Revista

Cadernos de Finanças Públicas

03 | 2026

TesouroNacional

Forecasting São Paulo's ICMS Revenue: An Experiment with Econometrics, Deep Learning, and *Foundation Models* Using *Top-Down* and *Bottom-Up* Approaches.

Vinícius Inamura Yosimoto

Cássio Machiaveli Oishi

ABSTRACT

This article evaluates seven forecasting methodologies for the monthly ICMS revenue of the State of São Paulo under *top-down* and *bottom-up* strategies: SARIMA, SARIMAX in differences, SARIMAX in levels, ECM, LSTM, TimesFM and SCINet. The experiment uses rolling-origin windows, annual and monthly accuracy metrics, and ordinal and cardinal performance comparisons. Results show that models with exogenous variables dominate the *ranking*. Bottom-up ECM and SARIMAX deliver the best annual forecasts, while *bottom-up* SARIMAX also leads monthly accuracy; LSTM is the best-performing *deep learning* model. Overall, the evidence favors econometric specifications with economic covariates and indicates that revenue disaggregation by collection regime improves ICMS forecasting and may inform future IBS forecasting.

Keywords: Forecasting; time series; tax policy; *machine learning*, econometrics.
JEL Classification: C53; H20; H71.

SUMMARY

FIGURES

TABLES

1. INTRODUCTION

Based on data from the São Paulo State Department of Finance and Planning[1] , it is estimated that revenue from the Tax on Operations Related to the Circulation of Goods and the Provision of Services (ICMS) accounted for approximately 82% of São Paulo's tax revenue in 2025. Thus, the ICMS currently stands as the state's primary tax. In addition, pursuant to the Tax Reform enacted by Constitutional Amendment No. 132/2023, the tax bases for the ICMS and the Tax on Services of Any Nature (ISS) will be incorporated into the scope of the Tax on Goods and Services (IBS), through a gradual transition lasting until 2033.

Consequently, sound practices for forecasting ICMS revenue are essential for the planning and management of the state's public accounts.

ICMS tax collection is a phenomenon that is both economic and legal in nature. On the one hand, the amount collected depends on the occurrence of the economic activities that constitute its taxable event, which makes it subject to systemic fluctuations in the economy. On the other hand, legislative changes—such as tax incentives and modifications to the rules governing the calculation and payment of the tax—directly impact the amount collected. As a result (Figure1), the monthly ICMS time series exhibit not only trends and seasonality but also volatility and structural breaks, which make forecasting a challenging task. The advance collection regimes described at Figure1 will be detailed in the section3.1 .

Broadly speaking, for budgetary purposes, the state must present an annual tax revenue forecast. In addition, cash management requires monthly forecasts to prepare the financial schedule and the flow of disbursements throughout the fiscal year. Therefore, this dual mandate demands a methodology that achieves predictive excellence in both its annual and monthly forecasts. Historically, econometric forecasting methodologies have been used for this purpose, particularly models from the SARIMAX family, whose performance has been considered satisfactory.

However, the predictive performance of the current methodology lacks a point of comparison. To the best of our knowledge, the literature does not offer a systematic comparison—validated across multiple forecast windows—that evaluates the predictive performance of econometric methodologies against *deep learning* and *foundation models* in the context of state tax

1 Chapter 3, Table 1.2a
https://portal.fazenda.sp.gov.br/acessoinformacao/Paginas/Relat%C3%B3rios-da-Receita-Tribut%C3%A1ria.aspx.
Accessed on: June 24, 2026.

revenue forecasting.

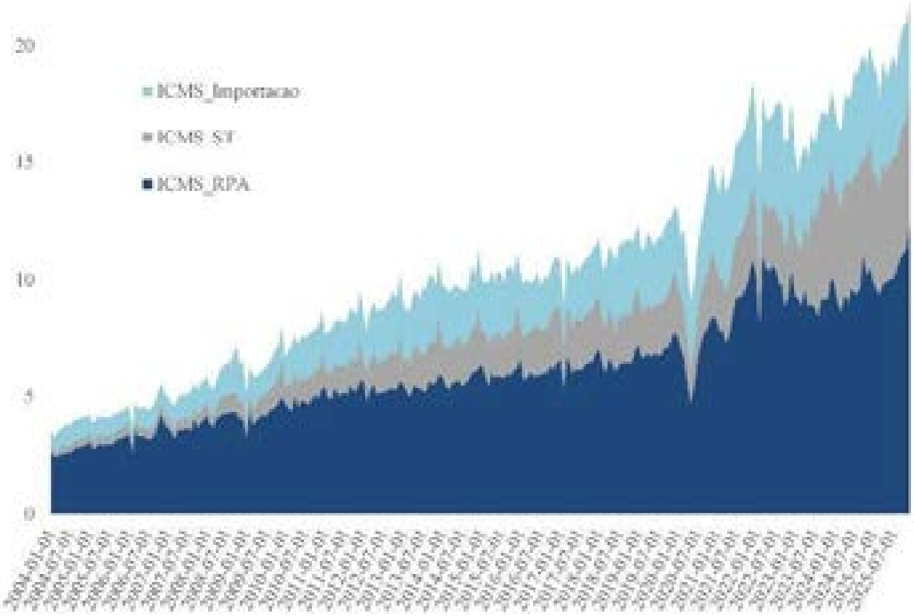
In this section2 , we reviewed recent literature on ICMS forecasting, as well as publications on time series forecasting in general. In the latter case, we focused on comprehensive reviews that compared various forecasting methodologies.

Thus, 14 forecasting methodologies were selected, consisting of the SARIMA, SARIMAX with differences, SARIMAX at level, ECM, LSTM recurrent neural network, SCINet, and TimesFM models, using both *top-down* and *bottom-up* approaches.

As detailed in the section3 , the 14 methodologies were evaluated in an experiment comprising 12 forecasting windows to mitigate any potential temporal selection bias. Subsequently, the resulting predictive performances (section4) were organized into two distinct *ranking* categories (section5).

Consequently, it was concluded that—in general—the most accurate forecasts were obtained by the ECM and SARIMAX econometric models, followed by the LSTM recurrent neural network, particularly under the *bottom-up* approach, which involves decomposing ICMS revenue into collection regimes, each with its own modeling and forecasting, followed by consolidation.

Figure1 —Monthly historical series of the ICMS, by component collection regime (R$ billion).



Source: SEFAZ-SP

2. LITERATURE REVIEW

In repositories such as those of the *Institute of Electrical and Electronics Engineers* (IEEE) and *arxiv.org*, as well as on platforms such as *Google Scholar*, we specifically searched for studies on ICMS revenue forecasting and, more broadly, articles on time series forecasting methodologies published since 2015. First, the objective was to identify time series forecasting methodologies that had already demonstrated promising performance when applied to ICMS revenue. Second, we sought techniques with promising predictive accuracy, even when applied to diverse time series.

The main forecasting methodologies identified can be classified into three broad categories: classical econometrics—which includes SARIMA and SARIMAX—*deep learning*—which encompasses the LSTM recurrent neural network and the SCINet network—and *foundation models*, exemplified by TimesFM.

Five publications on ICMS revenue forecasting were selected, using both classical statistical techniques and *deep learning* models, which will be presented below.

Bonato (2023) studied two approaches for forecasting ICMS tax revenue for the State of São Paulo. The study was based on the hypothesis that the process generating the ICMS time series is nonlinear, due to its structural breaks; therefore, nonlinear models would perform better in forecasting tax revenue. Without citing a specific reference, the author stated that—currently—the São Paulo State Department of Finance and Planning uses *Seasonal Autoregressive Integrated Moving Average Exogenous* (SARIMAX) econometric models to forecast ICMS revenue. This is a linear model in which structural breaks are handled using *dummy* variables. According to the author, this would result in an overfitted model, “ill-suited for long-term forecasts.” However, this hypothesis was not corroborated by the results of his study. As a counterpoint, the author proposed forecasting ICMS revenue using the nonlinear *Long Short-Term Memory* (LSTM) and *Markov-Switching* models. Thus, several forecasts were conducted, using both univariate and multiple regression versions for the two groups of models—linear and nonlinear. IGP-M, IBC-BR, and the U.S. dollar were used as exogenous variables in the multiple regression models. To compare the models’ performance, the *Model Confidence Set* (MCS)—as described by Hansen *et al*. (2011)—was used, with the mean absolute percentage error (, MAPE) as the loss function. Consequently, the resulting *ranking* identified the SARIMAX model with multiple regression as the best model, with a MAPE of 4.7%, followed by the univariate SARIMAX model in second place (MAPE: 4.78%) and the *Markov-switching model* with multiple regression (MAPE: 4.75%) in third. LSTM neural networks were ruled out due to their inconsistent performance and difficulty in parameterization.

Against the backdrop of a financial crisis in the State of Rio de Janeiro, in which the accuracy of ICMS revenue forecasts has become an even more critical factor for public finance planning, Dornelas *et al*. (2022) proposed an experiment using *Long Short-Term Memory* (LSTM) recurrent neural networks to improve the decision-making process of government officials. The dataset comprised monthly values for the variables ICMS revenue, GDP, and commercial electricity consumption in the Southeast Region, from January 2002 through December 2019. Data from 2002 to 2016 were included in the training set; data from 2017 and 2018 were used for validation; and data from 2019 were reserved for testing. The current forecasting methodology of the Rio de Janeiro Department of Finance, which is not specified in this study, yielded relative errors of 9% and 5.83% for the years 2017 and 2018, respectively. The multivariate LSTM model reduced the relative forecasting error to 0.41% and 1.65%, respectively. In a previous and related study, Silva and Figueiredo (2020) used a univariate LSTM network, fed only with values derived from Rio de Janeiro's ICMS tax revenue, such as moving averages and lagged values. Data from 2002 to 2016 comprised the training set, with validation using data from 2017 and 2018 and testing using values from 2019. As a result, a forecast for 2019 was obtained with a relative error of 0.13%, compared to a relative error of 1.49% from the official forecast by SEFAZ-RJ.

SEFAZ/RS published a discussion paper (Dornelles *et al*., 2022) describing the application of an LSTM network for revenue forecasting, inspired by the methodology presented in Silva and Figueiredo (2020). Monthly data from 2009 to 2017 were used to train the model, and data from 2018 to 2020 were used for testing. Finally, an out-of-sample forecast was performed for the period from Jan 2021 to Aug 2021, yielding a relative error of -2.3%.

In Fontenele (2017), SARIMA-type econometric models—selected using the Akaike and Schwarz criteria and estimated with monthly data from Jan 1998 to Dec 2015—were used to forecast 2016 ICMS revenue for the state of Ceará. The models showed greater predictive accuracy in the first 7 months, during which the mean absolute percentage errors (MAPE) were around 1%. Accuracy declined in the remaining months, when the MAPEs exceeded 6%. Thus, the resulting annual absolute relative error hovered around 3.2%.

In addition, studies were identified that used *deep learning* and *foundation models* applied to time series forecasting, as detailed below.

The study by Masini *et al*. (2023) provided a non-exhaustive review of time series forecasting using *machine learning* (ML) models and high-dimensional statistical models. Specifically, an empirical exercise was conducted to forecast the daily variance of the São Paulo Stock Exchange Index (IBOVESPA). Daily data from February 2, 2000, to May 21, 2020, were used, comprising a total of 4,200 observations of the IBOVESPA's daily variance, as well as stock indices from other countries such as the S&P 500 (U.S.), the FTSE 100 (United Kingdom), the DAX (Germany), the *Hang Seng* (Hong Kong), and the *Nikkei* (Japan), which were listed as candidate predictor variables. Additionally, for model estimation or training, a sliding window of 1,500 observations was used to forecast the variance one day ahead. A *Heterogeneous Autoregressive* (HAR) regression model was used as the *benchmark*. As competing models, we proposed a penalized HARX-LASSO regression model—which is an extension of HAR with exogenous variables—a Random Forest regression model, and neural networks with different numbers of hidden layers. To compare the forecast results, we used the QLIKE statistic, according to which no model was able to reject the null hypothesis of equal predictive power relative to the *benchmark*, and the Mean Squared Error (MSE), which indicated that neural networks performed better than the *benchmark* when considering the entire sample, though their predictive power was less pronounced during periods of higher volatility, such as the *Subprime* Crisis (2007–2008) and the COVID-19 pandemic (2020). The HARX-LASSO model was declared the winner among the contenders, as it demonstrated greater predictive accuracy both for the dataset as a whole and for the aforementioned periods of higher volatility. Furthermore, it was concluded that the Random Forest Regression model did not achieve good predictive performance in any period.

