is “dissipated over the years.” The author further advocates that, in this type of series, mathematical expectation and variance remain constant. By contrast, when there is an error component added to the value of the previous period of the series, the process is characterized as a random walk.

According to Morettin and Toloi (2018, p. 5), “one of the most frequent assumptions concerning a time series is that it is stationary, that is, that it evolves randomly over time around a constant average trend.” For the authors, time series generally present trends when related to economic and financial data series. They also point out that most forecasting methods have the premise that time series behavior patterns are based on previous data. Furthermore, they warn that non-stationary series must undergo a transformation process known as differencing in order to achieve stationarity. To detect the existence of non-stationarity, whether arising from deterministic trends or stochastic trends associated with unit roots, the authors list several tests. One of them is the Dickey-Fuller test, which employs
the least squares method to identify unit roots and remove trends from a series that, with stochastic trends being eliminated through differencing procedures.

The literature identifies two additional concerns in time-series analysis. The first is the need to stabilize variance, for which “a logarithmic transformation may be appropriate” (Morettin; Toloi, 2018, p. 9). The other point to be examined is whether the data are adequately represented by the selected model (Souza *et al*., 2022). To address this issue, an overfitting test may be applied. According to Morettin and Toloi (2018), the technique involves adding parameters to the model. If the residual variance does not decrease, the model initially identified is regarded as the most appropriate. Alternatively, residual autocorrelation detection techniques may be employed, such as the Ljung-Box test described by Ferreira and Mattos (2020). Regarding measures used to evaluate forecasting accuracy by comparing predicted and observed values, the Mean Absolute Percentage Error (MAPE) is widely recognized, as discussed by Kim and Kim (2016).

For the purposes of this study, the Forecast Accuracy Scale proposed by Klimberg, Sillup, Boyle and Tavva (2010) was adopted. According to its scale, a MAPE below 10% indicates highly accurate forecasts; values between 10% and 20% indicate good forecasts; and between 21% and 50% indicate reasonable forecasts. On the other hand, a MAPE exceeding 50% signifies an inaccurate forecasting model.

2.3 Predictive Processes and Techniques

Revenue forecasting based on time-series analysis is increasingly associated with technological development. In recent years, a concept that has gained prominence in data analysis is Business Analytics (BA), understood as the application of statistical and mathematical methods to extract knowledge from data and support the decision-making processes.

According to Araújo, Behr, and Schiavi (2023, p. 2), this concept was developed “to represent the main analytical component of Business Intelligence (BI),” whereas BI is more closely associated with the production and dissemination of reports intended to support decision-making.

In their review of the BA literature, Liu, Liu, and Chen (2023) identified four perspectives for defining the concept: technical, process-oriented, practical, and managerial. For the present study, the first two are particularly relevant as they are most directly related to the development and application of public revenue forecasting algorithms.

From a technical point of view, BA encompasses any application involving data science or analysis. From a process perspective, the authors define BA as the "encapsulation of tools to convert data into actionable insights through a scientific, mathematical, and intelligent process".

From a technical point of view, Hyndman and Rostami-Tabar (2024) analyzed time-series strategies with substantial changes in the series, including interruptions in data over specific periods. The authors proposed solutions for situations involving specific government interventions at particular points in time; disruptive events, such as the COVID-19 pandemic; or even missing data. To this end, and using Australian tourism data from 2019 to 2022, they analyzed, among others, the following modeling approaches: adaptive, ensemble, intervention-based, and what they termed "what-could-have-been estimates" (Hyndman; Rostami-Tabar, 2024, p. 7). The authors concluded, at the end of the study, that each situation may require a different solution. As an example, complex governmental interventions are more appropriately addressed through adaptive methods, whereas ensemble approaches are useful "when there is uncertainty about which approach to use and usually results in greater accuracy through the power of averaging" (Hyndman; Rostami-Tabar, 2024, p. 19). In the present study, the series examined did contain interruptions, even during the COVID-19 emergency period.

Additionally, in the context of the pandemic, an important tool in the modeling process was the X13-ARIMA-SEATS seasonal adjustment program (Sax; Eddelbuettel, 2018). This application was developed in 2012 by the U.S. Census Bureau (Mattos; Ferreira, 2020) and operates in conjunction with the Seasonal package, in R programming language. One of the of its major strengths, according to these authors, is the pre-adjustment feature, which enables the modeling of time series while accounting for trading days, holidays, and extreme observations (outliers), among other aspects. Prior to automated adjustments, users may refine the modeling specifications through manual adjustments, given that calendar structures differ across countries and years.

Another noteworthy study involving X13-ARIMA-SEATS is that of Abeln and Jacobs (2022), who investigated the impact of the COVID-19 pandemic on the Gross Domestic Product (GDP) of the Netherlands. n their analysis, they analyzed the pandemic as an outlier in the context of time-series data analysis. In Brazil, the X13-ARIMA-SEATS program is used by several institutions responsible for compiling official economic indicators, such as the Central Bank of Brazil (Brazil, 2021) and the Ministry of Economy (Brazil, 2020).

3 METHODOLOGY

With respect to its nature, this study adopts an applied research approach, since its objective is practical implementation (Gerhardt; Silveira, 2009). Regarding its purpose, the study is classified as descriptive research because it examines the application of an algorithm to carry out the process of comparing the revenues of two distinct municipalities. In terms of research approach, it is quantitative, as the phenomena under investigation refer to statistical analysis of numerical data. Concerning the methodological procedure, the study is characterized as a case study, since it seeks to address a budgetary data comparison problem through the development and application of an algorithm.

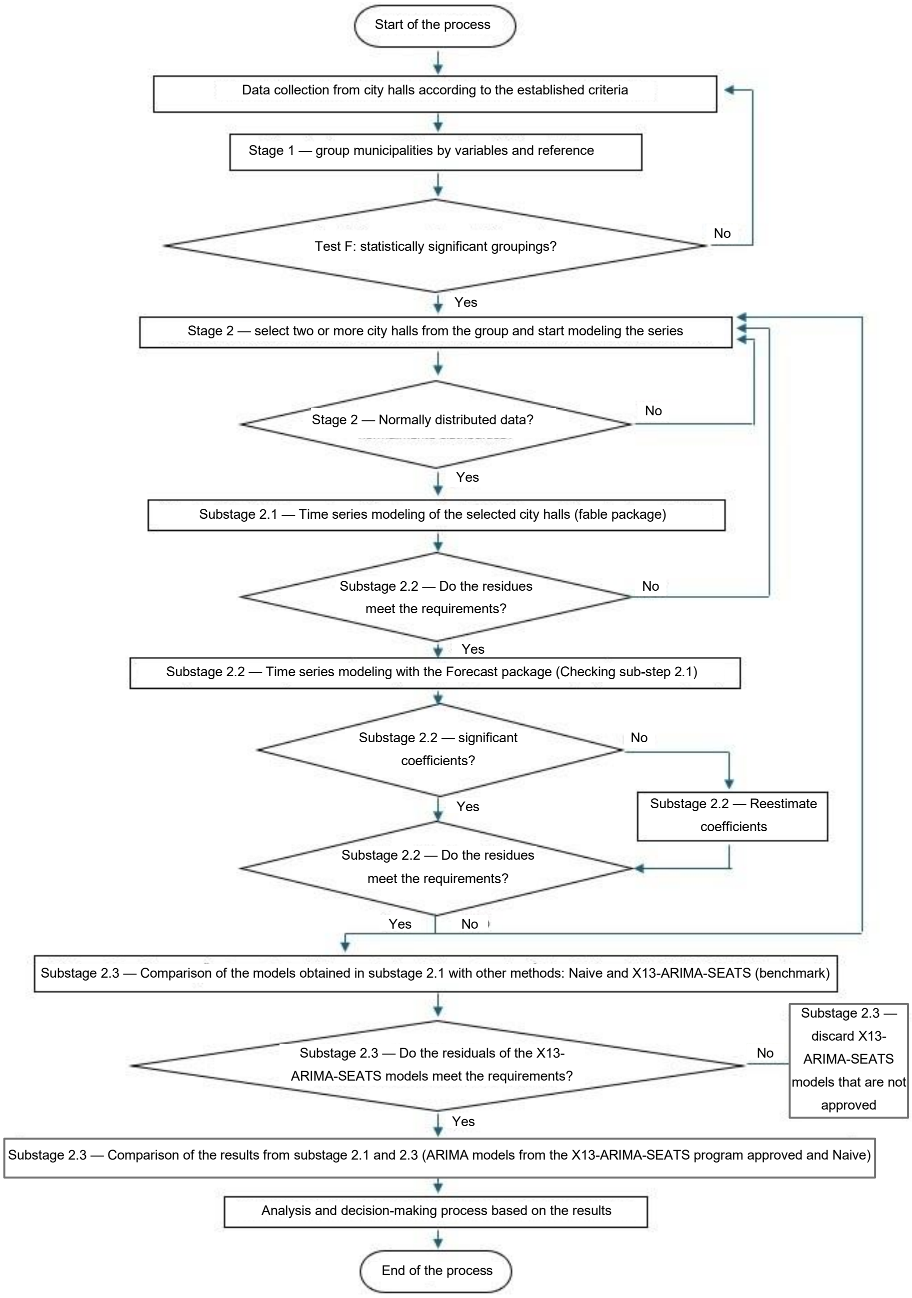
To ensure a logical organization of the content and facilitate a gradual understanding of the procedures performed, this section is divided into the subsections “characterization of the research subject,” “data collection,” and “data analysis,” following the framework proposed by Volpato (2015), with adaptations.

