Journal of Business Management and Economics
Vol.1(1), pp.001-009, November 2022
DOI: 10.5281/zenodo.7749556

http://www.sparklyresearchjournal.org
SPARKLY RESEARCH JOURNAL

Original Research Article

Statistical Model for Analyzing and Predicting Burundian Tax Revenues: Case Study of Burundi Revenue Authority

Received 5 October, 2022; Accepted 10 November, 2022

Byamungu Wanguwabo David*[1], Boyayo Théodette[2], Bourgeois Marc[3], Nahayo Fulgence[4], Traversa Edoardo[5], Louis Aimé Fono[6], Masabo Michel[7] and Manene M. Moses[8].

[1]Laboratoire Universitaire de Recherche en Modélisation et en Ingénierie de la Statistique appliquée "LURMISTA"-ISTA/University of Burundi in Doctoral School of University of Burundi, Lecturer at Université du Cinquantenaire and Université Libre des Pays des Grands-Lacs, RD Congo
[2]Doctoral School of University of Burundi, Burundi
[3] Faculty of Law at the University of Liège
[4]Ecole nationale d'Administration et Laboratoire Universitaire de Recherche en Modélisation et en Ingénierie de la Statistique appliquée "LURMISTA"-ISTA/University of Burundi in Doctoral School of University of Burundi, Burundi
[5]Faculty of Law and criminology, Université Catholique de Louvain-La-Neuve
[6]Laboratoire de mathématiques-Equipe de Recherche en Mathématiques Appliquées aux Sciences Sociales-EMASS, University of Douala, Cameroon
[7] Faculté de droit, Université du Burundi, Burundi
[8]Research Center in Applied Mathematics, University of Nairobi, Kenya
***Corresponding Author: byamungudavid@gmail.com**

To model and forecast monthly tax revenue collection, and propose policy recommendations for stabilization of revenue collection volatility are our objectives in this study. Data analysis of the evolution of tax revenues shows a slight growth and the Box.com test gives |λ|=0.07 which leads to a check logarithmic transformation. The study revealed a positive trend in data with an amplitude of variations increasing slightly over time. It showed also a quarterly seasonality, highlighting a peak in each third month of each year and a relative increase in the first two quarters. The correlogram and partial correlogram show respectively the lag of orders 3 and 12 and 3 and 9 which was significant and lead to conclude the seasonal autoregressive coefficient. Then, the Dickey-Fuller test (p-value < 0.01) and Phillips Perron test (p-value < 0.01) come up to confirm the stationarity of the series under study. ARIMA (0,0,0) (2,1,2) [4] and ARIMA (0,0,0) (2,1,1) [4] models have an AIC of 359.44 and 359.43 respectively, which is smaller than the other models and have been selected for analyzing and predicting the tax revenues of Burundi. But specifically, the study used ARIMA (0,0,0) (2,1,2) [4] which is better than ARIMA (0,0,0) (2,1,1) [4] because of its least parameters. The study thought that this model is recommendable for this institution, which supports the governmental constraints.

Keywords: ARIMA models, Burundian revenue authority, forecasting, tax revenue, time series.

INTRODUCTION

Financial resources are needed for each Government need to provide the city's public goods and services. That's why governments need to ascertain the availability of financial resources to determine the level of provision of these goods including tax and non-tax revenues. Thus, within governments, budgetary procedure frameworks, fiscal forecasting, and monitoring techniques have emerged as crucially important. Fiscal forecasts include revenue and expenditure forecast. In tax revenue, trends in government revenue collection in Burundi indicate persistent growth which can be explained by, among others, changing tax rates, increasing collection efforts widening the tax base, simplifying tax payment procedures, and promoting voluntary compliance (TRA, 2013)

Actual revenue collections have seldom met targets despite the growth and a result of higher targets driven by 'optimism' or ambition of reaching some percent of the GDP target can be the fact that explains the failure to meet these targets. It can be also explained by inefficient administration in tax collection or forecast accuracy when projected revenue collections are beyond potential realized collection.