A review of *deep learning* models for time series forecasting was presented in Han *et al*. (2021). In the steel industry, managing the flow of energy-carrying gases is a critical operational aspect. Therefore, it is important to forecast the volume of gases generated at each instant. Thus, recurrent neural network (RNN), convolutional neural network (, CNN), and LSTM models—classified as discriminative—were employed, as well as generative models such as *Generative Adversarial Networks* (GAN) and *Deep Belief Networks* (DBN), in addition to a hybrid model that combined an *Autoencoder* (AE) with an LSTM network, to forecast the generation of *Blast Furnace Gas* (BFG) and *Coke Oven Gas* (COG) over a 60-minute period. For BFG generation, the models with the best predictive performance were LSTM (MAPE: 1.42%), DBN (MAPE: 1.87%), and the hybrid model (MAPE: 3.2%). For COG generation, the models with the most accurate forecasts were LSTM (MAPE: 0.35%) and the hybrid model (MAPE: 0.45%).

Li and Law (2024) presented a review of recently developed *deep learning* models for time series forecasting. In addition to describing the models theoretically, the authors compared their predictive performance in an empirical experiment. To this end, two datasets of electrical transformer temperatures were selected: one with hourly values and the other with minute-by--minute values, both spanning a 2-year period. Thus, the Mean Squared Error and the Mean Absolute Error of 9 models with distinct designs were compared. The model with the best predictive performance across both datasets and according to both metrics was SCINet, which combines a convolutional neural network (CNN) and a Seq2Seq-type *encoder-decoder*.

In an approach distinct from those presented above, Das *et al*. (2024) presented a base model, pre-trained on a dataset comprising millions of time series—both historical and publicly available (80%) as well as synthetically generated (20%)—balanced by temporal granularity. Thus, this approach resembles that of *Large Language Models* (LLMs), which are trained on massive text corpora and—upon receiving a textual input (question) that provides context—return a textual output as a response to the posed question. Similarly, the so-called *Foundation Model for Time-Series Forecasting* (TimesFM), pre-trained on a massive corpus of time series, takes a historical time series as input and—given that context—generates a forecast for the specified horizon. To evaluate TimesFM's performance, the authors compared its forecasts with those of statistical and *deep learning* models, using three datasets that were intentionally excluded from its training corpus. It was concluded that TimesFM achieved higher forecast accuracy in two of the datasets and fell within the confidence interval of the three most accurate models in the third dataset.

Therefore, based on the literature review above, SARIMA, SARIMAX, LSTM, SCINet, and TimesFM were selected to be included in an evaluation experiment—in the context of forecasting the ICMS tax for the State of São Paulo—in which their predictive performance will be put to the test, within an isonomic framework consisting of multiple training and testing periods to mitigate any temporal selection bias, and evaluated according to the same predictive accuracy metrics. In addition, the *Error Correction Model* (ECM) was incorporated into the evaluation for the reasons presented in the section "3.2 ," despite not having been included in the literature review.

3. METHODOLOGY

As previously stated, the objective of this study is to evaluate time series forecasting methodologies when applied to the ICMS revenue series for the state of São Paulo. Thus, the predictive accuracy of the methodologies under analysis will be evaluated using increasing sliding windows, as detailed below. Broadly speaking, the idea is to repeat the methodological procedure—including data preparation, modeling or hyperparameter tuning, model estimation or training, and forecasting 12 months ahead—using distinct and progressively larger training sets. Thus, the evaluation is not restricted to a single selection of training and test sets, thereby mitigating any potential selection bias.

Below is a description of the database, a brief theoretical overview of the evaluated methodologies, and the criteria for evaluating the resulting forecasts.

3.1. Database

The selected database has a monthly frequency and spans from January 1, 2004, to December 1, 2025, totaling 264 periods. A summary of the variables included in the database is provided in Table 1 below.

Table1 —Database Used in the Experiment

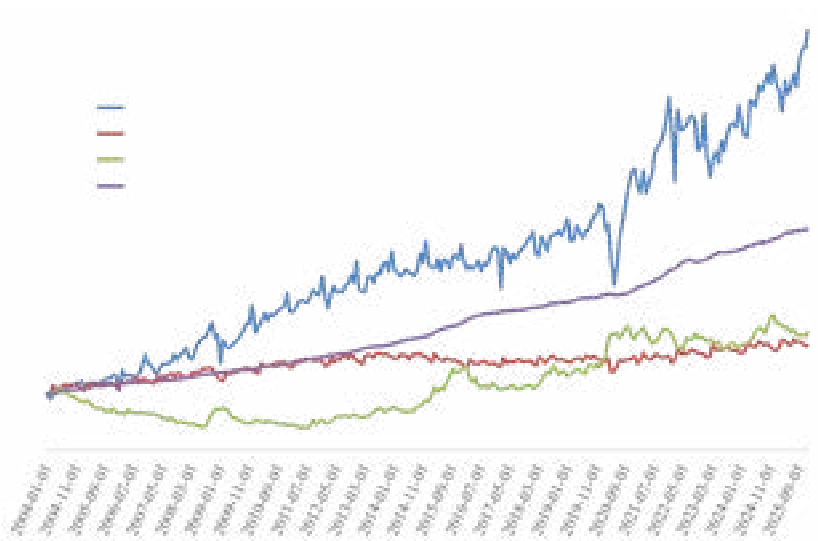
Variable	Description	Remarks	Source
ICMS	ICMS[2] revenue for the State of São Paulo.	Measured in current R$.	Sefaz/SP
ICMS_RPA	ICMS[2] revenue for the State of São Paulo under the Periodic Assessment Regime (RPA), combined with revenue from Simples Nacional and other less significant regimes.	Measured in current R$.	Sefaz/SP
ICMS_ST	ICMS[2] revenue for the State of São Paulo under the Tax Substitution Regime (ST).	Measured in current R$.	Sefaz/SP
ICMS_Importacao	ICMS revenue collected by the State of São Paulo from imports of goods from abroad.	Measured in current R$.	Sefaz/SP
PIBI	Monthly index of Brazil's Gross Domestic Product, calculated using quarterly GDP (IBGE, code 6612), monthly GDP (BCB, code 4380), and the IPCA (BCB, code 433).	Index number, where Jan 2004 = 100.	Compiled in-house, based on data from the IBGE and the BCB
IPCA	National Broad Consumer Price Index (BCB, code 433).	Index number, where Jan 2004 = 100.	IBGE
Exchange Rate	Exchange Rate - Free Market - U.S. Dollar (Selling Rate) - End of Period (code 3696)	Current currency unit (R$)/US$	Sisbacen PTAX800
DU	Monthly number of business days	SGS12_NDIASUTEISPAS12	IPEADATA
Various *Dummies*	Binary variables, which can only take the values 0 or 1.		

2 Amounts collected under Special Installment Payment Programs (PEP) were not taken into account.

Source: Prepared by the author

The ICMS is levied—broadly speaking—on the movement of goods, interstate and intermunicipal transportation services, and communication services (Law 6,374/89). Thus, its scope of application is quite broad, which gives rise to various collection regimes. It is a value-added tax; therefore—generally speaking—its collection is based on the difference between the tax amount levied on outflows (debits) and inflows (credits) under the Periodic Assessment Regime (RPA). Consequently, this system accounts for the majority of ICMS revenue , representing 40.7% of the annual total in 2025. Most taxpayers subject to the RPA pay the tax in the month following the reference month for calculating debits and credits. Consequently, RPA tax revenue for a given month tends to be influenced by economic activity indicators from the previous month. In addition, the ICMS_RPA variable includes payments from taxpayers under the Simples Nacional regime (3.2% of the total), due to the similarity in payment deadlines. Furthermore, less significant payment regimes—whose collected amount represented 9% of the total in 2025—were also consolidated with the RPA. The second most significant collection regime is Tax Substitution (ST), accounting for approximately 26% of tax revenue in 2025. In certain sectors, there is a significant difference in economic concentration among the various links in the value chain. Thus, more concentrated sectors, such as capital-intensive industries, tend to have a smaller number of taxpayers, whose oversight requires fewer public resources. Consequently, taxpayers in more concentrated segments are assigned the responsibility of collecting the tax levied on more dispersed segments, whose oversight would be more costly to the treasury. When this occurs, the ST regime is established. For example, when a business falls under the ST regime, it must collect in advance the ICMS that will be levied on more dispersed segments, such as retail. Consequently, when we take the collection deadline into account, revenue collection under the ST tends to occur simultaneously with the corresponding output. In fact, the ICMS_ST variable is found to be correlated with contemporary economic activity indicators. Finally, there is the regime pertaining to imports, which accounted for 21.1% of tax revenue in 2025. This regime stipulates that payment must be made upon customs clearance. Thus, the ICMS_Importacao variable also exhibits a contemporaneous relationship with economic activity indicators. Furthermore, specifically, revenue from imports is influenced by the exchange rate and the number of business days in the month. The PIBI variable is the economic activity index, constructed from the IBGE's quarterly GDP series, converted to monthly values using the proportion that each month's figure represents in its respective quarter, based on the nominal monthly GDP calculated by the BCB and adjusted by the IPCA through the most recent available date. Finally, the resulting monthly series was transformed into an index number, with the value for Jan/2004 set to 100. *Dummy* variables were used in the econometric models to account for exceptional events or regime shifts. These are binary variables that take on values of 0 or 1. In the case of economic shocks concentrated in specific periods, such as the COVID-19 pandemic, *Additive Outlier* (AO) *dummies* were used, which take on a value of 1 during the period of the event and 0 in the remaining months. To account for one-off calendar effects, in which a significant ICMS collection date is postponed to the following month, the one-off *dummy* difference (AO) was used. For example, in certain years, holidays with variable dates—such as Carnival—can cause the collection to be shifted from one month to the next. Thus, the difference *dummy* would take the value 1 in the month that experienced a reduction in revenue and -1 in the month that saw an increase in revenue. Consequently, its coefficient would be negative and would have an order of magnitude indicating the extent of the revenue transfer from one month to the next. Finally, *Level Shift* (LS) *dummies* were used to capture regime changes, such as amendments to tax legislation that impact ICMS revenue. Thus, the LS *dummy* would take on a value of 0 prior to the regime change and 1 thereafter. Other types of *dummies* will be accompanied by an *ad hoc* description when used.

Figure2 - Time series in index numbers (01/01/2004 = 100)



Source: Author's own analysis, using data from SEFAZ/SP.

To illustrate the monthly variation in the main time series from Table1 , from 2004 to 2025, index numbers were constructed for the variables ICMS (current R$), GDP (R$ Dec.

2025), exchange rate (current R$), and IPCA, where the first value is set to 100 and the others reflect subsequent changes, as shown in Figure2 .

3.2. Forecasting Models

According to the literature review presented in the section2 , econometric, *machine learning*, and foundational models stood out for their excellent predictive performance.

In classical econometrics, Fontenele (2017) achieved good results in forecasting ICMS tax revenue for the state of Ceará using a *Seasonal Autoregressive Integrated Moving Average* (SARIMA) model, which—as based on Box and Jenkins (1970)—includes only autoregressive (AR) and moving average (MA) components, including seasonal ones.

In Bonato (2023), which achieved good accuracy in forecasting ICMS revenue for the state of São Paulo, the *Seasonal Autoregressive Integrated Moving Average with exogenous regressors* (SARIMAX) model stood out; this model—in addition to the components of the SARIMA model—incorporates exogenous variables.

Although not included in the literature review, the *Error Correction Model* (ECM), proposed by Engle and Granger (1987), constitutes a natural alternative for the present forecasting experiment, as it allows for the explicit modeling of the long-run relationship between cointegrated time series through the error correction term (ECT).

It should be noted that only the predictive performance on data not used in the estimation will be considered to assess the value of a forecasting model. Indeed, the evaluation of the statistical properties of the model parameter estimators is relegated to a secondary role. Nevertheless, according to Hyndman and Athanasopoulos (2021), a good forecasting method will produce uncorrelated innovation residuals. This is because, if there is autocorrelation in the residuals, it implies that there is still information to be extracted from the data and used in the forecast. Furthermore, the mean of these residuals must be zero; otherwise, the forecast will be biased. If a forecasting method fails to satisfy these conditions, it indicates that it can be improved. Thus, these properties of the residuals will be considered necessary conditions in the context of model selection.