3.1 Characterization of the Research Subject

The object of analysis are Brazilian municipalities in the State of São Paulo (SP). To obtain financial resources, these entities must comply with Law 4,320 of March 17, 1964, and the LRF. Accordingly, they are required to issue, on an annual basis, the LDO and the LOA, regulations that govern revenue recognition and expenditure execution. Although this study focuses on municipalities, the proposed methodology may be applied to other governmental entities. For sample selection purposes, the concept of municipal federative entity according to the Brazilian Federal Constitution of 1988 was adopted. Consequently, revenue-collecting agencies such as pension funds, autonomous government agencies, and other municipal public entities legally authorized to collect revenues at the local level were also included. Therefore, the municipality is understood as the aggregate of all such entities.

The model development process, which consisted of designing the computational algorithm, is described below. Figure 1 presents a simplified flowchart of the modeling process.

Figure 1 — Flowchart of the multiple time-series modeling process



Source: Prepared by the authors.

3.2 Data Collection

Data were collected from the websites of the TCE-SP (State Court of Accounts of São Paulo) and the SEADE Foundation (State System Foundation for Data Analysis). Information on total municipal revenues from 2008 to 2023 was obtained from the TCE-

SP database (São Paulo, 2020c). Municipal Gross Domestic Product (GDP) for 2020 (São Paulo, 2020a) and estimated population of 2020 (São Paulo, 2020b) were collected from the SEADE Portal.

The total revenues spreadsheet contains revenue classifications in accordance with Summary Table of the Accounting Manual Applied to the Public Sector (MCASP) (Brazil, 2024). The data were organized in electronic spreadsheets and subsequently processed using the R programming language. The resulting values are presented in Table 1.

3.3 Data Analysis

After obtaining and preparing the data, the algorithm, also referred to as a script, was developed to generate the forecasting models. The algorithms were programmed in R (R Core Team, 2023) and are available in the Zenodo repository (https://zenodo.org/records/21877515).

3.3.1 Stage 1

Using the three types of data described above, a cluster analysis was performed, referred to as Stage 1 of the experiment, following the clustering concepts and techniques presented by Fávero and Belfiore (2021), described below.

- Standardization of variables (GDP, revenues, and population).
- Calculation of distances (or dissimilarities) between observations using the Euclidean distance measure.
- Adoption of a hierarchical clustering scheme using the single-linkage method.
- Generation of the dendrogram, a graph that allows the visualization of the clusters.

- Analysis of variance between groups (F test). The purpose of this test is to verify whether between-group variability exceeds within-group variability (Fávero; Belfiore, 2021, p. 338). In other words, it determines whether each variable chosen as a grouping parameter (GDP, revenue, and population) contributed to the formation of the clusters. This procedure is carried out through an Analysis of Variance (ANOVA). If, in the table with the results, the F-statistic is large and the p-value is less than 0.05 (the significance level adopted in this study), then the null hypothesis that "the variable under analysis has the same mean across all groups" can be rejected (Fávero; Belfiore, 2021, p. 338). The alternative hypothesis (H1) occurs when at least one group differs from the others. If confirmed, it can be concluded that the selected variables contributed to cluster formation in the sample. According to the authors, the test is strongly influenced by sample size; therefore, variables that do not contribute to clustering should be removed. Consequently, if the null hypothesis cannot be rejected, the municipality selection stage must be revisited.

3.3.2 Stage 2

Following the clustering process, Stage 2 began with the examination of municipalities belonging to the selected clusters. The activities performed during this stage are described below.

- Selection of two municipalities, whose revenue time series from 2008 to 2021 exhibited normal distributions at the 5.0% significance level (Morettin; Toloi, 2018). To assess normality, the Kolmogorov-Sminoff test (KS Method) was conducted using the Statistic Kingdom platform (2017). Revenue data from the selected municipalities were then classified according to the budget revenue-by-nature framework established by Brazil (Brazil, 2024), particularly the categories known as "economic categories" and "source". An important adjustment involved excluding intra-budgetary current and capital revenues from the "category" variable, since these include transfers between agencies within the same municipal budget structure. This procedure avoided double counting. Intra-budgetary revenues originating from external entities were also excluded. It is worth noting that, in the revenue spreadsheets obtained from TCE-SP, the "source" classification corresponds to the column labeled "subcategory," as prescribed by the accounting recording standards of TCE-SP (São Paulo, 2007, p. 9).

- Importing of the datasets into RStudio and loading of the required packages, particularly forecast (Hyndman *et al*., 2024) and Fable (O´Hara-Wild *et al*., 2023), which enables the "stacking" of multiple time series for simultaneous analysis. In addition, the Fable package (O´Hara-Wild *et al*., 2023) incorporates parameterizations from the Forecast package and therefore uses AIC and BIC statistics to calculate the most appropriate model for each time series (O´Hara-Wild *et al*., 2023).

- Division of samples into training (January 2008 to December 2021) and testing (January 2022 to December 2023).

3.3.2.1 Substage 2.1 – Modeling Using the Fable Package

Went through the actions listed below.

- Generation of ARIMA and Exponential Smoothing models.

- Examination of model coefficients and selection of models with the lowest AIC, AICc, and BIC values.

- Performing the Ljung-Box test to verify the absence of residual autocorrelation.

- Application of the KS test to assess residual normality (Lopes; Branco; Soares, 2013).

- Application of the ARCH test to assess variance stationarity.

- Calculation of the Mean Absolute Percentage Error (MAPE) for forecasts.

Ferreira and Mattos (2020) demonstrate that if the assumptions of residual independence, normality, and homoscedasticity (constant variance) are not satisfied, the model is unsuitable for forecasting because its linear component has not been properly specified. Therefore, whenever residuals fail any of these three diagnostic tests, new municipalities must be selected and the process repeated.

3.3.2.2 Substage 2.2 – Validation Using the Forecast Package

To assess the accuracy of the models obtained in Substage 2.1, modeling was carried out using the Forecast package, following the identification, estimation, diagnostic, and forecasting procedures described by Ferreira and Mattos (2020). The following analyses were performed.

- Identification and estimation of ARIMA models: identification and estimation were performed using the auto.arima function, obtaining the autoregressive “p” (AR), integrated “d” (I) and moving averages “q” (MA) components, as well as the corresponding seasonal parameters (P, D, and Q).

- Assessment of coefficient significance: the statistical significance of the estimated coefficients and the adequacy of the model were evaluated.

- Residual diagnostics: the absence of residual autocorrelation, residual normality, and homoscedasticity were examined.

- Forecasting and accuracy assessment: forecasts were generated and the MAPE error statistic values were calculated.

- Exponential Smoothing Modeling: the forecast.ets function was used, selecting the model with the lowest AIC value (Hyndman; Khandakar, 2008).

- Comparison of information criteria: it was verified whether the AIC, AICc and BIC obtained were the same as those generated in substage 2.1.