Tax administrations, as well as the government, are stressed because of failure to meet revenue collection and failure to meet some expenditure commitments. However, revenue collection and its dynamics can be well predicted, if proper forecasting techniques are employed, and so have accounted with reasonable accuracy. The relevance of this problem lies in the fact that the redistributive function of the tax system is undermined by the lack of reliable tax revenue forecasts.

Apart from the judgemental methods in forecasting tax revenue, quantitative methods based on static and dynamic models are also used. Static models include GDP-based models. For example, GDP-based models use tax elasticity and buoyancy to assume linearity - tax revenue responds at a certain percentage with changes in GDP. They, sometimes, fail to capture dynamics in tax revenue and may produce less accurate forecasts despite several merits of dynamic models. Because they are capable of using past information to predict dynamics in revenue collection, in this study, dynamic time series models have been chosen; in particular the issue of volatility. This paper, therefore, undertakes to forecast revenue using dynamic models (ARIMA and combined forecasts) and volatility models. Further, ARIMA models are ideal for high-frequency data like those recorded monthly over a long period of exactly 48 months.

To model and forecast monthly tax revenue collection and propose policy recommendations for stabilization of revenue collection volatility based on the forecasting model are our objectives in this study. This study used monthly tax revenue data spanning June 2017 to June 2021. The study focuses on ARIMA models to forecast tax revenue and its volatility respectively. Tax revenue was found to increase steadily over the period, although with persistent volatility that increases over time.

LITERATURE REVIEW

Various tax revenue forecasting techniques divided into qualitative or quantitative are available (Jenkins, Kuo and Shukla 2000). Referred to as judgmental forecasts, qualitative methods are based on human judgment that's why they prone to bias and conservatism. However, they are used in situations where historical data are no longer representative, where data are scarce or non-existent, or such as in cases of structural change. That's why consensus forecasting is a special case of judgmental forecasting.

Then, in setting revenue targets, this method is determined by institutional setup which comprises the parties involved. As argued in (Voorhees, 2004) studies, forecasting accuracy and institutional arrangements imply consensus forecasting which diminishes forecast bias and increases forecast accuracy as it takes the politics out of the revenue forecast accuracy on one hand. On the other side, lags between forecast preparation and forecast use can be increased by the time-consuming nature of consensus forecasting, in turn potentially reducing accuracy. Tax revenue forecasting errors are not solely caused by the fluctuations in the economy, but also are attributable to the institutional structures and the degree of consensus required for the forecast. "It is good practice to try to insulate forecasting from the political process and consensus forecasting can help achieve that" claimed Borrowing from (Boyd and Dadayan, 2014).

Both causal and extrapolation are included in quantitative methods. Causal methods, however static and hence low ability to capture dynamics in data collected over a long period, involve simple and multiple regression of a dependent variable (tax revenue) and some other independent variables (e.g., income, imports, consumption). Extrapolation techniques, commonly referred to as time series techniques, make use of past data to predict the future and are accurate compared to the former making them more popular. The most widely used extrapolative techniques are: moving averages and exponential smoothing, naïve models which assume the current situation is the same as the previous; which use averages of the most recent data to calculate forecasts; trend line analysis which regresses a variable on some function of quadratic, logarithmic, time – linear, etc.; autoregressive models which, on its past values, regress a variable; and Box-Jenkins models are considered quite an accurate approach to forecasting. Box-Jenkins, by combining autoregressive, Integrating, and moving average processes (ARIMA), provide more objective forecasts because they are able of revealing regularities in the data that would be overlooked by other methods.

In developing countries like Burundi, major challenges to revenue collection are the prevalence of discretionary changes in tax systems, linked with the inadequacy of data and limited forecasting skills. For example, (Fjeldstad, Jensen and Paulo, 2014) reported the underdeveloped administrative capacities and influence of political and economic factors as the major challenges for sound fiscal policy in Angola where the application of any sophisticated forecasting methods or adherence to economic growth projections by the revenue forecasters found no evidence.

Date adequacy and skills form major impediments to the use of sophisticated time series models in revenue forecasting in developing countries as mentioned challenges, literature shows that time series models, like

ARIMA, are more extensively used in developed countries than in developing countries due to the said challenges. Since this study focuses on monthly revenue forecasts it benefits from the advantage of a large sample size.