Among the *deep learning* models, the *Long Short-Term Memory* (LSTM) recurrent neural network and the SCINet network (Liu *et al.*, 2022) stood out. The LSTM network (Hochreiter and Schmidhuber, 1997) achieved remarkable results in forecasting ICMS tax revenue for the states of Rio de Janeiro (Dornelas *et al.*, 2022) and Rio Grande do Sul (Dornelles *et al.*, 2022), as well as in forecasting steel industry gas emissions (Han *et al*., 2021). It incorporates information persistence through a feedback loop, which makes it well-suited for use with data structures that exhibit some sequential pattern, such as time series. The SCINet network, on the other hand, was designed to capture time-series characteristics such as trend and seasonality. Based on a hierarchical tree structure with subsampling, convolution, and interaction operations, it stood out in Li and Law (2024) by presenting the lowest mean absolute error among multiple compared models.

Finally, TimesFM (Das *et al*., 2024) is a *foundation model*, pre-trained on a time series corpus consisting of 300 billion data points, built to perform *zero-shot* forecasts—that is, without the need for any input data preparation, hyperparameter tuning, or even training. The model simply receives a time series as input, which provides it with context, and returns the forecast for the defined time horizon. In Das *et al*. (2024), comparative performance results are presented, in which TimesFM's forecasts rank among those with the lowest mean absolute errors obtained.

3.3. Hyperparameter Tuning and Training of Deep Learning Models

Prior to training, it is necessary to establish the hyperparameter values for *deep learning* models such as LSTM and SCINet networks, which define their architecture. To this end, the *Bayesian Optimization* algorithm (Jones *et al*., 1998) was used, whose search is—roughly speaking—guided by a probabilistic model, with the aim of balancing the exploration of the performance of combinations of hyperparameter values in less-known neighborhoods and the refinement of the search in neighborhoods where the probabilistic model estimates good performance.

In general, once the hyperparameter values have been established, the training of *deep learning* models such as LSTM and SCINet networks occurs iteratively, over successive epochs, during which the model parameters are adjusted to minimize the loss function. To achieve greater robustness and faster convergence, we used the *Adaptive Moment Estimation* (Adam) algorithm (Kingma and Ba, 2015), which sets adaptive and efficient learning rates for each parameter and training epoch.

In addition, to mitigate overfitting, early stopping was employed, which halts training when the marginal decrease in the loss function is consistently negligible. This reduces the risk of the model becoming overfitted to the training set, thereby improving its predictive performance on out-of-sample data.

3.4. Evaluation of Predictive Performance

It should be reiterated that the primary criterion for evaluating forecasting methodologies is their predictive performance on data outside the sample used for estimation or training. The following section presents the metrics used to assess forecasting performance, the increasing sliding window framework—whose purpose is to mitigate selection bias arising from specific training or estimation periods versus testing or forecasting periods—as well as the methodology for *ranking* model performance.

With regard to fiscal management, two areas rely heavily on revenue forecasting. For budgetary purposes, the annual tax revenue forecast is a central element of planning, both materially and legally. Therefore, the accuracy of the annual forecast becomes a critical factor. Consequently, the 12-month Aggregate Percentage Error (APE) (Equation 2) was adopted to assess the performance of the annual forecast. In addition, cash management is directly impacted by the accuracy of the monthly revenue forecast. Consequently, the mean absolute percentage error (Equation 1), known as MAPE (*Mean Absolute Percentage Error*), was also assessed to reflect the monthly accuracy of the forecast.

$$MAPE = \frac{1}{H}\sum_{H}^{1}\left(\frac{Y_t - \hat{Y}_t}{Y_t}\right)$$

(Equation 1)

$$EAPA = \frac{\sum_{H}^{1} Y_t - \sum_{H}^{1} \hat{Y}_t}{\sum_{H}^{1} Y_t}$$

(Equation 2)

H: number of periods in the forecast horizon (H=12)

Y_i: value of the series at time t

$\hat{Y}_t$: value predicted by the model for time t

When evaluating the performance of forecasting models—whether econometric or *machine learning*—there is a training dataset and a test dataset. Thus, broadly speaking, the training dataset is used to estimate or train the models, and the test dataset is then used to make forecasts and evaluate their accuracy. However, the evaluation result depends on the choice of training and test sets, which can lead to selection bias.

When dealing with a time-series data structure, in which the data has a chronological order or temporal dependence, it is not possible to randomly segment the data into training and test sets without the risk of temporal leakage. This is because, in a random selection of training and test sets, there is a risk of violating the chronological order of the data. Thus, there is a risk of using future data to predict the past.

Consequently, validation is performed using sliding windows of increasing size. In this approach, an initial training window is defined from period t_1 to period t_J. The model is then trained using the data from this window, a forecast is made H points ahead—for the period from t_{J+1} to t_{J+H}, and the error is calculated. Next, the estimation window shifts forward by one period, extending from the period t_1 to the period t_{J+1}, and the training process is repeated. Meanwhile, the forecast window shifts forward by one period to the interval from t_{J+2} to t_{J+2+H} and so on. The iterations continue until the last period of the forecast horizon reaches t_T, where T is the total number of periods in the time series. Finally, the average of the errors resulting from each iteration is calculated. This mitigates temporal selection bias and provides a more robust assessment of the models' predictive performance.

In addition to the aforementioned predictive accuracy metrics, to rank the performance of the forecasting methodologies, we considered the relative position of each methodology's forecast in relation to the *ranking* of their performance—from the smallest to the largest predictive error—at each forecasting step, across all the windows described in 3.5.

For each model, 12 forecasts of 12 steps each were obtained. Thus, for example, in the case of monthly forecasting, using a loss function that measures forecast error, the performance of each methodology was ranked at each step of each window.

Furthermore, to at least partially mitigate the effect of having relatively few forecast windows—and thereby improve the quality of the inference—stationary *bootstrap* resampling was performed (Politis and Romano, 1994).

Thus, based on the *ranking*s obtained, the *Model Confidence Set* (MCS), proposed in Hansen *et al*. (2011), was used to identify the set of models whose predictive performance is statistically indistinguishable at a given significance level. Next, the Friedman test (1937) was performed to assess the homogeneity of performance within this set, and the mean rank of each model was used to order the remaining forecasts when heterogeneity was observed. The post hoc comparison by Nemenyi (1963) was calculated, though it had low power to distinguish the predictive performances of the models.

3.5. Experiment

The experiment was designed to ensure the comparability of the forecasting methodologies by establishing uniform parameters for the training window, forecast horizon, number of sliding steps, and total number of slides. In addition, we managed to obtain the maximum number of slides, given the computational constraints, so that the experiment could be carried out. The forecasting methodologies using the LSTM and SCINet *deep learning* models proved to be computationally intensive, with execution times of up to 4 hours for tests with 12 training and test windows. Consequently, the experiment to evaluate predictive performance was limited to 12 increasing sliding windows.

Consequently, the initial training window spanned from January 1, 2004, to January 1, 2024, and the first 12-month forecast covered the period from February 1, 2024, to January 1, 2025. , the training window was extended by one period, reaching February 1, 2024, and the 12-month-ahead forecast shifted forward, beginning on March 1, 2024. And so on, until the forecast horizon reached December 1, 2025, so that the final testing period coincided with the 12 months of 2025, the last calendar year for which revenue data was fully available. Consequently, the experiment consisted of 12 increasing estimation or training windows and 12 sliding forecast and test windows, with 1-month shifts and 12-month forecast horizons.

By design, the SARIMA econometric model and the TimesFM *foundation model* do not allow for exogenous variables, as described in the section “3.2 ,” and, for the SCINet network, they were not used due to computational constraints. In contrast, the SARIMAX and ECM econometric models, as well as the LSTM recurrent neural network, allowed for the use of exogenous variables. Thus, in these cases, the variables described in the section3.1 were considered as potential exogenous variables.

The target variable is ultimately the total ICMS revenue for the State of São Paulo. Thus, its forecast can be obtained directly by explicitly using it as the target variable in the models, which constitutes the *top-down* or aggregate approach. Alternatively, it is possible to decompose ICMS revenue into the ICMS_RPA, ICMS_ST, and ICMS_Importacao collection regimes, as previously described in the section3.1 , which constitutes the *bottom-up* approach. By establishing a separate model for each regime, this approach allows us to capture the pattern of each specific revenue component, as well as highlight its relationship with distinct sets of exogenous variables. Thus, the forecasts resulting from the three component regimes would be aggregated to obtain the ICMS forecast. Both forecasting strategies were incorporated into the experiment.

In addition, the configuration of the hyperparameter tuning for the LSTM and SCINet networks is detailed below. For both, the *Bayesian optimization* algorithm described in the section3.3 was used. The Gaussian process was initialized with 5 combinations of hyperparameters, and the optimization was limited to 25 additional points. Early stopping was also used, with a patience of 5 iterations.

In the case of the LSTM network, based on (Dornelas *et al*., 2022), the hyperparameter space was restricted to the intervals listed below. The search for the number of neurons in the hidden layer was limited to between 32 and 700, the dropout rate to between 5% and 75%, and the sequence size between 3 and 24. In addition, the *ReLU* activation function and the Adam optimizer with an initial learning rate of 0.001 were used (section3.3).

The structure of the SCINet network imposes an intrinsic constraint on two hyperparameters. The size of the input sequence, or *look-back window,* must be divisible by 2^L, where L is the number of levels in the binary tree of *SCI-Blocks*. Thus, a range for L between 1 and 3 was established. Consequently, the *look-back window* must be divisible by 8 (2^3). Accordingly, an input sequence size of 24 was adopted. To find the values of the other hyperparameters that minimize the loss function, the *Bayesian optimization* algorithm was used. The search space for these hyperparameters was delimited according to the following intervals, based on (Liu *et al*., 2022). Within the SCI block, the size of the convolution filter and the hidden layer was limited to between 3 and 12, and the dropout rate to between 20% and 72%. The initial learning rate was restricted to between 0.00005 and 0.002, and the *batch* size to between 2 and 10. In theory, it is possible to stack more than one SCINet network. However, the implementation provided by the authors (Liu *et al*., 2022) only allows for a stack with a single SCINet network. Thus, it was not possible to conduct experiments with a stack size greater than 1.

The TimesFM model does not allow for hyperparameter tuning or training, as it is a pre--trained base model.

The SARIMA, SARIMAX, and ECM econometric models were implemented using components from the *Statsmodels* library. Exogenous variables were selected based on economic relevance and statistical significance, when applicable, provided that the cointegration relationship was established. ICMS tax revenue presupposes the conduct of economic activities that fall under the tax's taxable event, which is reflected in the Gross Domestic Product (GDP). In addition, changes in the prices of taxed goods and services also impact the amount of tax collected. In the case of imports, the exchange rate directly affects the amount of tax collected. Thus, the inclusion of these three macroeconomic variables—GDP, the IPCA, and the dollar exchange rate—was considered during the econometric modeling. In addition, autoregressive (AR) and moving average (MA) components, as well as their seasonal variants (SAR and SMA), were tested based on the autocorrelation function (ACF) and the partial autocorrelation function of the residuals (PACF). As a result, the models should produce residuals that behave like white noise. However, the ultimate test of a predictive model's value is the accuracy of its forecast for data outside the dataset used to estimate its parameters.

The LSTM recurrent neural network was implemented using components from the *tensorflow.keras* library (*Sequential, Input*, LSTM, *Dropout,* and *Dense*). Hyperparameter tuning utilized the *Bayesian Optimization* classes from the *bayes_opt* library. In addition, the *Early Stopping* class from the *tensorflow.keras* library was used during model training to incorporate early stopping.

The foundational TimesFM model is implemented using the *timesfm* library, whose classes load the pre-trained model from a *Hugging Face* repository, take a time series as input, and perform the prediction.

For the SCINet network, the authors' own implementation (Liu *et al*., 2022) was used, available at https://github.com/cure-lab/SCINet (accessed on: 12/17/2025). In addition, the *torch* library was used to prepare the input tensors and to incorporate the Adam optimizer.

Finally, it should be noted that simulating a real-world forecasting scenario was established as an axiom of the experiment. Thus, in forecasts using methodologies that incorporated exogenous variables, their expected future values at the time were considered, as recorded by market expectations in the Brazilian Central Bank's (BCB) Focus Report.

4. RESULTS

As described in the section3 , 14 forecasting methodologies were evaluated, comprising 7 models—SARIMA, SARIMAX with differences, SARIMAX at the level, ECM, LSTM network, SCINet network, and TimesFM—using both *top-down* and *bottom-up* approaches across 12 training and testing windows. The results obtained across the 12 windows were recorded in Table8 and summarized as the mean and standard deviation observed for EAPA and MAPE, as presented in Table7 , both of which are located in the Appendix (section7).

Next, we analyze the characteristics of the time series included in the study and discuss the specifications of the models used, as well as the predictive performance observed for each of them.