- Residual diagnostics for Exponential Smoothing models: residual independence, normality, and variance stationarity were verified.

- Forecasting using Exponential Smoothing: forecasts were generated and MAPE statistics were calculated.

3.3.2.3 Substage 2.3 – Comparison of Results with Alternative Methods

The results obtained in Substages 2.1 and 2.2 were compared with other methods to assess the adequacy of the model.

- Application of reference methods (benchmarks): revenue forecasts were initially generated using the less complex benchmark methods Seasonal Naïve, Naïve, Drift, and Mean, following Dhakal (2017).

- Accuracy assessment: MAPE values were calculated for each benchmark model.

- X13-ARIMA-SEATS modeling: an ARIMA model was generated using the X13-ARIMA-SEATS program.

- Sample division: the time series was divided into training and test datasets for forecast validation.

- Model generation and adjustment: multiple time-series models were generated, and each series was individually adjusted using theseas function.

Regarding the X13-ARIMA-SEATS program, Mattos and Ferreira (2020, p. 146) assert that one of its features that “deserves to be highlighted is the pre-adjustment of the time series, that is, a correction performed before seasonal adjustment actually takes place.” This functionality allows for the treatment of non-seasonal outlying observations that may nevertheless affect seasonality patterns. One limitation of the present study is that no prior adjustment for working days (among other specific dates) was performed before the automatic adjustment procedure. Consequently, reversing the sequence of applying the Fable and Seasonal packages should be considered in future research. It is important to note that, if, although the X13-ARIMA-SEATS framework models only one series at a time (Mattos; Ferreira, 2020), its advantage lies in the fact that all residual diagnostics are performed within the same function series function (“basededados_ajustada”, forecast.forecasts).

It is noteworthy, in this context, that the modeling process must be iterative the agencies and respective time series selected for analysis should undergo continuous testing, particularly with respect to residual diagnostics. Moreover, the entire modeling process should be based on the “underlying theory and the researcher's experience” (Fávero; Belfiore, 2021, p. 511) and evaluated considering the specific circumstances faced by each user of the methodology.

4 RESULTS AND DISCUSSION

This section is divided into two subsections: the first presents the results of the municipality clustering process, while the second reports the results of the time-series modeling of revenues for the municipalities analyzed.

4.1 Results obtained in Stage 1 – Economic Clustering of Municipalities

The 15 municipalities in the State of São Paulo with the highest Gross Domestic Product (GDP) values are presented in Table 1, in descending order.

Table 1 — Municipalities with the highest GDPs in the State of São Paulo (2020)

Continues

Municipality	GDP (BRL)	Revenue (BRL)	Population
Osasco	63.377.281.663,00	2.855.078.622,94	680.964
Guarulhos	54.753.119.623,00	4.261.460.107,26	1.351.275

Conclusion

Municipality	GDP (BRL)	Revenue (BRL)	Population
Campinas	53.364.538.386,00	5.905.702.404,37	1.175.501
São Bernardo do Campo	39.856.385.608,00	4.363.769.356,61	812.086
Barueri	39.482.997.035,00	3.226.248.352,97	264.390
Jundiaí	39.278.730.712,00	2.284.188.239,98	407.016
São José dos Campos	34.197.969.838,00	3.013.041.783,68	710.654
Paulínia	32.690.552.549,00	1.347.499.218,87	105.037
Ribeirão Preto	31.423.776.118,00	2.949.437.391,30	683.777
Sorocaba	29.554.630.061,00	2.957.377.427,20	658.547
Santo André	25.828.979.709,00	2.638.303.559,53	693.867
Piracicaba	21.130.510.915,00	1.780.043.134,66	389.873
Santos	18.513.709.213,00	2.960.581.983,75	428.703
São José do Rio Preto	16.848.532.921,00	1.996.643.669,15	447.924
Cajamar	16.523.071.678,00	594.102.652,04	77.627

Source: SEADE Foundation (São Paulo, 2020a, 2020b) and TCE-SP (São Paulo, 2020c).

Subsequently, according to what was proposed by Fávero and Belfiore (2021), the clustering results were obtained from the tabulated data. The first five clustering stages are presented in Table 2. Negative numbers represent observation positions in alphabetical order. Positive numbers represent clusters according to the stage in which they were formed. Thus, the first cluster was formed by the municipalities of Ribeirao Preto (position 9 in the spreadsheet) and Sorocaba (position 15) during Stage 1. In Stage 2, the cluster created in Stage 1 was expanded to include São José dos Campos (observation 12). Subsequently, Santo André (observation 13) was incorporated into the same cluster.

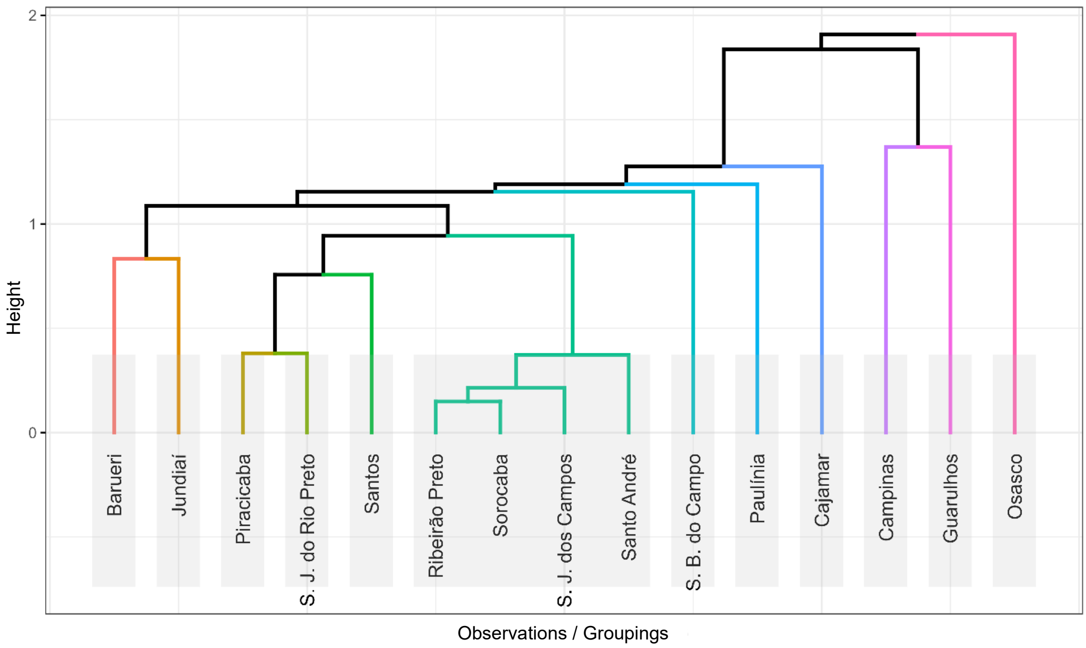
Table 2 — Clustering Stages of the Municipalities with the Highest GDPs in the State of São Paulo (2020)

Stage	Observation 1 / Cluster 1	Observation 2 / Cluster 2	Coefficient
1	-9 (Ribeirão Preto)	-15 (Sorocaba)	0,149
2	1	-12 (São José dos Campos)	0,214
3	2	-13 (Santo André)	0,372
4	-8 (Piracicaba)	-11 (São José do Rio Preto)	0,380
5	4	-14 (Santos)	0,756

Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

The dendrogram presented in Figure 2 graphically summarizes these results.

Figure 2 — Dendrogram of the Clustering of the 15 Municipalities with the Highest GDPs in the State of São Paulo (2020)



Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

The results of the F-test are presented in Table 3.