As measured by the size of forecast error, the accuracy of forecasts is very important and represents a key element for the design and execution of sound fiscal policies. Budget management problems arise from forecasting errors and Auerbach (1995) distinguishes between three types of errors: policy errors, economic errors, and technical (behavioral) errors. A model needs to perform better both in-sample and out-of-sample while forecast errors can never be entirely avoided. However, better performance of most models is seen in-sample forecasts than out-of-sample forecasts. Then, the Customary for forecasters is to estimate several models and compare them in terms of forecast performance. To check forecast accuracy by measuring the size and distribution of the errors, several measures can be used such as Mean Absolute Error - MAE, Mean Absolute Percentage Error - MAPE, Mean Square Error - MSE, and Root Mean Square Error – RMSE (Johnston and DiNardo, 1997; Leal, Perez, Tujula and Vidal, 2007).

The puzzle of which forecasting model performs better remains unresolved. From several kinds of literature, it is not clear which method fiscal and monetary authorities, international economic organizations, financial market analysts, rating agencies, or research institutes should be adopting when preparing their forecasts. There is no single model that outperforms others universally because of the different situations between economies to which the models have to be applied.

Although the recommendation of a single forecasting method is prominent in the forecasting literature, it is sometimes possible to achieve a more accurate forecast by using a combination of several forecasts' methods.

In many studies, combined forecasts have shown accurate results, and elements such as simple average forecasts, weighted forecasts, and linear combination forecasts belong to the variants of combined forecasts.

Bunn, 1985 states that the most widely used combination method is still the simple average. However, as it does not take into account the dynamics in the forecasts, it becomes contested.

This study adopts linear combination forecasts which are ARIMA, used because of out-of-sample forecasts and it seems to be very optimistic compared to trends of in-sample forecasts.

For dynamic models, more are data better forecasts and need to be updated from time to time, and must be data intensive. Model updating is important because an unexpected event can change the whole calculation of the predicted value and therefore forecasting any series is a continuous process rather than a single calculation (Nandi, Chaudhury and Hasan, 2014).

A discretionary question remains as to how often and when these methods should be updated as the literature does not refer to this. It should also be noted that the statistics do not prove anywhere that updates lead to greater accuracy, and this assertion is consistent with the study of (Boyd and Dadayan, 2014). From this perspective, our paper does not specify such a recommendation for the models we have used.

Materials and methods

Research Design

Since 2010 the Burundi Revenue Authority has been operating under the strategic planning model in all its activities to better accomplish its missions. Currently, the Burundian Revenue Office (BRO) has a strategic plan that covers the period from 2018 to 2022. In this strategic plan, the BR0 attempts to achieve revenue mobilization through the assessment, collection, and accounting of taxes through the administration of tax laws fairly and equitably. But to optimize revenue collection the tool of forecasting is indispensable. To make forecasts, time series theory is applied when one is interested in the evolution of a phenomenon over time.

Source and method of Data collection

This study mainly employs a time series approach using Box-Jenkins models to model and forecast monthly taxes revenues. The study uses secondary data which represent monthly and annual total tax revenue collection data published by Burundian customs officers. The monthly revenue data covers the period of 48 months from June 2017 to June 2021. This sample size is adequate for the estimation of ARIMA models; according to (Garrett and Leatherman, 2000), the generally accepted threshold for ARIMA estimation is 54 data points

Method of Data analysis

Data analysis undertaken involves descriptive analysis, stationarity tests, model fitting, and forecasting. Descriptive analyses are used to explore the internal properties of data. Analytical models used include ARIMA with seasonality. This study followed the standard procedure for the estimation of ARIMA models which has five steps: stationarity test, identification of the model, Selection of the model, Validation of the model, and forecasting of the future values (Johnston and DiNardo, 1997). Identification involves checking for stationarity and determination of the order of the model. The best models are selected by using Akaike Information Criteria (AIC) and Bayes-Schwarz Information Criteria (BIC), forecast performance measures, and other statistical criteria. AIC is an asymptotically model selection criterion. AIC provides a trade-off between goodness of fit and the complexity of model specification (Akaike, 1974). For this purpose, ARIMA-type models consist of removing obvious trends and seasonality (or periodicity) from the series and modeling the remaining residual. These methods are more sophisticated and numerically more cumbersome than the previous ones, but also more efficient. These are the ones that have been used in our study.