4.1. Econometrics

Based on the literature review (section2) and the time series comprising the database used (section3.1), three econometric models were selected: SARIMA, SARIMAX, and ECM.

As will be shown below, all ICMS revenue time series are first-order integrated. Thus, their first differences are stationary, which makes it feasible to estimate SARIMA models, whose parameter estimators are consistent, asymptotically unbiased, and efficient (Box and Jenkins, 1970). Furthermore, for combinations of time series that exhibit a cointegration relationship, it is appropriate to use the SARIMAX model, both for first differences and for the level. On the one hand, at the first-difference level, the variables become stationary, and consequently, the parameter estimators benefit from consistency and asymptotic efficiency; however, the long-run relationship between the variables is no longer explicitly modeled. On the other hand, at the level, non-stationarity leads to superconsistency—that is, the estimators of the cointegration coefficients converge to the true value at a faster rate than normal—yet they are biased, meaning they are no longer efficient, and inference using the t and F statistics is no longer valid (Engle and Granger, 1987). If the focus were on the parameter values, the level-based SARIMAX model would be ruled out. However, since the present study aims to evaluate predictive performance, the level-based SARIMAX model will be included in the experiment and evaluated for its predictive accuracy outside the estimation sample. Finally, by explicitly modeling the long-run relationship between the cointegrated variables, as well as the short-run deviations, the ECM emerges as a natural choice, whose cointegration relationship estimators are superconsistent, asymptotically unbiased, and efficient (Engle and Granger, 1987), provided that weak exogeneity holds.

According to Hyndman and Athanasopoulos (2021), a good forecasting method will produce uncorrelated innovation residuals. If there is autocorrelation in the residuals, this implies that there is still information to be extracted from the data and used in the forecast. Furthermore, the mean of these residuals must be zero; otherwise, the forecast will be biased. Failure to satisfy these two conditions indicates that a forecasting method can be improved. Thus, these properties of the residuals were considered necessary conditions in the model selection process.

As previously described in the section "3.1 ," the time series included as dependent variables are the ICMS tax revenue for the State of São Paulo, as well as revenue broken down by the component regimes: RPA (ICMS_RPA), Imports (ICMS_Importacao), and ST (ICMS_ST). In addition, the GDP, IPCA, and exchange rate time series were incorporated as exogenous variables.

First, the order of integration of the aforementioned time series was assessed using the ADF (Said and Dickey, 1984), PP (Phillips and Perron, 1988), and KPSS (Kwiatkowski *et al*., 1992) hypothesis tests. Thus, a time series was considered stationary when the null hypotheses of the ADF and PP tests were rejected and the null hypothesis of the KPSS test was not rejected at a 5% significance level.

It was concluded that the time series in the database described in the section "3.1 " are integrated of order 1, I(1), with the exception of the IPCA, for which the tests yielded contradictory and, consequently, inconclusive results.

In models with exogenous variables, it is essential to ensure cointegration between the endogenous and exogenous variables to avoid spurious regression. Thus, the combinations of dependent and independent variables used below were evaluated using the Johansen (1988) test. The results can be found in Table9 (Appendix).

In addition, the nonlinearity of the time series was assessed using the BDS test (Brock *et al*., 1987), applied to the residuals of ARIMA models preliminarily fitted to the time series based on information criteria. Furthermore, their volatility or heteroscedasticity was examined, since—in addition to affecting the statistical properties of the estimators—it is also a characteristic that can be mitigated by logarithmic transformation. The assessment of heteroscedasticity was conducted at two levels: the first using the ARCH-LM test (Engle, 1982) —which detects autoregressive conditional heteroscedasticity—and the second by the White test (1980)—which captures general heteroscedasticity—applied only when the first level does not reject the null hypothesis of homoscedasticity for all lags.

Thus, evidence was found that the time series exhibit nonlinearities, as the null hypothesis that the resulting residuals behave as white noise was rejected. Furthermore, the logarithmic transformation did not affect the result of the BDS test. With regard to heteroscedasticity, the null hypothesis of homoscedasticity was rejected at the 5% significance level in all cases. However, the logarithmic transformation proved effective in mitigating the heteroscedasticity, either by increasing the observed p-value or by failing to reject the hypothesis of homoscedasticity.

Thus, based on the above analysis of the time series, the ARIMA, SARIMAX, and ECM models will be specified below for the experiment to forecast ICMS revenue in the State of São Paulo.

The application of the Box and Jenkins (1970) methodology assumes that the time series are stationary. Through the tests described above, it was found that all ICMS revenue series (ICMS, ICMS_RPA, ICMS_Importacao, and ICMS_ST) are integrated of order 1. Thus, the first difference of each series was taken to estimate the SARIMA models. In addition, to mitigate the effects of time series volatility, a logarithmic transformation was applied.

We began with the *top-down* approach, in which the SARIMA model was estimated for the total ICMS revenue series specified at Equation 3 (TD_SARIMA).

$$\left(1-\Phi_1 L^{12}\right)(1-L)ln\left(ICMS_t\right)=c+\left(1+\theta_1 L\right)\left(1+\Theta_1 L^{12}\right)\varepsilon_t+(1-L)Dummies+\delta_5 D_{202004}$$
$$Dummies=\delta_1 D_{200602}+\delta_2 D_{200901}+\delta_3 D_{201703}+\delta_4 D_{202202}$$

(Equation 3)

The *dummies* modeling specific events have the syntax D_{AAAAMM}, where AAAA represents the year and MM the month for which the *dummy* takes the value 1, and 0 in all other periods.

Across the 12 rolling windows (Table8) described in the section3.5 , the SARIMA model achieved an average EAPA of 3.13% and an average MAPE of 3.63% (Table7).

Next, the performance of the SARIMA model was evaluated using a *bottom-up* approach, with the forecast segmented into three models—one for each revenue collection regime—according to the specifications contained in Equations 4 ,5 , and 6 (BU_SARIMA).

$$\left(1-\phi_1 L\right)\left(1-\Phi_1 L^{12}\right)(1-L)ln\left(ICMS\_RPA_t\right)=c+\left(1+\theta_1 L\right)\left(1+\Theta_1 L^{12}\right)\varepsilon_t+(1-L)D_{RPA}$$
$$D_{RPA}=\delta_1 D_{200602}+\delta_2 D_{200901}+\delta_3 D_{201703}+\delta_4 D_{202202}$$

(Equation 4)

$$(1-L)ln\left(ICMS\_Importacao_t\right)=c+\left(1+\theta_1 L\right)\left(1+\Theta_1 L^{12}\right)\varepsilon_t+\delta(1-L)D_{200403}$$

(Equation 5)

$$\left(1-\phi_2 L^2\right)\left(1-\Phi_1 L^{12}\right)(1-L)ln\left(ICMS\_ST_t\right)=c+\left(1+\Theta_1 L^{12}\right)\varepsilon_t+\delta(1-L)D_{200901}$$

(Equation 6)

As a result, a deterioration in predictive performance was observed, with the average EAPA rising to 4.11% and the MAPE to 4.71%.

Under the *top-down* approach, using first differences, the resulting SARIMAX model specification was that shown in Equation 7 (TD_SARIMAX_d). Since the time series—both endogenous and exogenous—exhibit volatility, their logarithmic transformations were used, a technique that will be repeatedly employed throughout the study, though without further mention for the sake of conciseness.

$$\left(1-\Phi_1 L^{12}\right)(1-L)ln\left(ICMS_t\right)=c+L(1-L)\left[\beta_1 ln\left(PIBI_t\right)+\beta_2 ln\left(IPCA_t\right)\right]+\left(1+\theta_1 L\right)\left(1+\Theta_1 L^{12}\right)\varepsilon_t$$

(Equation 7)

After running the 12 increasing sliding forecast windows, an average EAPA of 1.95% was observed, with a standard deviation of 1.63% (Table7), which represents a significant improvement in the accuracy of the annual forecast compared to the SARIMA model. The MAPE for the monthly forecast, on the other hand, had a mean of 3.48% and a standard deviation of 0.99%.

As previously noted, the ICMS is levied on a broad spectrum of economic activities across multiple links in the value chain. Furthermore, in principle, it is a value-added tax (VAT). Thus, changes in ICMS revenue are closely related to changes in the Gross Domestic Product (GDP) of the State of São Paulo, which—in turn—is *proxied* by changes in Brazil's GDP. Consequently, the GDP index (PIBI) is used as an explanatory variable for the three tax regimes, although with different time lags, as detailed in the section3.1 . Furthermore, the ICMS levied on imports is directly and indirectly affected by the exchange rate. On the one hand, an increase in the exchange rate leads to a simultaneous increase in the amount of ICMS collected. On the other hand, the same increase may also reduce the volume of imports in subsequent periods. For this reason, the econometric model for the ICMS_Importacao regime includes the exchange rate variable in two distinct components: one contemporaneous and one lagged by four periods. The number of business days in the month (BD) also serves as a *driver* in the model relevant to the ICMS_Importacao regime.

Thus, moving on to the *bottom-up* approach, the models for each tax collection system were specified according to the equations8 , 9 , and10 (BU_SARIMAX_d).

$$\left(1-\Phi_1 L^{12}\right)(1-L)ln\left(ICMS\_RPA_t\right)=c+L(1-L)\beta_1 ln\left(PIBI_t\right)+\left(1+\theta_1 L\right)\left(1+\Theta_1 L^{12}\right)\varepsilon_t+(1-L)D_{RPA}+\delta_5 D_{202004}$$
$$D_{RPA}=\delta_1 D_{200602}+\delta_2 D_{200901}+\delta_3 D_{201703}+\delta_4 D_{202202}$$

(Equation 8)

$$\left(1-\Phi_1 L^{12}\right)(1-L)ln\left(ICMS\_Importacao_t\right)=c+(1-L)\left[\beta_1 ln\left(PIBI_t\right)+\beta_2 ln\left(Câmbio_t\right)+\beta_3 ln\left(DU_t\right)+\beta_4 L^4 ln\left(Câmbio_t\right)\right]+\left(1+\theta_1 L\right)\varepsilon_t$$

(Equation 9)

$$\left(1 - \phi_1 L\right)\left(1 - \Phi_1 L^{12}\right)(1 - L) ln\left(ICMS\_ST_t\right) = c + (1 - L)\left[\beta_1 ln\left(PIBI_t\right) + \delta D_{200901}\right] + \left(1 + \theta_1 L\right)\left(1 + \Theta_1 L^{12}\right)\varepsilon_t$$
(Equation 10)

Consequently, in first differences, the *Bottom-Up* approach resulted in a deterioration in predictive performance (Table7), with an increase in the average EAPA to 2.81% and a slight decrease in the average MAPE to 3.42% across the 12 estimation and forecasting windows.

The ICMS, PIBI, and IPCA time series are cointegrated (Table9). Thus, as explained earlier, this allows for the specification of a level-based SARIMAX model, as per Equation 11 (TD_SARIMAX), despite the non-stationarity of the time series.

$$\left(1 - \phi_1 L\right)\left(1 - \Phi_1 L^{12}\right) ln\left(ICMS_t\right) = c + \beta_1 ln\left(PIBI_t\right) + \beta_2 ln\left(IPCA_t\right) + \left(1 + \theta_1 L\right)\left(1 + \Theta_1 L^{12}\right)\varepsilon_t$$
(Equation 11)

In fact, under the *Top-Down* approach, the level-based SARIMAX model produced more accurate forecasts (Table8) than those based on differences, with a mean EAPA of 1.74% and a mean MAPE of 2.6%.