Table 3 — Analysis of Within-Group and Between-Group Variability for the Clustering of the 15 Municipalities with the Highest GDPs in the State of São Paulo (2020)

Clustering Variable	F-test value	P-valor
GDP	20,8	0,01470
Revenue	73,44	0,00231
Population	332,9	0,00024

Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

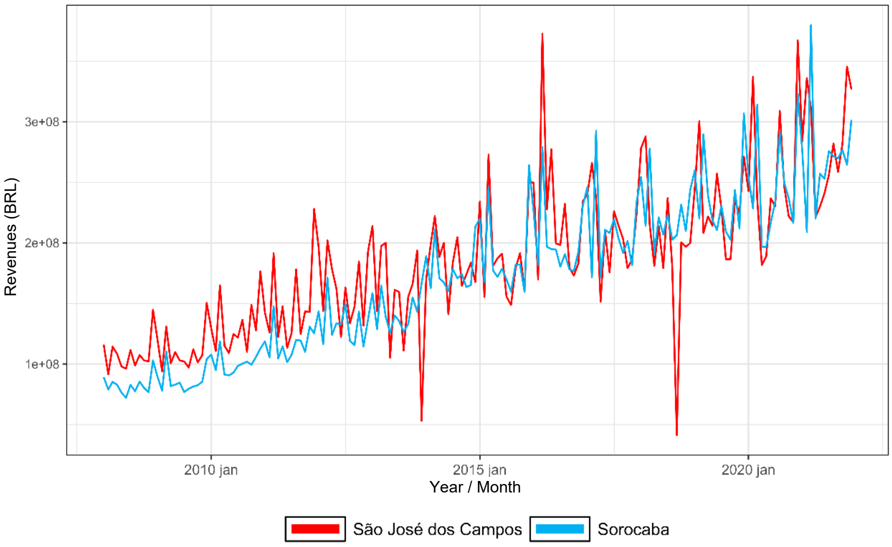
Since all p-values were below 0.05, the selected variables were found to justify the adequacy of the clusters obtained, as proposed by Fávero and Belfiore (2021).

4.2 Results Obtained in Stage 2 – Development of the Forecasting Models

Based on the cluster identified in Stage 1, the municipalities of Sao Jose dos Campos and Sorocaba were selected for the multiple time-series analysis. Ribeirao Preto was excluded because, although its revenue series exhibited a normal distribution, its residuals did not satisfy the normality assumption (Morettin; Toloi, 2018). This requirement is essential to ensure that the generated models capture the full information contained in the dataset while minimizing information loss in the residual component (Ferreira; Mattos, 2020).

Further analysis revealed that the revenues of the two municipalities were relatively similar when compared with those of the remaining municipalities. In December 2020, São José dos Campos recorded monthly revenue of approximately BRL 367 million, whereas Sorocaba recorded approximately BRL 322 million. Based on the economic characteristics of both municipalities, the evolution of revenues from January 2008 to December 2021 was observed. The graphical analysis of the time series (Figure 3) indicates a growth trend in both municipalities. In 2020, there was a drop in revenues, possibly associated with the economic crisis caused by the COVID-19 pandemic, followed by recovery in subsequent years.

Figure 3 — Revenue Time Series of the municipalities of Sao Jose dos Campos and Sorocaba (2008-2021) (in BRL)



Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

Based on the graphical representation of revenue evolution from 2008 to 2021, the forecasting models described in the following sections were generated.

4.2.1 Substage 2.1 – Development of ARIMA and Exponential Smoothing Models

The execution of the Fable package produced the models presented in Table 4.

Table 4 — ARIMA and Exponential Smoothing Models for the Revenue Time Series of Sao Jose dos Campos and Sorocaba (2008-2021)

Continues

Municipality	Modeling Approach	Model Specification
São José dos Campos	ARIMA	(0,1,1) (0,0,1) [12]

		Conclusion
São José dos Campos	Exponential Smoothing	Multiplicative, Additive, Multiplicative
Sorocaba	ARIMA	(1,0,1) (0,1,1) [12] with constant
Sorocaba	Exponential Smoothing	Multiplicative, Additive, Multiplicative

Source: prepared by the authors, based on calculations performed in R (R Core Team, 2023).

For the ARIMA models, São José dos Campos exhibited a first-order integration in the non-seasonal component, whereas Sorocaba exhibited a first-order integration in the seasonal component. This indicates that differencing was required to eliminate unit roots in both seasonal and non-seasonal components, thereby stabilizing the mean and variance of the series and improving forecast accuracy. Subsequently, the algorithm identified the lowest AIC, AICc, and BIC values required for model selection and specification (O'Hara-Wild *et al*., 2023), as reported in Table 5.

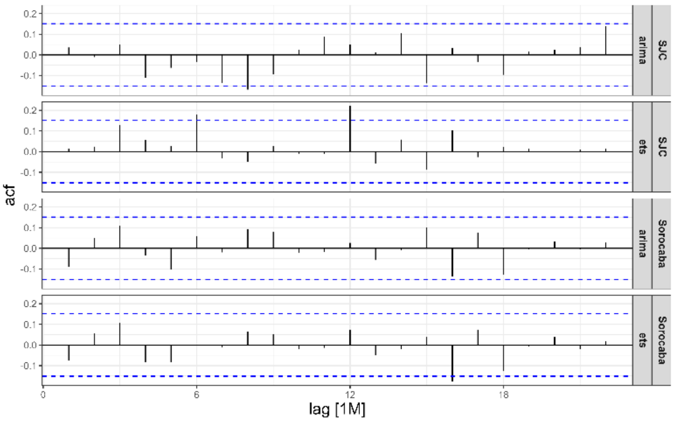
Table 5 — AIC, AICc, and BIC Values for ARIMA and Exponential Smoothing Models of Sao Jose dos Campos and Sorocaba (2008-2021)

Municipality	Modeling Approach	AIC	AICc	BIC	Parameters
São José dos Campos	ARIMA	6313	6313	6322	Ma1=-8,58e-1; Sma1=3,83e-1
São José dos Campos	Exponential Smoothing	6678	6682	6732	Alpha=1,00e-2; Beta=1,04e-4; Gama=1,00e-4.
Sorocaba	ARIMA	5650	5651	5666	Ar1=9,17e-1; Ma1=-8,32e-1; Sma1=-5,78e-1; Constant = 1,22e+6
Sorocaba	Exponential Smoothing	6379	6383	6432	Alpha=1,28e-1; Beta=7,35e-3; Gama=1,77e-4.

Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

Following the verification, the correlogram (Figure 4) showed that the residuals of the ARIMA models for both municipalities crossed the statistical significance bounds. In other words, there was an indication of residual correlation (Ferreira, 2020).

Figure 4 — Correlogram of Residuals of ARIMA and Exponential Smoothing Time-Series Models (2008-2021)



Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

To determine whether this indication would be confirmed, residual diagnostic tests were subsequently performed. As shown in Table 6, all values in the Ljung-Box Test column were greater than 0.05. Therefore, the residuals were not autocorrelated. Next, the Kolmogorov-Smirnov test indicated that the residuals were normally distributed. Finally, the ARCH test, conducted at the 5.0% significance level, showed that the null hypothesis of homoscedastic residual variance could not be rejected. Therefore, for Sao Jose dos Campos, it was not possible to reject the null hypothesis of the absence of linear autocorrelation, nor the assumptions of residual normality and variance homoscedasticity. For Sorocaba, the only difference was that the non-rejection of the homoscedasticity assumption was verified at the 1.0% significance level. The results are reported in Table 6.

Table 6 — Residual Diagnostic Results for ARIMA and Exponential Smoothing Time-Series Models (2008-2021)

Municipality	Modeling Approach	Ljung-Box Test	Kolmogorov-Smirnov Test	ARCHTest
São José dos Campos	ARIMA	0,11	0,08	0,99
São José dos Campos	Exponential Smoothing	0,57	0,48	0,99
Sorocaba	ARIMA	0,31	0,06	0,02
Sorocaba	Exponential Smoothing	0,38	0,88	0,25

Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

In view of these values, all above 1.0%, Mattos and Ferreira (2020) argue that the linear component of the time series may be considered adequately modeled, thereby allowing forecasts to be generated. Consequently, the fourth and final procedure of the modeling process using the Fable package consisted of generating forecasts for 2022 and 2023. The corresponding MAPE values are presented in Table 7.