Furthermore, the SARIMAX model was estimated at the level using a *bottom-up* approach, according to Equations12 ,13 , and14 (BU_SARIMAX). It should be noted that the *dummy* variable D_{ST} reflects a change in the payment deadlines for the Tax Substitution regime that occurred in 2016.

$$\left(1 - \phi_1 L\right)\left(1 - \Phi_1 L^{12}\right) ln\left(ICMS\_RPA_t\right) = c + \beta_1 L ln\left(PIBI_t\right) + \left(1 + \theta_1 L\right)\varepsilon_t + (1 - L) D_{RPA} + \delta_4 D_{201703} + \delta_5 D_{202004}$$
$$D_{RPA} = \delta_1 D_{200602} + \delta_2 D_{200901} + \delta_3 D_{202202}$$
(Equation 12)

$$\left(1 - \phi_1 L\right)\left(1 - \Phi_1 L^{12}\right) ln\left(ICMS\_Importacao_t\right) = c + \beta_1 ln\left(PIBI_t\right) + \left[\beta_2 + \beta_4 L^4\right] ln\left(Câmbio_t\right) + \beta_3 ln\left(DU_t\right) + \left(1 + \theta_1 L\right)\varepsilon_t + \delta_1 D_{200805} + \delta_2 D_{202011}$$
(Equation 13)

$$\left(1 - \phi_1 L - \phi_2 L^2 - \phi_3 L^3\right)\left(1 - \Phi_1 L^{12}\right) ln\left(ICMS\_ST_t\right) = c + \beta_1 ln\left(PIBI_t\right) + \delta_1 D_{200901} + \delta_2 D_{ST} + \delta_3 D_{12} + \varepsilon_t$$
(Equation 14)

Consequently, the predictive performance of the above models represented an improvement over the *Top-Down* approach, with an average EAPA of 1.4% and an average MAPE of 2.03%, as shown in Table8 (Appendix). It should be noted that, in addition to being more accurate, the forecast from the SARIMAX model using the *bottom-up* approach was also more stable, with standard deviations for EAPA and MAPE of 0.6% and 0.46%, respectively.

Once cointegration between the time series is established, it is known that the cointegration vector estimator is superconsistent (Engle and Granger, 1987), as confirmed by the Johansen test (1988), as shown in Table9 (Appendix). Furthermore, weak exogeneity (Engle *et al.*, 1983) ensures the validity of the inference of the t and F statistics regarding the long-run and short-run parameters, as well as the asymptotic efficiency of their estimators. Considering the time series that make up the cointegration relationship, once a full *Vector Error Correction Model* (VECM) has been estimated, if the coefficient of a variable under test is not statistically significant, it is concluded that the variable is weakly exogenous. However, even if statistically significant, if such a coefficient is small enough, the effects of endogeneity on the efficiency of the estimators become negligible (Stock, 1987; Phillips and Hansen, 1990; Phillips, 1991). In economic terms, it is reasonable to expect that changes in indicators such as GDP, the IPCA, and the exchange rate would have a significant impact on ICMS revenue; however, the reverse is not true. Variations in São Paulo's ICMS revenue are unlikely to be a major explanatory factor for changes in GDP, the dollar exchange rate, or the national price index, a finding corroborated by the small values of their coefficients (Table2). Therefore, even setting aside the hypothesis of weak exogeneity for economic variables such as GDP, the IPCA, and the exchange rate in most models—as can be seen at Table2 — the negligible magnitude of their coefficients (α) suggests a negligible impact of endogeneity on the efficiency of the estimators—a phenomenon we can refer to as "quasi-weak exogeneity"—which justifies the use of the ECM.

Table 2 – Exogeneity

Model	**Variable**	***Rank* r**	**α**	**p-value α**
ICMS	PIBI	3	0.0000	0.00%
ICMS	IPCA_ind	3	0.0000	0.01%
ICMS	Exchange Rate	3	0.0000	0.24%
ICMS_RPA	log(PIBI)	1	0.0133	0.00%
ICMS_RPA	log(IPCA_ind)	1	0.0007	0.29%
ICMS_Importacao	PIBI	3	0.0000	0.08%
ICMS_Importacao	Exchange Rate	3	0.0000	0.00%
ICMS_Importacao	IPCA_ind	3	0.0000	2.49%
ICMS_ST	PIBI	3	0.0000	3.08%
ICMS_ST	IPCA_ind	3	0.0000	0.01%
ICMS_ST	Câmbio	3	0.0000	11.47%

Source: Author's own work

Starting with the *top-down* approach, the error correction term (ECT) and the ECM were modeled using the specifications expressed in Equations15 and16 (TD_ECM), respectively.

$$ECT_t = ICMS_t - \beta_0 - \beta_1 PIBI_t - \beta_2 IPCA_t - \beta_3 Câmbio_t - \beta_4 SLP_{200909} \cdot t - \beta_5 SLP_{202008} \cdot t$$
(Equation 15)

$$\left(1 - \phi_1 L - \phi_2 L^2 - \phi_{12} L^{12}\right)(1 - L) ICMS_t = c + \alpha ECT_{t-1} + L(1 - L)\left[\theta_1 PIBI_t + \theta_2 SLP_{202008} \cdot t\right] + \varepsilon_t$$
(Equation 16)

It should be noted that the variables SLP_{200909} and SLP_{202008} are slope *dummies*, which take the value 1 as of the date they represent and 0 prior to that, depending on the number of periods elapsed since the specified date (t). These variables were used to model structural breaks that occurred following the *Subprime* Crisis and the COVID-19 Pandemic, respectively. In addition, it should be noted that the autoregressive $\Phi_2 L^2$ e component was retained despite its statistical insignificance, as it neutralizes the autocorrelation of the residuals.

In fact, forecasts were obtained for the 12 increasing sliding windows, with an average EAPA of 3.05% and an average MAPE of 3.87% (Table8).

Next, using the *bottom-up* approach, the ECM models were established for the RPA, Import, and ST regimes, according to Equations18 ,20 , and22 (BU_ECM), respectively. It is observed that, for ICMS_ST, there was a structural break around June 2023, resulting from changes in collection rules related to fuel transactions, which led to an increase in tax revenue collected through Tax Substitution. Thus, to model this structural break, the *dummy* variable LS_{202306}, was used, which takes the value 1 starting in June 2023 and zero before that.

$$ECT_t^{RPA} = ICMS\_RPA_t - \beta_0 - \beta_1 PIBI_t - \beta_2 IPCA_t$$
(Equation 17)

$$\left(1 - \phi_1 L - \phi_{12} L^{12}\right)(1 - L) ICMS\_RPA_t = c + \alpha ECT_{t-1}^{RPA} + L(1 - L)\, \theta_1 PIBI_t + \varepsilon_t$$
(Equation 18)

$$ECT_t^{Importacao} = ICMS\_Importacao_t - \beta_0 - \beta_1 PIBI_t - \beta_2 IPCA_t - \beta_3 Câmbio_t - \beta_4 SLP_{201104}$$
(Equation 19)

$$\left(1 - \phi_1 L - \phi_2 L^2 - \phi_{12} L^{12}\right)(1 - L) ICMS\_Importacao_t = c + \alpha ECT_{t-1}^{Importacao} + (1 - L)\left[L^{12}\theta_1 PIBI_t + L\theta_1 Câmbio_t\right] + \varepsilon_t$$
(Equation 20)

$$ECT_t^{ST} = ICMS\_ST_t - \beta_0 - \beta_1 PIBI_t - \beta_2 IPCA_t - \beta_3 LS_{202306}$$
(Equation 21)

$$\left(1 - \phi_1 L - \phi_3 L^3 - \phi_{12} L^{12}\right)(1 - L) ICMS\_ST_t = c + \alpha ECT_{t-1}^{ST} + L(1 - L)\, \theta_1 PIBI_t + \delta_1 D_{12} + \varepsilon_t$$
(Equation 22)

At the end of the 12 rolling windows, it became clear that the forecasting performance of the ECM under the *bottom-up* approach (Table8) surpassed the annual accuracy of the SARIMAX model, with an average EAPA of 1.2%; however, its monthly accuracy deteriorated relative to SARIMAX (2.03%), with an average MAPE of 3.27%. This result is consistent with econometric intuition, given that the ECM explicitly models the long-term relationship between time series through the ECT, while the SARIMAX model features specialized capabilities to capture short-term patterns such as seasonality.

4.2. Deep Learning

Deep learning models have demonstrated good performance in forecasting financial time series (Masini *et al*., 2023), particularly in contexts involving massive amounts of data, whether daily or even hourly. Even in forecasts of ICMS tax revenue, such models have shown promising performance (Dornelas *et al*., 2022). However, the literature still lacks validated evaluations across multiple forecast horizons using monthly data, which tends to be on a smaller scale. The following experiments aim to assess the predictive performance of SCINet and LSTM networks using a dataset with only 264 periods.

The SCINet network has 7 hyperparameters, but only 5 are adjusted for each training window using the *Bayesian Optimization* algorithm, as described in the section3.2 . The resulting hyperparameter values are presented in Table14 , Table15 , Table16 , and Table17 (Appendix).

Under the *Top-Down* approach, the SCINet network achieved an average EAPA of 2.69%, which ranks as the second-best annual performance among univariate models, and an average MAPE of 4.61%, as shown in Table8 (section7).

Under the *bottom-up* approach, however, the SCINet network's predictive performance deteriorated, with an average EAPA of 5.48% and an average MAPE of 7.28%.

The LSTM recurrent neural network has three hyperparameters, which can be considered a modest number compared to SCINet's seven. However, it was scaled with a hyperbolic number of neurons in the hidden layer (*units*), which—through *Bayesian optimization*—can reach a limit of 700, as detailed in the section3.5 , resulting in high processing costs.

For the *top-down* approach, an LSTM network was considered, in which the inputs are the variables presented in Table3 .

Table3 - Exogenous variables of the LSTM network

Variable	Description
PIBI(-1)	GDP index number, lagged by 1 period.
ICMS(-1)	ICMS tax revenue in São Paulo, lagged by 1 period.
MA(12)	Moving average of ICMS revenue over the last 12 months
DU	Business days in the current month
Exchange rate	Current exchange rate

Source: Prepared by the author

Consequently, the predictive performance was not very accurate across the 12 windows (Table8), with an average EAPA of 4.77% and an average MAPE of 5.49%, as shown in Table10 (Appendix), which also contains the hyperparameter values.

Under the *bottom-up* approach, the exogenous variables listed in Table4 were considered for each tax collection regime.

Table4 - Exogenous variables for each tax collection regime

System	Exogenous Variables
ICMS_RPA	PIBI(-1), ICMS_RPA(-1)
ICMS_Importação	PIBI, Exchange Rate, DU, ICMS_Importação(-1)
ICMS_ST	PIBI, ICMS_ST(-1)

Source: Author's own analysis

After running the 12 forecast windows, the predictive performance of the LSTM network using the *bottom-up* approach demonstrated improved accuracy compared to the *top-down* approach, with an average EAPA of 1.69% and an average MAPE of 3.12% (Table7). Also noteworthy is the low standard deviation shown by the EAPA (0.89%) and MAPE (0.46%) metrics, which demonstrates the stability of the predictive performance. The hyperparameter values are listed in Tables11 ,12 , and13 on the web (section7).

4.3. Foundation Model

Massive artificial intelligence systems, trained on vast amounts of data—such as large language models (LLMs) for natural language processing—are called *Foundation Models*. Recently, this same framework has been applied to large time-series datasets, with the aim of generating predictions using no input other than the time series itself (*zero-shot*).

The TimesFM model (Das *et al*., 2024) is pre-trained on a time-series corpus consisting of 300 billion data points, designed to make predictions without requiring any input data preparation, hyperparameter tuning, or even training. The model simply takes a time series as input, which provides it with context, and returns a forecast for a defined time horizon. It has 500 million parameters and supports an input of up to 2,048 data points. Thus, on the one hand, processing time is drastically reduced, as this model requires no hyperparameter tuning or training. On the other hand, this model requires more memory—about 4 GB—which is consistent with the space-time dilemma in computing.

Given that TimesFM requires minimal analytical effort and little processing, without any training on the input time series, the performance of its *top-down* forecast was surprising, with an average EAPA of 2.66%—the best annual performance among univariate models—and an average MAPE of 3.72% (Table8).

In contrast, under the *bottom-up* approach, the predictive performance of the TimesFM model deteriorated significantly, with an average EAPA of 4.25% and an average MAPE of 4.58% (Table8).

5. DISCUSSION

The comparison of the predictive performance of the 14 forecasting methodologies, whose results were presented in the section4 , was conducted in two ways. The first used the *Model Confidence Set* (Hansen *et al*., 2011), in conjunction with the Friedman test (1937), to rank the performance of the forecasts based on the relative position of their errors compared to the errors of the other forecasts, which was referred to as an ordinal performance *ranking*. The second method, on the other hand, used the average of the accuracy metrics themselves, presented in the section3.4 , to rank the predictive performance of the methodologies; this was termed “cardinal performance *ranking*.”

5.1. Ordinal Performance Ranking

Considering the 12 forecast windows, for each of the 12 forecast steps, the monthly performance of the 14 methodologies listed in Section 4 was ranked at each step of forecast. Furthermore, the sum of the 12 steps in each forecast window was considered to rank the annual forecast performance of the same methodologies.

Regarding the annual forecast, the *ranking* shown in Table 5 was obtained, which ordered the predictive performance according to the average of the ranks obtained across the 12 windows. Furthermore, the p-value of the EPA (*Equal Predictive Ability*) test from the *Model Confidence Set* (Hansen *et al*., 2011) is presented. Thus, at a significance level of 25%, models with a p-value greater than this threshold would fall within the residual set.

For this set, the Friedman test (1937) was performed. With a p-value of 0.0001, the null hypothesis of equivalence in predictive performance was rejected. Next, the average rank—calculated across the 12 forecast windows—was used to rank the performance of the annual forecasts. The post hoc comparison by Nemenyi (1963) was not informative, as it only allowed us to conclude that the performance of the methodologies occupying the top 10 positions in the annual *ranking* was superior to that of the bottom 4 positions.

Table 5 - Annual Ordinal Performance Ranking

Experiment	Average *Rank*	MCS p-value
BU_ECM	3,583	1
TD_SARIMAX	4,5	0,72
BU_SARIMAX	4,667	0,97
TD_SARIMAX_d	4,917	0,475
BU_LSTM	5,75	0,72
TD_SCINet	6,667	0,38
TD_TimesFM	6,667	0,38
BU_SARIMAX_d	7,25	0,32
TD_ECM	7,667	0,21
TD_SARIMA	8,75	0,04
TD_LSTM	9,75	0,035
BU_SCINet	10,25	0
BU_TimesFM	11	0
BU_SARIMA	13,583	0

Source: Author's own calculations

In this instance, it was found that the *Error Correction Model* forecast, using the *Bottom-Up* approach (BU_ECM), achieved the best annual forecasting performance, *ranking* first, followed by SARIMAX, using both the *Top-Down* and *Bottom-Up* approaches (TD_SARIMAX and BU_SARIMAX), in second and third place, respectively.

In addition, the performance of the monthly forecast was evaluated based on the relative *ranking* of each of the 14 methodologies in each of the 12 steps across the 12 forecast windows. Thus, the average rank of the " was obtained (6). The same table shows the number of forecast horizons for which each model was part of the MCS residual group (Hansen *et al*., 2011).

Consequently, the *Bottom-Up* SARIMAX (BU_SARIMAX) emerged as having the best monthly predictive performance, followed by TD_SARIMAX and the *Bottom-Up* LSTM network (BU_LSTM) in second and third place, respectively.

Table 6 —Ranking by monthly ordinal performance

Methodology	Average *Rank*	No. of Horizons in the MCS (α=0,25)
BU_SARIMAX	4,62	12
TD_SARIMAX	5,35	12
BU_LSTM	5,81	12
BU_ECM	6,55	12
TD_SARIMAX_d	6,67	12
BU_SARIMAX_d	6,8	12
TD_TimesFM	6,85	11
TD_SARIMA	7,03	12
TD_ECM	7,19	12
TD_SCINet	7,9	8
BU_TimesFM	8,76	5
TD_LSTM	9,23	4
BU_SCINet	9,99	3
BU_SARIMA	12,25	1

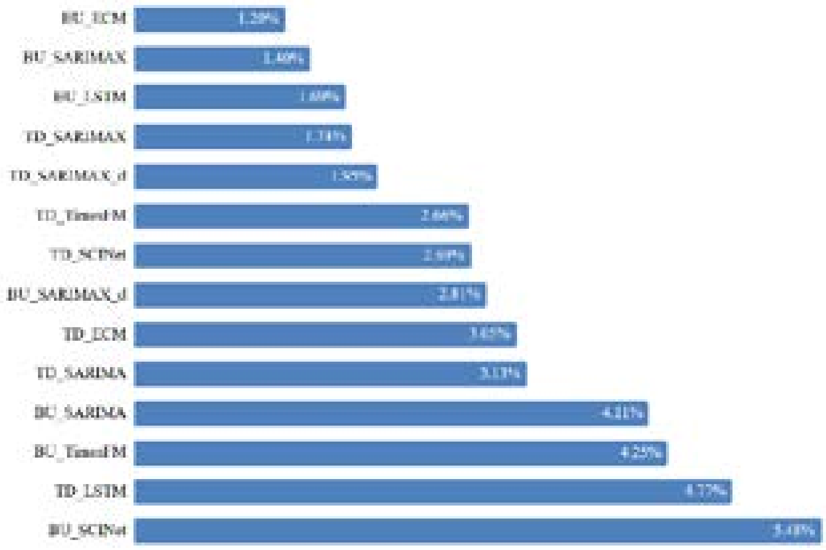
Source: Compiled by the author

5.2. Ranking by Cardinal Performance

In addition, the annual (EAPA) and monthly (MAPE) predictive accuracy metrics, presented in the section3.4 , were used to compare the predictive performance of the 14 methodologies.

Once again, as shown in Figure3 , BU_ECM ranked first in the annual forecast *ranking*, with an average EAPA of 1.2%, followed by BU_SARIMAX (1.4%) and BU_LSTM (1.69%). The ECM explicitly models the long-term relationship of cointegrated time series. Thus, its annual predictive performance, compared to the others, confirms the econometric intuition.

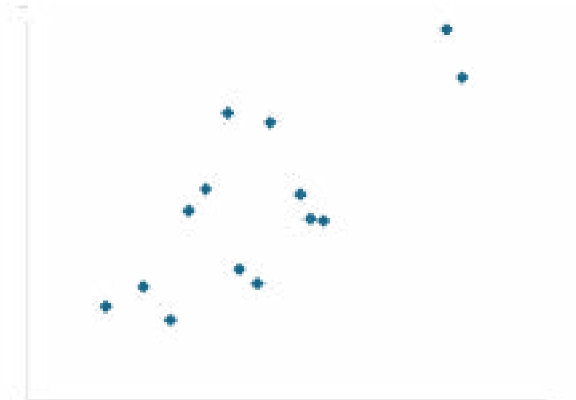
Figure 3 - Annual *ranking*, based on the average EAPA.



Source: Author’s own work

However, when we also look at the standard deviation of the annual EAPA, we find that the performance of BU_SARIMAX was more stable, with a standard deviation of 0.6%, compared to BU_LSTM and BU_ECM, whose EAPAs had standard deviations of 0.89% and 1.1%, respectively (Figure4).

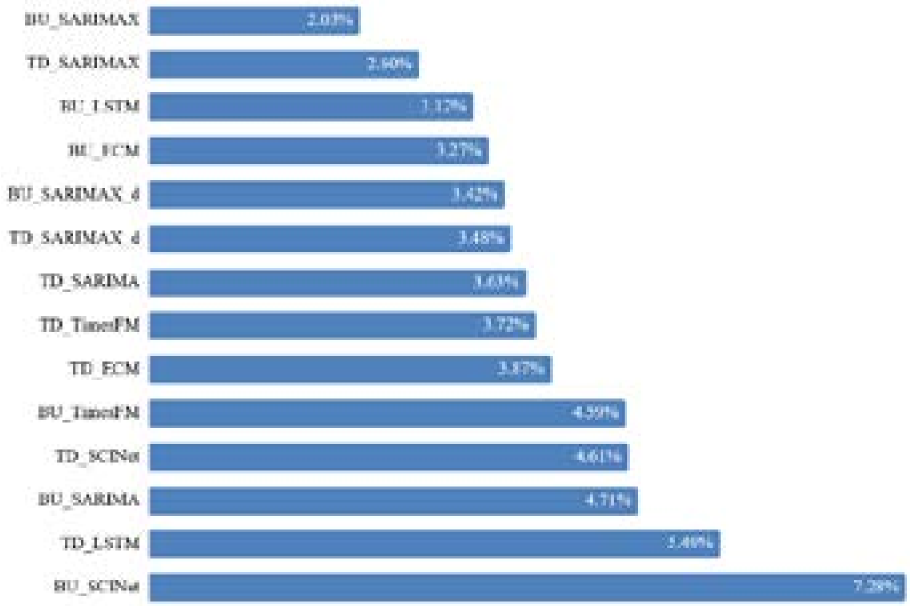
Figure 4 - Annual predictive performance: Dispersion of means and standard deviations of the EAPA.



Source: Author’s own work

Consistent with the ordinal performance *ranking* in the section5.1 , the BU_SARIMAX forecast showed the best monthly performance, as measured by MAPE, with an average of 2.03%, followed by the TD_SARIMAX and BU_LSTM forecasts, whose average MAPE values were 2.6% and 3.12%, respectively (Figure5).

Figure 5 - Monthly *ranking*, based on the average MAPE.



Source: Author’s own work

In terms of monthly forecast stability, the BU_SARIMAX model had a MAPE with a standard deviation of 0.46%. Thus, it was very close to the BU_LSTM (0.46%) and the BU_ECM (0.40%), as shown in Figure6 .

Figure6 - Monthly predictive performance: Scatter plot of MAPE means and standard deviations.

Source: Author’s own work

5.3. Conclusions

Based on the *ranking*s by ordinal performance (section5.1) and cardinal performance (section5.2), both in terms of monthly and annual forecasting, it was observed that the models incorporating exogenous variables (ECM, SARIMAX, SARIMAX_d, and LSTM) exhibited more accurate predictive performance, occupying the top positions. In contrast, the univariate models (SARIMA, SCINet, and TimesFM) ranked lower. It should be noted that the forecasts were made using only the information available at the time of the models' last estimation or training period, to simulate a real-world forecasting scenario, as described in the section3.5 . Thus, the expected values of the exogenous variables were used, as recorded in the expectations collected by the BCB (FOCUS Report), which ensures a level playing field between the univariate models and those that incorporated exogenous variables.

Furthermore, it was found—in general—that the *Bottom-Up* (BU) approach improved the predictive performance of models incorporating exogenous variables compared to the *Top-Down* (TD) approach. For univariate models, however, the BU strategy led to a decline in predictive accuracy when compared to the TD approach.

In fact, with regard to annual forecasting, the ECM and SARIMAX econometric models stood out, *ranking* among the top positions in both the ordinal and cardinal performance *ranking*s, particularly under the BU approach. Among the *machine learning* models, the LSTM network under the BU approach exhibited the highest predictive accuracy. Although it did not rank among the top models, TimesFM under the TD approach stood out for its median performance, even while requiring negligible analytical effort and processing costs.

As for monthly forecasting, the methodologies that consistently dominated the top positions in both the ordinal and cardinal performance *ranking*s were BU_SARIMAX, TD_SARIMAX, BU_LSTM, and BU_ECM, in that order.

Finally, the methodologies that showed the most promising overall performance consisted of the SARIMAX and ECM econometric models and the LSTM recurrent neural network ( ), particularly when using the BU approach. The decomposition of ICMS revenue into ICMS_RPA, ICMS_Importacao, and ICMS_ST—each regime with its own specific model, exogenous variables, and lags—appears to have improved the accuracy of the forecasts, both on an annual and monthly basis. Therefore, within the scope of the fiscal responsibilities discussed in the introduction (section1), these methodologies appear to be the most suitable for guiding both annual budget planning and the scheduling of disbursements throughout the fiscal year.

Furthermore, in light of the tax reform enacted by Constitutional Amendment No. 132/2023—under which the tax bases for the ICMS and the Tax on Services of Any Nature (ISS) will be consolidated and expanded within the scope of the Tax on Goods and Services (IBS), in a gradual transition through 2033— the *bottom-up* approach tends to emerge as a natural alternative for forecasting future IBS revenue. This is because, given the lack of historical data for the IBS, segmented forecasts for the ICMS and ISS—which are subsequently consolidated—may serve as an initial estimate of revenue from the new tax.

BIBLIOGRAPHICAL REFERENCES

BONATO, A. S. F. **Utilização de Modelos Não Lineares para Previsão do ICMS Paulista**. 2023. 170 p. Dissertação (Dissertação (Mestrado em Economia)) — Fundação Getúlio Vargas/ Escola de Economia de São Paulo, São Paulo, 2023. Disponível em: https://repositorio.fgv.br/items/15a860b4-0458-4c68-b5aa-5f9b0789a5e4. Acesso em: 24/06/2026.

BOX, George E. P.; JENKINS, Gwilym M. **Time Series Analysis**: Forecasting and Control. San Francisco: Holden-Day, 1970.

BROCK, W.; DECHERT, W.; SCHEINKMAN, J. **A test for independence based on the correlation dimension**. SSRN Working Paper, 1987. Disponível em: https://doi.org/10.1080/07474939608800353. Acesso em: 24/06/2026.

DAS, A. *et al*. **A decoder-only** *foundation model* **for time-series forecasting**. In: PROCEEDINGS OF THE 41st INTERNATIONAL CONFERENCE ON MACHINE LEARNING. [S. l.]: PMLR, 2024. p. 10148-10167. Disponível em: https://proceedings.mlr.press/v235/das24c.html. Acesso em: 24 jun. 2026.