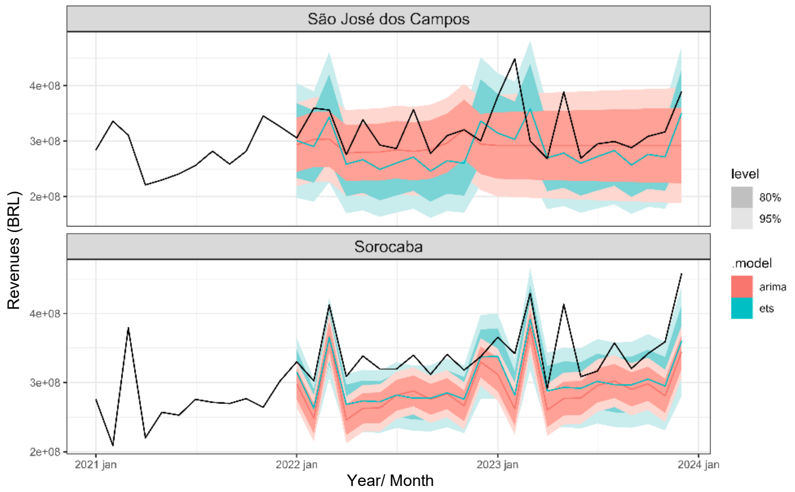
Table 7 — Comparison of Forecast Errors (MAPE) Between ARIMA and Exponential Smoothing Models for Sao Jose dos Campos and Sorocaba (2022-2023)

Municipality	Modeling Approach	MAPE
São José dos Campos	ARIMA	9,75%
São José dos Campos	Exponential Smoothing	13,2%
Sorocaba	ARIMA	15,3%
Sorocaba	Exponential Smoothing	12,3%

Source: prepared by the authors, based on calculations in the R language (R Core Team, 2023).

Based on the results presented in Table 7, it can be concluded that the ARIMA model provided the most accurate revenue forecasts for Sao Jose dos Campos, as evidenced by its MAPE of 9.75%. For Sorocaba, however, the Exponential Smoothing model proved to be the most suitable forecasting approach, achieving a MAPE of 12.3%. The forecast trajectories are illustrated graphically in Figure 5. The black line represents the actual observed values of the time series. For visualization purposes, observations prior to 2021 were omitted from the figure.

Figure 5 — Forecasts of the Sao Jose dos Campos and Sorocaba Time Series Using ARIMA and Exponential Smoothing Models (2022-2023)



Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

After completing all these procedures, it was verified that the Fable package provides all the functionality required to perform multiple time-series modeling. Accordingly, Substage 2.1 was concluded.

4.2.2 Substage 2.2 – Assessment and Validation of Results

In Substage 2.2, the first verification performed after executing the auto.arima function, it revealed that the AIC, AICc, and BIC values obtained through the forecast package were identical to those generated by the Fable package. Next, Subsequently, the statistical significance of the coefficients (Ferreira; Mattos, 2020) was assessed using the BETS:t_test function, from the BETS package (Ferreira *et al*., 2022). The results indicated that all coefficients of the ARIMA models satisfied the significance requirement, as illustrated in Figure 6. It can be observed that the critical value “t Crit.

Value" of 1.97 falls within the interval defined by the statistics calculated for the standard errors "Std. Errors", which prevents rejection of the null hypothesis that the ARIMA model coefficients are statistically equal to zero. Consequently, the reported result is "True."

Figure 6 — Analysis of the Statistical Significance of the ARIMA Model Coefficients

```
> BETS::t_test(city1.treino.autoarima) # # Análise dos Coeficientes
          Coeffs Std.Errors          t Crit.Values Rej.H0
ma1  -0.8579742 0.04284415 20.025470    1.974446   TRUE
sma1  0.3828907 0.06756902  5.666661    1.974446   TRUE
> BETS::t_test(city2.treino.autoarima) # Análise dos Coeficientes
              Coeffs   Std.Errors          t Crit.Values Rej.H0
ar1     9.168114e-01 7.113143e-02 12.888977    1.974535   TRUE
ma1    -8.324925e-01 9.058996e-02  9.189677    1.974535   TRUE
sma1   -5.783509e-01 7.489816e-02  7.721830    1.974535   TRUE
drift   1.220164e+06 1.031064e+05 11.834029    1.974535   TRUE
```

Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

An important caveat should be noted regarding the constant term of the linear function (the drift component). The value estimated by the Fable package may differ from that obtained through the forecast package. This discrepancy is explained in the Fable Package Manual (2023, p. 59): "Forecast standard errors allow for uncertainty in estimating the drift parameter (unlike the corresponding forecasts obtained by fitting an ARIMA model directly)." Furthermore, whenever one or more coefficients are found to be statistically insignificant, the model may be re-estimated through the ARIMA function (Mattos; Ferreira, 2020) by removing the non-significant coefficients from the specification.

Subsequently, the assumptions of residual independence, normality, and homoscedasticity were examined. Using the accuracy function from the Forecast Package, Sao Jose dos Campos presented a MAPE of 9.75%, while Sorocaba presented a MAPE of 12.3% for the ARIMA models. These findings were consistent with those obtained through the Fable modeling procedure. The validation of the Exponential Smoothing models involved a simpler procedure. The models were generated using the ets function, after which the corresponding residual diagnostic tests were performed.

As observed, both the fitted models and the forecasts generated using the Fable package produced identical values and conclusions to those obtained through the forecast package. This result corroborates the fact that both packages rely on the same underlying parameterization framework (O'Hara-Wild *et al*., 2023).

4.2.3 Substage 2.3 – Benchmark Models and Alternative Methods

In Substage 2.3 of the algorithm developed in RStudio, the benchmark forecasting methods Drift, Mean, Naïve, and Seasonal Naïve produced the MAPE values for the 2022–2023 period shown in Table 8.

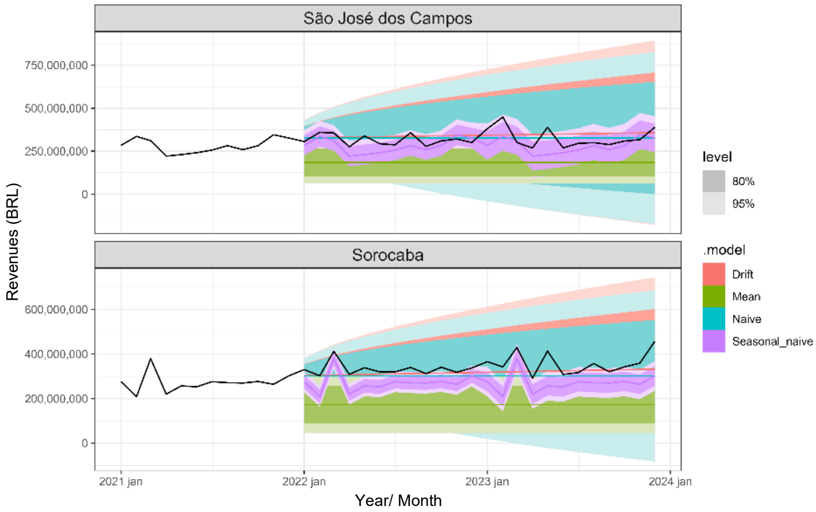
Table 8 — MAPE Statistics for the Revenue Time Series of Sao Jose dos Campos and Sorocaba Using Naïve Methods (2022-2023)

Municipality	Drift	Mean	Naive	Seasonal Naive
São José dos Campos	14,4	41,6	11,7	14,4
Sorocaba	8,69	49,2	11,7	21,3

Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

Considering the results shown in Table 8, the Naïve method, with a MAPE of 11.7%, was the closest to the ARIMA model result for Sao Jose dos Campos, which achieved a MAPE of 9.75%. The Naïve approach uses the most recent observed value in the series as the forecast for future periods. For Sorocaba, the Drift method, with a MAPE of 8.69%, was the closest to the Exponential Smoothing result, which yielded a MAPE of 12.3%. This method incorporates the trend observed in the series when generating forecasts, as illustrated in Figure 7.

Figure 7 — Forecasts of the Revenue Time Series of São José dos Campos and Sorocaba Using Naïve methods (2022-2023)



Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

As can be observed, ARIMA and Exponential Smoothing models are more sophisticated forecasting approaches than the Naïve models, corroborating the findings of Dhakal (2017). Regarding the application of the X13-ARIMA-SEATS program through the Seasonal package, the results are presented in Table 9.