DICKEY, David A.; FULLER, Wayne A. **Distribution of the estimators for autoregressive time series with a unit root**. Journal of the American Statistical Association, [s. l.], v. 74, n. 366, p. 427-431, 1979. DOI: https://doi.org/10.1080/01621459.1979.10482531. Acesso em: 24/06/2026.

DORNELAS, A. F.; CAMPOS, L. D.; FIGUEIREDO, K. **Modelos para previsão tributária utilizando redes neurais lstm**. In: Anais do XIX Encontro Nacional de Inteligência Artificial e Computacional. SBC, 2022. p. 95–105. Disponível em: https://doi.org/10.5753/eniac.2022.227595. Acesso em: 24/06/2026.

DORNELLES, G. Z.; SCHWARTZER, F. R.; BRAATZ, J. **Redes Neurais Aplicadas na Previsão de Receita de ICMS no Rio Grande do Sul**. Textos para Discussão TE/RS Nº 19. Rio Grande do Sul, Brasil, 2022. Disponível em: https://tesouro.fazenda.rs.gov.br/upload/arquivos/202508/29094521-1643376127-artigo-modelo-redes-neurais-texto-de-discussao.pdf. Acesso em: 30/06/2026.

ENGLE, Robert F. **Autoregressive conditional heteroscedasticity with estimates of the variance of United Kingdom inflation**. Econometrica, v. 50, n. 4, p. 987–1007, 1982. DOI: 10.2307/1912773. Disponível em: https://doi.org/10.2307/1912773. Acesso em: 24/06/2026.

ENGLE, Robert F.; GRANGER, Clive W. J. **Co-integration and Error Correction: Representation, Estimation, and Testing**. Econometrica, [s. l.], v. 55, n. 2, p. 251-276, mar. 1987. Disponível em: https://doi.org/10.2307/1913236. Acesso em: 28 maio 2026.

ENGLE, R. F.; HENDRY, D. F.; RICHARD, J.-F. **Exogeneity**. Econometrica, [S. l.], v. 51, n. 2, p. 277–304, mar. 1983. DOI: 10.2307/1911990. Disponível em: https://doi.org/10.2307/1911990. Acesso em: 10 jun. 2026.

FONTENELE, S. V. P. **Arrecadação Tributária: Modelos de Previsão para o Estado do Ceará**. 2017. Dissertação (Dissertação (Mestrado Profissional em Economia do Setor Público – MESP)) — Programa de Pós-Graduação em Economia (CAEN), Universidade Federal do Ceará, Fortaleza, 2017.

FRIEDMAN, M. **The use of ranks to avoid the assumption of normality implicit in the analysis of variance**. Journal of the American Statistical Association, v. 32, n. 200, p. 675–701, 1937. DOI: https://doi.org/10.1080/01621459.1937.10503522. Acesso em: 24/06/2026.

HAMILTON, James D. **Time Series Analysis**. Princeton: Princeton University Press, 1994.

HAN, Z. *et al*. **A review of** ***deep learning*** **models for time series prediction**. IEEE Sensors Journal, v. 21, n. 6, p. 7833–7848, 2021. Disponível em: https://doi.org/10.1109/JSEN.2019.2923982. Acesso em: 24/06/2026.

HANSEN, P. R.; LUNDE, A.; NASON, J. M. **The model confidence set**. Econometrica, v. 79, n. 2, p. 453–497, 2011. DOI: https://doi.org/10.3982/ECTA5771.

HOCHREITER, S.; SCHMIDHUBER, J. **Long short-term memory**. Neural Computation, v.

9, n. 8, p. 1735–1780, 1997.

HYNDMAN, R. J.; KOEHLER, A. B. **Another look at measures of forecast accuracy**. International Journal of Forecasting, v. 22, n. 4, p. 679–688, 2006. DOI: https://doi.org/10.1016/j.ijforecast.2006.03.001. Acesso em: 24/06/2026.

HYNDMAN, R. J.; ATHANASOPOULOS, G. **Forecasting: principles and practice**. 3rd. ed. Melbourne, Australia: OTexts, 2021. Disponível em: OTexts.com/fpp3. Acesso em: 23 out. 2025.

JOHANSEN, Søren. **Statistical analysis of cointegration vectors. Journal of Economic Dynamics and Control**, [s. l.], v. 12, n. 2-3, p. 231-254, 1988. DOI: https://doi.org/10.1016/0165-1889(88)90041-3. Acesso em: 24/06/2026.

JOHANSEN, Søren. **Estimation and hypothesis testing of cointegration vectors in Gaussian vector autoregressive models**. Econometrica, [s. l.], v. 59, n. 6, p. 1551-1580, 1991. DOI: https://doi.org/10.2307/2938278. Acesso em: 24/06/2026.

JOHANSEN, S. **Likelihood-based inference in cointegrated vector autoregressive models**. Oxford: Oxford University Press, 1995.

JONES, D. R.; PERTTUNEN, C. D.; STUCKMAN, B. E. **Efficient global optimization of expensive black-box functions**. Journal of Global Optimization, Kluwer Academic Publishers, v. 13, n. 4, p. 455–492, 1998.

KINGMA, D. P.; BA, J. **Adam: a method for stochastic optimization**. In: INTERNATIONAL CONFERENCE ON LEARNING REPRESENTATIONS (ICLR), 3., 2015, San Diego. Proceedings [...]. San Diego: ICLR, 2015. p. 1–15. Disponível em: https://*arxiv.org*/abs/1412.6980. Acesso em: 30 jun. 2026.

KWIATKOWSKI, Denis; PHILLIPS, Peter C. B.; SCHMIDT, Peter; SHIN, Yongcheol. **Testing the null hypothesis of stationarity against the alternative of a unit root: how sure are we that economic time series have a unit root?** Journal of Econometrics, [s. l.], v. 54, n. 1-3, p. 159-178, 1992. DOI: https://doi.org/10.1016/0304-4076(92)90104-Y. Acesso em: 24/06/2026.

LI, W.; LAW, K. L. E. *Deep learning* **models for time series forecasting: A review**. IEEE Access, Piscataway, v. 12, p. 92306–92327, 2024. Disponível em: https://doi.org/10.1109/ACCESS.2024.3422528. Acesso em: 24/06/2026.

LIU, M. *et al*. **SCINet: Time Series Modeling and Forecasting with Sample Convolution and Interaction**. 2022. ArXiv. Disponível em: http://*arxiv.org*/abs/2106.09305v3. Acesso em: 8 set. 2025.

MASINI, R. P.; MEDEIROS, M. C.; MENDES, E. F. **Machine learning advances for time series forecasting**. Journal of Economic Surveys, v. 37, n. 1, p. 76–111, fev. 2023.

NEMENYI, P. **Distribution-free multiple comparisons**. 1963. Tese (Doutorado) — Princeton University, Princeton, 1963.

PHILLIPS, Peter C. B.; PERRON, Pierre. **Testing for a unit root in time series regression**. Biometrika, [s. l.], v. 75, n. 2, p. 335-346, 1988. DOI: https://doi.org/10.1093/biomet/75.2.335. Acesso em: 24/06/2026.

PHILLIPS, P. C. B.; HANSEN, B. E. **Statistical inference in instrumental variables regression with I(1) processes**. Review of Economic Studies, v. 57, n. 1, p. 99–125, 1990.

PHILLIPS, P. C. B. **Optimal inference in cointegrated systems**. Econometrica, v. 59, n. 2, p. 283–306, 1991.

POLITIS, D. N.; ROMANO, J. P. **The stationary** *bootstrap*. Journal of the American Statistical Association, v. 89, n. 428, p. 1303–1313, 1994. DOI: https://doi.org/10.1080/01621459.1994.10476870. Acesso em: 24/06/2026.

SAID, Said E.; DICKEY, David A. **Testing for unit roots in autoregressive-moving average models of unknown order**. Biometrika, [s. l.], v. 71, n. 3, p. 599-607, 1984. DOI: https://doi.org/10.1093/biomet/71.3.599. Acesso em: 24/06/2026.

SILVA, P.; FIGUEIREDO, K. **Aprendizado profundo aplicado na previsão de receita tributária utilizando variáveis endógenas**. In: Anais do XVII Encontro Nacional de Inteligência Artificial e Computacional. Porto Alegre, RS, Brasil: SBC, 2020. p. 414–425. ISSN 2763-9061. Disponível em: https://sol.sbc.org.br/index.php/eniac/article/view/12147. Acesso em: 24/06/2026.

STOCK, J. H. **Asymptotic properties of least squares estimators of cointegrating vectors**. Econometrica, v. 55, n. 5, p. 1035–1056, 1987.

WHITE, H. **A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity**. Econometrica, v. 48, n. 4, p. 817–838, 1980. DOI: 10.2307/1912934. Disponível em: https://doi.org/10.2307/1912934. Acesso em: 24/06/2026.

APPENDIX

Table 7 - Summary of the performance of forecasting methodologies

Methodology	EAPA		MAPE	
	Average	Standard Deviation	Mean	Standard Deviation
TD_SARIMA	3.13%	1.37%	3.63%	0.98%
BU_SARIMA	4.11%	1.87%	4.71%	1.42%
TD_SARIMAX_d	1.95%	1.63%	3.48%	0.99%
BU_SARIMAX_d	2.81%	1.24%	3.42%	0.96%
TD_SARIMAX	1.74%	1.77%	2.60%	1.20%
BU_SARIMAX	1.40%	0.60%	2.03%	0.46%
TD_ECM	3.05%	2.10%	3.87%	1.49%
BU_ECM	1.20%	1.10%	3.27%	0.40%
TD_SCINet	2.69%	2.18%	4.61%	1.45%
BU_SCINet	5.48%	3.23%	7.28%	2.35%
TD_LSTM	4.77%	3.35%	5.49%	2.57%
BU_LSTM	1.69%	0.89%	3.12%	0.46%
TD_TimesFM	2.66%	2.28%	3.72%	1.66%
BU_TimesFM	4.25%	1.54%	4.58%	1.33%

Source: Prepared by the author

Table 8 - Performance of methodologies by forecast window

	TD_SARIMA		BU_SARIMA		TD_SARIMAX_d	
Forecast Window	EAPA	MAPE	EAPA	MAPE	EAPA	MAPE
02/01/2024	2.36%	3.11%	1.04%	2.71%	5.07%	5.42%
03/01/2024	2.40%	3.21%	2.64%	3.76%	3.51%	4.57%
04/01/2024	5.00%	5.34%	2.66%	4.54%	4.56%	5.06%
05/01/2024	0.16%	1.90%	6.06%	6.21%	2.67%	3.69%
06/01/2024	3.58%	3.69%	7.82%	7.87%	0.58%	2.75%
07/01/2024	2.36%	2.98%	3.56%	3.95%	0.89%	2.73%
08/01/2024	4.52%	4.57%	4.54%	4.72%	0.61%	2.51%
09/01/2024	3.73%	3.99%	4.62%	4.96%	0.42%	2.75%
10/01/2024	3.59%	4.00%	4.76%	5.17%	0.55%	2.79%
11/01/2024	4.85%	4.89%	5.41%	5.47%	1.82%	3.32%
12/01/2024	2.86%	3.20%	2.00%	2.90%	1.34%	3.20%
01/01/2025	2.19%	2.65%	4.18%	4.24%	1.32%	2.98%

	BU_SARIMAX_d		TD_SARIMAX		BU_SARIMAX	
Window	EAPA	MAPE	EAPA	MAPE	EAPA	MAPE
02/01/2024	2.00%	2.25%	4.77%	4.78%	0.38%	1.59%
03/01/2024	2.54%	2.77%	3.68%	3.89%	0.25%	1.60%
04/01/2024	3.29%	3.46%	4.74%	4.72%	1.51%	2.01%
05/01/2024	3.57%	3.72%	2.62%	2.72%	1.70%	2.02%
06/01/2024	4.38%	4.43%	1.17%	1.85%	2.32%	2.38%
07/01/2024	1.51%	2.14%	1.38%	1.88%	0.89%	1.24%
08/01/2024	1.41%	2.48%	0.18%	1.37%	1.46%	1.69%
09/01/2024	1.80%	3.38%	0.49%	1.78%	1.75%	2.14%
10/01/2024	1.87%	3.60%	0.13%	1.67%	1.78%	2.20%
11/01/2024	5.31%	5.45%	1.05%	1.98%	1.56%	2.13%
12/01/2024	2.31%	3.23%	0.63%	1.95%	1.65%	2.34%
01/01/2025	3.73%	4.18%	0.00%	2.64%	1.60%	3.00%