Table 9 — Modeling of the Revenue Time Series of Sao Jose dos Campos and Sorocaba Using X13-ARIMA-SEATS (2008-2021)

Continues

Modeling Element	São José dos Campos	Sorocaba
Model Specification	(0,1,1) (0,1,1)	(0,1,1) (0,1,1)
Estimated Coefficients	Non-seasonal: MA=0.9 Seasonal: MA=0.61	Non-seasonal: MA=0.71 Seasonal: MA=0.73
Information Criteria AIC; AICc; BIC	-; 5793; 5830	-; 5498; 5527
QS Test – Was seasonality detected in the adjusted series?	No	No
Ljung-Box Test – Is residual autocorrelation present?	No	No

		Conclusion
Shapiro-Wilk Test – Are residuals normally distributed?	No	Yes
MAPE Forecasts (2022 and 2023)	7,54%	11,78%

Source: Prepared by the authors, based on calculations performed in R (R Core Team, 2023).

As shown, differences exist between the models identified by the Fable (or Forecast) package and those generated by Seasonal (X13-ARIMA-SEATS). First, the X13-ARIMA-SEATS procedure applied both seasonal and non-seasonal differencing to the series of each municipality, whereas the Fable package required only differencing for the time series of each municipality. Second, the autoregressive component was estimated as zero in the Sorocaba model produced by X13-ARIMA-SEATS. Regarding forecast accuracy, the ARIMA models generated through Fable produced a MAPE of 9.75% for Sao Jose dos Campos and 15.3% for Sorocaba. However, according to the Shapiro-Wilk test, the residuals of the Sao Jose dos Campos model generated by X13-ARIMA-SEATS were not normally distributed at the 5% significance level, indicating the need for additional adjustments under that methodology. For Sorocaba, the Seasonal package automatically applied a logarithmic transformation. Consequently, the ARIMA models generated by the Fable package exhibited higher MAPE values and less favorable information criteria than those obtained with Seasonal (X13-ARIMA-SEATS) for both municipalities. Nevertheless, the Sao Jose dos Campos model produced by X13-ARIMA-SEATS could not be considered directly comparable with the corresponding Fable model because it failed to satisfy the residual normality assumption. Likewise, it could not be used for forecasting or MAPE calculation without additional manual adjustments (Mattos; Ferreira, 2020), which were beyond the scope of this study.

These findings highlight that each forecasting technique is subject to specific assumptions, including data quality and availability, the need to account for anomalous observations (outliers), and the complexity of the algorithms involved. The study also showed that the X13-ARIMA-SEATS models required fewer lines of code than those implemented through the Fable package. However, depending on the characteristics of the series, model specification may become more complex. For Sao Jose dos

Campos, although the ARIMA model estimated through Fable was less accurate than the model generated by seasonal, it was the only model that satisfied all diagnostic assumptions. For Sorocaba, the Seasonal package provided more accurate forecasts. Most importantly, the proposed algorithm was capable of simultaneously modeling the revenue series of two municipalities using two distinct forecasting approaches, ARIMA and Exponential Smoothing. This considerably increased the efficiency of the modeling process and facilitated comparative analyses between governmental entities. Regarding forecasting performance, based on the Forecast Accuracy Scale proposed by Klimberg, Sillup, Boyle, and Tavva (2010), Sao Jose dos Campos achieved high forecast accuracy, with a MAPE of 9.75%, whereas Sorocaba achieved good forecast accuracy, with a MAPE of 12.31%. Furthermore, the MAPE values obtained from both the ARIMA and Exponential Smoothing models developed in this study were lower than the average forecasting error of 17.5% per fiscal year reported by Nascimento and Boente (2022).

5 FINAL CONSIDERATIONS

This study demonstrated the feasibility of a predictive model for the simultaneous analysis of multiple public revenue time series, thereby facilitating budget planning and decision-making in revenue-collecting agencies. The results showed that the Forecast and Fable packages are fully compatible. However, the latter offers the additional advantage of enabling the simultaneous generation of multiple forecasting models, including Exponential Smoothing and ARIMA models. In consequence, rather than fitting Exponential Smoothing and ARIMA models separately to determine which produces the lowest AIC, AICc, BIC, or MAPE values for the data under analysis, users can perform all these procedures through a single algorithm implemented in the Fable package. Furthermore, this process can be applied to multiple municipalities within the same code framework. Another important advantage is algorithmic transparency, which allows analysts to understand each stage of the modeling process and every intermediate result produced.

From a statistical perspective, the conclusions indicate that even municipalities with similar economic characteristics, as identified through cluster analysis, may produce different AIC, AICc, BIC, and MAPE values and, consequently, may require different forecasting approaches. Such a finding may serve as a starting point for further investigations into the reasons why municipalities with similar historical revenue trajectories can present different outcomes, as may be observed in financial audit procedures.

Future studies are recommended to explore alternative research hypotheses, such as the application of the proposed methodology to time series that do not exhibit normal distributions. Likewise, from the perspective of cluster formation, other variables may be added to the analysis, in addition to diversifying the sample size.

It should be emphasized that the modeling process is useful not only for generating forecasts but also for understanding the behavior of time series. Moreover, it constitutes an important decision-support tool. Therefore, the use of these forecasting technologies should be complemented by professional expertise, the conceptual foundations of economics, data science principles, and, in the case of governmental applications, compliance with the applicable legal and regulatory framework.

REFERENCES

ABELN, Barend; JACOBS, Jan P. A. M. COVID-19 and Seasonal Adjustment. **Journal of Business Cycle Research**, v. 18. p. 159-169, July 2022. DOI: https://doi.org/10.1007/s41549-022-00071-z. Available at: https://link.springer.com/article/10.1007/s41549-022-00071-z. Accessed: Sept. 23, 2024.

ARAÚJO, Letícia; BEHR, Ariel; SCHIAVI, Giovana Sordi. Adoption of business analytics in accounting. **Revista Contabilidade & Finanças**, São Paulo, v. 34, n. 93, e1771, Dec. 2023. DOI: https://doi.org/10.1590/1808-057x20231771.en. Available at: https://www.scielo.br/j/rcf/a/h8BSKTCgthvCfwCf9dNSCkH/?lang=en. Accessed: Sept. 23, 2024.

BATÓG, Barbara; BATÓG, Jacek. Regional government revenue forecasting: risk factors of investment financing. **Risks**, v. 9, n. 12, p. 1-15, Nov. 2021. DOI: https://doi.org/10.3390/risks9120210. Available at: https://www.mdpi.com/2227-9091/9/12/210. Accessed: Sept. 23, 2024.

BRAZIL. Central Bank of Brazil. **Efeitos da pandemia sobre o ajuste sazonal dos indicadores econômicos**. Special Study No. 108/2021. Brasília, DF: BCB, 2021. Available at: https://www.bcb.gov.br/conteudo/relatorioinflacao/EstudosEspeciais/EE108_efeitos_pandemia_sobre_ajuste_sazonal.pdf. Accessed: Sept. 23, 2024.

BRAZIL. [Constituição (1988)]. **Constituição da República Federativa do Brasil de 1988**. Brasília, DF: Presidência da República, 1988. Available at: http://www.planalto.gov.br/ccivil_03/Constituicao/Constituicao.htm. Accessed: July 27, 2026.

BRAZIL. **Lei No. 4.320, de 17 de março de 1964**. Estatui normas gerais de Direito Financeiro para elaboração e controle dos orçamentos e balanços da União, dos estados, dos municípios e do Distrito Federal. Brasília, DF: Presidência da República, 1964. Available at: https://www.planalto.gov.br/ccivil_03/leis/l4320.htm. Accessed: July 28, 2026.

BRAZIL. **Lei Complementar No. 101, de 4 de maio de 2000**. Estabelece normas de finanças públicas voltadas para a responsabilidade na gestão fiscal e dá outras providências. Brasília, DF: Presidência da República, 2000. Available at: https://www.planalto.gov.br/ccivil_03/leis/lcp/lcp101.htm. Accessed: July 28, 2026.

BRAZIL. **Metodologia**: ajuste sazonal. Brasília, DF: ME, 2020. Available at: https://balanca.economia.gov.br/balanca/IPQ/arquivos/metodologia_ajuste_sazonal.pdf. Accessed: Sept. 23, 2024.

BRAZIL. Secretaria do Tesouro Nacional. **Manual de contabilidade aplicado ao setor público**. 11. ed. Brasília, DF: STN, 2024. Available at: https://www.gov.br/tesouronacional/pt-br/contabilidade-e-custos/manuais/manual-de-contabilidade-aplicada-ao-setor-publico-mcasp-1. Accessed: Feb. 6, 2026.

DHAKAL, Chuda Prasad. A naïve approach for comparing a forecast model. **International Journal of Thesis Projects and Dissertations (IJTPD)**, v. 5, n. 1. p. 1-3, Jan./Mar. 2017. Available at: https://www.researchpublish.com/papers/a-nave-approach-for-comparing-a-forecast-model. Accessed: Sept. 23, 2024.

FAVERET, Guilherme; AGUILAR, Soraida; BAIÃO, Fernanda; SIMÕES, Paulo Henrique; MAÇAIRA, Paula. Estudo de similaridade da evolução da covid-19 nos estados brasileiros. In: Simpósio Brasileiro de Pesquisa Operacional, 53, Joao Pessoa, 2021. **Anais** [...], v. 53. Joao Pessoa: Sociedade Brasileira de Pesquisa Operacional (SOBRAPO), 2021. Available at: https://proceedings.science/sbpo/sbpo-2021/trabalhos/estudo-de-similaridade-da-evolucao-da-covid-19-nos-estados-brasileiros?lang=pt-br. Accessed: Sept. 23, 2024.

FÁVERO, Luiz Paulo; BELFIORE, Patrícia. **Manual de análise de dados**: estatística e modelagem multivariada com Excel®, SPSS® e Stata®. 1. ed. Rio de Janeiro: LTC, 2021.

FERREIRA, Pedro Costa; SPERANZA, Talitha; COSTA, Jonatha; TEIXEIRA, Fernando; MARCOLINO, Daiane. Package 'BETS: Brazilian Economic Time Series'. **The comprehensive R archive network**. Documentation. Reference manual. Version 0.4.9. Sept. 2018. Available at: https://cran.r-project.org/src/contrib/Archive/BETS/. Accessed: Sept. 23, 2024.

FERREIRA, Pedro Guilherme Costa; DUCA, Victor Eduardo Leite de Almeida. Modelos de suavização exponencial. In: FERREIRA, Pedro Guilherme Costa (ed.). **Análise de séries temporais em R**: curso introdutório. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 75-93.

FERREIRA, Pedro Guilherme Costa. **Processos não estacionários**. In: FERREIRA, Pedro Guilherme Costa (ed.). In: FERREIRA, Pedro Guilherme Costa (ed.). **Análise de séries temporais em R**: introductory course. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 95-120.

FERREIRA, Pedro Guilherme Costa; MATTOS, Daiane Marcolino de. **Modelo SARIMA**. In: FERREIRA, Pedro Guilherme Costa (ed.). **Análise de séries temporais em R**: curso introdutório. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 121-143.

GERHARDT, Tatiana Engel; SILVEIRA, Denise Tolfo (ed.). **Métodos de Pesquisa**. Série Educação a distância. Universidade Aberta do Brasil (UAB); Universidade Federal do Rio Grande do Sul (UFRGS) (coord.). Porto Alegre: Editora da UFRGS, 2009. Available at: https://www.ufrgs.br/cursopgdr/downloadsSerie/derad005.pdf. Accessed: July 27, 2026.

HYNDMAN, Rob; ATHANASOPOULOS, George; BERGMEIR, Christoph; CACERES, Gabriel; CHHAY, Leanne; KUROPTEV, Kirill; MÜCKE, Maximilian; O'HARA-WILD, Mitchell; PETROPOULOS, Fotios; RAZBASH, Slava; WANG, Earo; YASMEEN, Farah; GARZA. Forecast: forecasting functions for time series and linear models. **The comprehensive R archive network**. Documentation. Reference manual. Version 8.21.1. 2024. DOI: https://doi.org/10.32614/CRAN.package.forecast. Available at: https://cran.r-project.org/web/packages/forecast/index.html. Accessed: Sept. 23, 2024.

HYNDMAN, Rob J.; KHANDAKAR, Yeasmin. Automatic time series forecasting: the forecast package for R. **Journal of Statistical Software**, v. 27, n. 3, p. 1-22, July 2008. DOI: https://doi.org/10.18637/jss.v027.i03. Available at: https://www.jstatsoft.org/article/view/v027i03. Accessed: Sept. 23, 2024.

HYNDMAN, Rob. J.; ROSTAMI-TABAR, Bahman. Forecasting interrupted time series. **Journal of the Operational Research Society**, v. 76, n. 4, p. 790-803, Sept. 2024. DOI: https://doi.org/10.1080/01605682.2024.2395315. Available at: https://www.tandfonline.com/doi/full/10.1080/01605682.2024.2395315. Accessed: Sept. 23, 2024.

KIM, Sungil; KIM, Heeyoung. A new metric of absolute percentage error for intermittent
demand forecasts. **International Journal of Forecasting**, v. 32. n. 3. p. 669-679, July/Sept. 2016. DOI: https://doi.org/10.1016/j.ijforecast.2015.12.003. Available at: https://www.sciencedirect.com/science/article/pii/S0169207016000121?via%3Dihub. Accessed: March 14, 2025.

KLIMBERG, Ronald K.; SILLUP, George P.; BOYLE, Kevin J.; TAVVA, Vinay. Forecasting performance measures: what are their practical meaning? In: LAWRENCE, Kenneth D.; KILMBERG, Ronald K (eds.). **Advances in business and management forecasting**. 7. ed. Leeds: Emerald Books, 2010. p. 137-147. DOI: https://doi.org/10.1108/S1477-4070(2010)0000007012. Available at: https://www.emerald.com/books/edited-volume/11151/chapter-abstract/80949903/Forecasting-performance-measures-what-are-their?redirectedFrom=fulltext. Accessed: March 14, 2025.

KONIAGINA, Maria N. Forecast of budget revenues from taxes in the context of economy digitalization. **IOP Publishing**, IOP conference series: materials science and engineering, v. 940, 012040. 2020. DOI: https://doi.org/10.1088/1757-899X/940/1/012040. Available at: https://iopscience.iop.org/article/10.1088/1757-899X/940/1/012040. Accessed: Sept. 23, 2024.

LEVY, Karen; CHASALOW, Kyla E.; RILEY, Sarah. Algorithms and decision making in the public sector. **Annual review of law and social science**, v. 17, p. 309-334, Oct. 2021. DOI: https://doi.org/10.1146/annurev-lawsocsci-041221-023808. Available at: https://www.annualreviews.org/content/journals/10.1146/annurev-lawsocsci-041221-023808. Accessed: March 14, 2025.

LIU, Shiyu; LIU, Ou; CHEN, Junyang. Review on business analytics: definitions, techniques, applications and challenges. **Mathematics**, v. 11, n. 4, p. 1-20, Feb. 2023. DOI: https://doi.org/10.3390/math11040899. Available at: https://www.mdpi.com/2227-7390/11/4/899. Accessed: Sept. 23, 2024.

LOPES, Manuela de Mesquita; BRANCO, Verônica Teixeira Franco Castelo; SOARES, Jorge Barbosa. Utilização dos testes estatísticos de Kolmogorov-Smirnov e Shapiro-Wilk para verificação da normalidade para materiais de pavimentação. **Revista Transportes**, São Paulo, v. 21, n. 1, p. 59-66, Apr. 2013. DOI: https://doi.org/10.4237/transportes.v21i1.566. Available at: https://transportes.anpet.org.br/anpet/article/view/566. Accessed: Sept. 23, 2024.

MATTOS, Daiane Marcolino de; FERREIRA, Pedro Guilherme Costa. Seasonal adjustment. In: FERREIRA, Pedro Guilherme Costa (ed.). **Análise de séries temporais em R**: curso introdutório. 1. ed. São Paulo: GEN Atlas; FGV IBRE, 2020. p. 145-163.

MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais**: modelos lineares univariados. Volume 1. 3. ed. São Paulo: Bluncher, 2018.

MORETTIN, Pedro A.; TOLOI, Clélia M. C. **Análise de séries temporais**: modelos multivariados e não lineares. Volume 2. 1. ed. São Paulo: Bluncher, 2020.