	TD_ECM		BU_ECM		TD_SCINet	
Window	EAPA	MAPE	EAPA	MAPE	EAPA	MAPE
02/01/2024	3.60%	4.77%	0.04%	3.70%	2.72%	5.77%
03/01/2024	5.78%	6.21%	1.89%	3.26%	0.13%	3.94%
04/01/2024	6.95%	6.88%	3.35%	3.67%	5.69%	6.34%
05/01/2024	4.59%	4.68%	2.01%	3.06%	1.99%	3.62%
06/01/2024	2.96%	3.51%	1.69%	3.37%	0.93%	4.71%
07/01/2024	4.52%	4.47%	1.14%	2.97%	3.19%	3.84%
08/01/2024	1.88%	2.39%	0.04%	2.60%	1.42%	4.16%
09/01/2024	2.53%	2.70%	0.53%	2.98%	5.25%	5.38%
10/01/2024	1.72%	2.38%	0.05%	2.80%	2.17%	3.02%
11/01/2024	0.17%	2.11%	2.59%	3.62%	6.86%	7.69%
12/01/2024	1.38%	2.77%	0.39%	3.30%	0.26%	4.07%
01/01/2025	0.47%	3.56%	0.73%	3.89%	1.61%	2.74%

	BU_SCINet		TD_LSTM		BU_LSTM	
Window	EAPA	MAPE	EAPA	MAPE	EAPA	MAPE
02/01/2024	1.71%	3.32%	8.56%	8.39%	1.60%	3.04%
03/01/2024	4.91%	6.18%	9.95%	9.87%	0.90%	2.31%
04/01/2024	5.77%	6.29%	8.87%	8.80%	0.91%	3.23%
05/01/2024	10.32%	10.59%	8.06%	7.99%	0.76%	2.44%
06/01/2024	6.82%	7.03%	4.52%	5.08%	2.90%	3.53%
07/01/2024	4.60%	6.03%	4.96%	5.29%	2.46%	3.40%
08/01/2024	3.86%	6.55%	3.85%	4.36%	1.53%	2.80%
09/01/2024	2.00%	4.81%	2.70%	3.26%	0.07%	2.70%
10/01/2024	1.68%	8.07%	1.33%	3.13%	2.63%	3.41%
11/01/2024	3.90%	7.45%	0.31%	2.81%	2.58%	3.76%
12/01/2024	9.60%	9.49%	0.90%	3.05%	2.17%	3.38%
01/01/2025	10.55%	11.56%	3.21%	3.86%	1.74%	3.46%

	TD_TimesFM		BU_TimesFM	
Window	EAPA	MAPE	EAPA	MAPE
02/01/2024	4.75%	5.27%	5.05%	5.57%
03/01/2024	5.25%	5.85%	5.93%	6.39%
04/01/2024	7.52%	7.49%	6.79%	6.75%
05/01/2024	3.25%	3.44%	4.53%	4.54%
06/01/2024	1.65%	2.72%	5.71%	5.64%
07/01/2024	3.46%	3.76%	5.29%	5.24%
08/01/2024	0.31%	1.82%	2.87%	3.34%
09/01/2024	0.35%	2.30%	2.38%	3.01%
10/01/2024	0.28%	2.67%	2.05%	2.79%
11/01/2024	0.78%	2.60%	3.55%	3.72%
12/01/2024	2.16%	3.34%	2.86%	3.61%
01/01/2025	2.19%	3.39%	4.02%	4.42%

Source: Prepared by the author

Table 9 - Johansen Cointegration Test

Endogenous	Exogenous	*Lags* (BIC)	*Rank* (dash)	*Rank* (λmax)
ICMS	GDP, CPI	2	1	1
ICMS	PIBI, Exchange Rate, IPCA	1	3	3
ICMS_RPA	PIBI	1	2	1
ICMS_RPA	PIBI, IPCA	1	1	1
ICMS_Import	PIBI, Exchange Rate	2	3	2
ICMS_Import	PIBI, Exchange Rate, IPCA	1	3	3
ICMS_ST	PIBI, IPCA, LS202306	2	1	2
ICMS_ST	PIBI, Dst	1	1	1

Source: Prepared by the author

Table 10 - LSTM network forecast, using a top-down approach

Window	units	dropouts	seq_length	EAPA	MAPE
02/01/2024	585	0.05	6	8.56%	8.39%
03/01/2024	556	0.45	4	9.95%	9.87%
04/01/2024	409	0.05	7	8.87%	8.80%
05/01/2024	587	0.24	6	8.06%	7.99%
06/01/2024	594	0.05	7	4.52%	5.08%
07/01/2024	434	0.57	9	4.96%	5.29%
08/01/2024	499	0.05	24	3.85%	4.36%
09/01/2024	587	0.45	6	2.70%	3.26%
10/01/2024	142	0.05	19	1.33%	3.13%
11/01/2024	506	0.05	20	0.31%	2.81%
12/01/2024	504	0.06	23	0.90%	3.05%
01/01/2025	504	0.06	23	3.21%	3.86%

Source: Prepared by the author

Table 11 - LSTM network forecast, using a bottom-up approach, ICMS_RPA regime

Window	units	dropouts	seq_length	MAPE	EAPA
02/01/2024	588	17.54%	6	3.95%	0.62%
03/01/2024	512	65.91%	10	4.43%	0.79%
04/01/2024	431	15.92%	6	4.42%	0.03%
05/01/2024	692	5.00%	17	4.18%	0.23%
06/01/2024	588	19.86%	6	4.54%	0.90%
07/01/2024	431	15.92%	6	4.82%	1.69%
08/01/2024	594	75.00%	23	3.44%	2.27%
09/01/2024	601	5.00%	16	3.46%	0.98%
10/01/2024	504	6.44%	23	4.32%	0.57%
11/01/2024	430	5.00%	9	4.38%	1.00%
12/01/2024	433	5.00%	6	4.50%	0.49%
01/01/2025	432	5.00%	10	4.82%	0.65%

Source: Prepared by the author

Table 12 - LSTM network forecast, using a bottom-up approach, ICMS_Importacao regime

Window	units	dropouts	seq_length	MAPE	EAPA
02/01/2024	607	19.29%	4	7.14%	2.59%
03/01/2024	182	5.00%	3	6.13%	3.46%
04/01/2024	569	25.15%	3	6.48%	2.70%
05/01/2024	619	5.71%	3	7.25%	2.88%
06/01/2024	592	5.00%	3	7.11%	0.68%
07/01/2024	589	63.89%	4	5.92%	0.24%
08/01/2024	504	6.44%	23	5.91%	1.52%
09/01/2024	465	47.68%	3	6.87%	1.57%
10/01/2024	428	5.00%	8	7.06%	2.44%
11/01/2024	431	15.92%	6	7.78%	6.25%
12/01/2024	419	5.00%	3	7.95%	6.41%
01/01/2025	156	27.51%	3	9.11%	6.96%

Source: Prepared by the author

Table 13 - LSTM network forecast, using a bottom-up approach, ICMS_ST regime

Window	units	dropouts	seq_length	MAPE	EAPA
02/01/2024	518	75.00%	3	8.06%	2.75%
03/01/2024	544	75.00%	3	8.12%	2.12%
04/01/2024	552	75.00%	3	8.79%	1.17%
05/01/2024	218	5.00%	3	8.98%	4.91%
06/01/2024	374	5.00%	3	11.16%	8.87%
07/01/2024	610	5.00%	3	9.42%	6.35%
08/01/2024	431	15.92%	6	12.77%	11.86%
09/01/2024	178	55.07%	3	7.51%	0.39%
10/01/2024	365	5.00%	3	9.77%	7.07%
11/01/2024	317	5.00%	3	5.43%	2.74%
12/01/2024	460	5.00%	3	5.27%	2.08%
01/01/2025	466	5.00%	3	4.25%	2.37%

Source: Prepared by the author

Table 14 - SCINet network forecast, using a top-down approach

Window	kernel_size	dropout	batch_size	learning_rate	num_levels	EAPA	MAPE
02/01/2024	10	32.30%	5	0.51%	1	2.72%	5.77%
03/01/2024	6	32.40%	7	0.92%	1	0.13%	3.94%
04/01/2024	8	61.30%	5	0.76%	2	5.69%	6.34%
05/01/2024	10	32.30%	5	0.51%	1	1.99%	3.62%
06/01/2024	11	39.50%	4	0.62%	2	0.93%	4.71%
07/01/2024	9	63.20%	5	0.75%	2	3.19%	3.84%
08/01/2024	10	72.00%	7	1.00%	3	1.42%	4.16%
09/01/2024	10	32.30%	5	0.51%	1	5.25%	5.38%
10/01/2024	12	20.00%	6	0.67%	1	2.17%	3.02%
11/01/2024	8	20.00%	5	0.77%	1	6.86%	7.69%
12/01/2024	9	72.00%	5	1.00%	3	0.26%	4.07%
01/01/2025	8	62.70%	5	0.91%	2	1.61%	2.74%

Source: Prepared by the author

Table 15 - SCINet network forecast, using a bottom-up approach, ICMS_RPA regime

Window	kernel_size	dropout	batch_size	learning_rate	num_levels	MAPE	EAPA
02/01/2024	7	54.22%	3	0.53%	2	16.31%	16.25%
03/01/2024	6	40.00%	5	0.96%	2	10.26%	9.78%
04/01/2024	5	56.04%	4	0.85%	2	6.06%	4.85%
05/01/2024	4	60.00%	5	1.00%	2	5.78%	3.94%
06/01/2024	7	60.00%	5	1.00%	3	7.63%	5.03%
07/01/2024	4	60.00%	5	1.00%	3	4.10%	1.44%
08/01/2024	6	41.27%	4	0.54%	2	3.99%	1.68%
09/01/2024	5	43.47%	4	1.00%	2	5.32%	0.12%
10/01/2024	6	44.72%	4	0.60%	2	5.79%	2.23%
11/01/2024	6	40.00%	3	0.19%	2	7.07%	5.73%
12/01/2024	7	40.00%	4	1.00%	2	7.73%	6.61%
01/01/2025	5	54.64%	4	0.32%	2	4.92%	1.77%

Source: Prepared by the author

Table 16 - SCINet network forecast, using a bottom-up approach, ICMS_Importacao regime

Window	kernel_size	dropout	batch_size	learning_rate	num_levels	MAPE	EAPA
02/01/2024	6	44.58%	4	0.56%	2	9.33%	6.87%
03/01/2024	7	40.80%	3	0.80%	2	31.45%	32.47%
04/01/2024	7	40.00%	3	1.00%	3	30.31%	30.89%
05/01/2024	6	41.27%	4	0.54%	2	23.55%	24.15%
06/01/2024	7	40.80%	3	0.80%	2	13.73%	14.12%
07/01/2024	7	40.00%	3	1.00%	3	11.11%	10.27%
08/01/2024	7	40.80%	3	0.80%	2	14.44%	14.71%
09/01/2024	5	42.45%	3	0.32%	2	13.31%	13.82%
10/01/2024	7	40.00%	3	1.00%	3	8.04%	3.38%
11/01/2024	6	41.27%	4	0.54%	2	9.17%	6.29%
12/01/2024	6	56.91%	5	0.66%	3	7.93%	3.78%
01/01/2025	7	40.00%	3	1.00%	3	8.70%	5.99%

Source: Prepared by the author

Table 17 - SCINet network forecast, using a bottom-up approach, ICMS_ST regime

Window	kernel_size	dropout	batch_size	learning_rate	num_levels	MAPE	EAPA
02/01/2024	6	44.72%	4	0.60%	2	19.64%	20.27%
03/01/2024	6	44.72%	4	0.60%	2	14.10%	11.08%
04/01/2024	6	56.91%	5	0.66%	3	23.74%	5.81%
05/01/2024	6	60.00%	5	1.00%	3	27.82%	27.10%
06/01/2024	6	58.52%	5	0.62%	2	25.67%	24.59%
07/01/2024	7	40.00%	3	1.00%	3	10.64%	6.10%
08/01/2024	6	59.11%	4	0.67%	3	23.66%	5.79%
09/01/2024	7	43.56%	5	0.94%	2	13.17%	4.25%
10/01/2024	7	40.00%	3	1.00%	2	27.55%	0.94%
11/01/2024	6	49.74%	3	0.90%	3	33.11%	31.79%
12/01/2024	6	44.72%	4	0.60%	2	29.21%	26.67%
01/01/2025	6	59.11%	4	0.67%	3	40.95%	42.04%

Source: Prepared by the author