MUÑOZ, Laura Alcaide; BOLÍVAR, Manuel Pedro Rodríguez; ARELLANO; Cinthia L. Villamayor. Factors in the adoption of open government initiatives in Spanish local governments. **Government information quarterly**, v. 39, n. 4, p. 1-13, Oct. 2022. DOI: https://doi.org/10.1016/j.giq.2022.101743. Available at: https://www.sciencedirect.com/science/article/pii/S0740624X2200079X?via%3Dihub. Accessed: March 5, 2025.

NASCIMENTO, Maria de Nazaré Souza; BOENTE, Diego Rodrigues. Fatores associados aos erros de previsão orçamentária da receita do setor público. **Administração pública e gestão social**, Universidade Federal de Viçosa, v. 14, n. 2, p. 1-15, Apr./June 2022. Available at: https://doi.org/10.21118/apgs.v14i2.12945. Accessed: Sept. 23, 2024.

O'HARA-WILD, Mitchell; HYNDMAN, Rob; WANG, Earo; CACERES, Gabriel; BERGMEIR, Christoph; HENSEL, Tim-Gunnar; HYNDMAN, Timothy. Fable: forecasting models for tidy time series. **The comprehensive R archive network**. Documentation. Reference manual. Version 0.3.3. 2023. DOI: https://doi.org/10.32614/CRAN.package.fable. Available at: https://cran.r-project.org/web/packages/fable/index.html. Accessed: Sept. 23, 2024.

OLIVEIRA JUNIOR, João Paulo de; LEVINO, Natallya de Almeida. Impacto da pandemia covid-19 nas finanças dos municípios do Nordeste: uma análise da efetividade dos critérios de distribuição do auxílio. **Revista Brasileira de Planejamento e Desenvolvimento**, Curitiba, v. 13, n. 2, p. 369-400, May/Aug. 2024. Available at: https://periodicos.utfpr.edu.br/rbpd/article/view/16231/10175. Accessed: Sept. 23, 2024.

SÁ, Rogério de Oliveira; AMORIM, Junior José; SANTOS, Cleidiany Dias dos; ZAMBRIM, Dejane Arruda de Carli; GOMES, Vallência Maira. **Metodologia utilizada nas estimativas de receitas e despesas (2021-2025)**. Cuiabá: Secretaria de Estado de Fazenda, 2020. Available at: https://www5.sefaz.mt.gov.br/documents/6071037/15333979/Metodologias+utilizadas+nas+proje%C3%A7%C3%B5es+04.09.2020.pdf/38604051-d9da-f912-e9e0-e0150e1c1148. Accessed: Sept. 23, 2024.

SÃO PAULO. Fundação Sistema Estadual de Análise de Dados. **Tabela – PIB 2020**. São Paulo, SP: Fundação Seade, 2020a. Available at: https://repositorio.SEADE.gov.br/dataset/pib-municipal-ate-2021/resource/13af6a0f-e731-4fc7-8664-73e57de8f465. Accessed: Sept. 23, 2024.

SÃO PAULO. Fundação Sistema Estadual de Análise de Dados. **Taxa de crescimento populacional – 2000 a 2050. São Paulo, SP: Fundação Seade**, 2020b. Available at: https://repositorio.SEADE.gov.br/dataset/populacao-por-idade-sexo/resource/2a0551df-ec74-473c-b0c3-387f0f128523. Accessed: Sept. 23, 2024.

SÃO PAULO. Court of Accounts of the State of São Paulo. **Conjunto de dados**. São Paulo, SP: TCE-SP, 2020c. Available at: https://transparencia.tce.sp.gov.br/conjunto-de-dados. Accessed: Sept. 23, 2024.

SÃO PAULO. Court of Accounts of the State of São Paulo. **Projeto Audesp**. Accounting bookkeeping table. São Paulo, SP: TCE-SP, 2007. Available at: https://www.tce.sp.gov.br/sites/default/files/audesp-documentacao/tabelas_escrituracao_contabil_sistema_audesp_versao_janeiro_2007.pdf. Accessed: Sept. 23, 2024.

SARTORIS, Alexandre. **Estatística e introdução à econometria**. 1. ed. São Paulo: Saraiva, 2003.

SAX, Christoph; EDDELBUETTEL, Dirk. Seasonal adjustment by X-13ARIMA-SEATS in R. **Journal of Statistical Software**, Austria, v. 87. n. 11, p. 1-17, Dec. 2018. DOI: http://doi.org/10.18637/jss.v087.i11. Available at: https://www.jstatsoft.org/article/view/v087i11. Accessed: Sept. 23, 2024.

SOUZA, Celso Correia; CARNEIRO JUNIOR, João Bosco Arbués; CRISTALDO, Márcia Ferreira; CASTELÃO, Raul Asseff; FRAINER, Daniel Massen; VIAGNÓ, Hevelyne Henn da Gama. Previsão de produção de cana-de-açúcar e seus derivados, através de ARIMA, no período de 2020 a 2025, na macrorregião Centro-Sul do Brasil. **Research, society and development**, Vargem Grande Paulista, v. 11, n. 4, e6211427048, March 2022. DOI: https://doi.org/10.33448/rsd-v11i4.27048. Available at: https://rsdjournal.org/rsd/article/view/27048. Accessed: March 5, 2025.

STATISTICS KINGDOM. Multiple linear regression calculator. **Statistics Kingdom**, 2017. Application. Available at: https://www.statskingdom.com/410multi_linear_regression.html. Accessed: Sept. 27, 2024.

TOLEDO JÚNIOR, Flávio Corrêa de. Os conteúdos da Lei Orçamentária Anual. **Revista do TCU**, Brasilia, v. 118, n. 1, p. 35-44, May/Aug. 2010. Available at: https://revista.tcu.gov.br/ojs/index.php/RTCU/article/view/231. Accessed: Sept. 28 2024.

VOLPATO, Gilson Luiz. O método lógico para redação. **Revista Eletrônica de Comunicação Informação & Inovação em Saúde**, Rio de Janeiro, v. 9, n. 1, p. 1-14, Jan./Mar. 2015. DOI: https://doi.org/10.29397/reciis.v9i1.932. Available at: https://www.reciis.icict.fiocruz.br/index.php/reciis/article/view/932. Accessed: Nov. 1, 2024.

AUTHOR CONTRIBUTION (CRediT)

Otoniel Arruda Costa: conceptualization; methodology; investigation; data curation; formal analysis; software; writing – original draft; project administration; resources; and visualization.

Henrique Raymundo Góia: methodology; formal analysis; software; review and editing; project administration; supervision; and validation.

FUNDING

This research did not receive funding or specific support from a funding agency, public or private institution, company or non-profit organization.

CONFLICT OF INTEREST

We declare that there is no conflict of interest related to the development, results or publication of this work.

DATA AVAILABILITY

The data that supported the choice of the municipalities object of the research are available in two repositories. The first one is the SEADE Repository, from which the population data of the municipalities of the State of São Paulo in 2020 were obtained, which can be accessed through the link https://repositorio.SEADE.gov.br/dataset/populacao-por-idade-sexo/resource/2a0551df-ec74-473c-b0c3-387f0f128523 (Population Growth Rate – 2000 to 2050), and the GDP data of the municipalities of São Paulo in 2020, through the link https://repositorio.SEADE.gov.br/dataset/pib-municipal-ate-2021/resource/13af6a0f-e731-4fc7-8664-73e57de8f465 (GDP Table - 2020). The second was the portal of the Court of Accounts of the State of São Paulo, through which the data of the annual revenues of the municipalities of São Paulo were obtained, which can be accessed at the link https://transparencia.tce.sp.gov.br/conjunto-de-dados. The programming codes (script) of the modeling proposed in this article are in the Zenodo repository (https://zenodo.org/records/21877515).

ETHICAL APPROVAL

This research does not involve any procedure, data, or participant requiring ethical approval.

PREPRINT PUBLICATION

This work has not been previously made available on any preprint server or platform.

USE OF AI

No generative artificial intelligence tool was used in the development of this work.

WORK REVIEW HISTORY

Submission date: 11/4/2024

Publication date: 9/24/2026

Responsible editor: Alessandra Caballero Brügger Freitas (executive editor)

Link to the full evaluation history (The provided evaluation history is available only in Portuguese.)