ARIMA (Auto-Regressive-Integrated-Moving-Average) models, popularised and formalized by Box and Jenkins (1976), have as their main objective to allow a prediction of the future evolution of a phenomenon. They seek to determine each value of the series according to the values that precede it ($y_t = f(y_{t-1}, y_{t-2},...)$).

Autoregressive (AR) processes assume that each point can be predicted by the weighted sum of a set of previous points, plus a random error term tainting the previous points, plus its error. Moving average (MA) methods assume that each point is a function of the errors. The integration process (I) assumes that each point has a constant difference from the previous point.

An ARIMA model is labeled as an ARIMA (p, d, q) model, where: p is the number of autoregressive terms, d is the number of differences and q is the number of moving averages. Stationarity is the rule that the series must satisfy to be estimated by ARIMA models. This means that the mean and variance of the series are constant over time. To eliminate the trend, differentiation is applied, i.e replacing the original series with the series of adjacent differences. The differentiated series to verify stationarity is considered as an integrated version of a stationary series (hence the term Integrated). Mathematically, a first-order differentiation is given by the difference between two successive values and this difference is constant. We have $y_t - y_{t-1} = \mu + \varepsilon t$ where μ *is* the constant of the model and represents the average difference in y. Such a model is an ARIMA (0;1;0). The second-order model operates on a difference of a difference and not on a raw difference.

An undifferentiated white noise ARIMA (0,0,0) process suggests random fluctuations around a reference value. This value can be considered a stable characteristic of the system under study.

Auto-regression

Autoregressive models assume that Y_t is determined by past values.

$$y_t = \mu + \varphi 1 y t-1 + \varphi 2 y t-2 + \varphi 3 y t-3 + ... + \varphi p y t-p + \varepsilon t \text{ ...... (1)}$$

$$\Leftrightarrow y_t = \mu + \phi_1 L y_t + \phi_2 L^2 y_t + \phi_3 L^3 y_t + \cdots + \phi_p L^p y_t + \epsilon_t \text{ ...... (2)}$$

Where L is a delay operator and ϕ_1; ϕ_2; ϕ_3; are the autoregression coefficients. If differentiation was necessary, the autoregressive model ARIMA (1,1,0) is given by:

$$y_t - y_{t-1} = \mu + \varphi(y_{t-1} - y_{t-2}) + \varepsilon_t \text{ ....(3)}$$

The autoregressive model is therefore multiplied by the value (1-L). Each observation in the autoregressive model consists of a random component (ϵ_t) and a linear combination of past observations.

An autoregressive process suggests that the phenomenon under study is not determined by a reference value. It is the previous performance (or performances) that fully determine the present performance.

Moving average

Moving average models assume that the series is subject to fluctuation around the mean. The series is therefore estimated by the weighted average of a certain number of previous values, which amounts to considering that the estimate is equal to the true average to which is added a sum of the errors that have affected the previous values.

$$Y_t = \mu - \sum_{j=0}^{q} \theta_j \epsilon_{t-j} = \mu - \sum_{j=0}^{q} \theta_j L^j \epsilon_t \quad \forall j < q \,;\, \theta_j \in R, \theta_0 = 1 \; et \; \theta_q \in R^*, \text{ .... (4)}$$

avec ϵ_t i.i.d $(0, \sigma^2)$ et θ_j the moving average coefficients of the model

The model consists of a random error component and a linear combination of past random errors. The moving average process suggests that the reference value evolves from one measurement to the next. Specifically, the reference value is a function of the previous reference value and the error in the previous measurement.

Mixed model

The addition of an autoregressive (AR) component and a moving average (MA) component provides mixed ARIMA models. In the presence of a series with a seasonal component, the SARIMA-seasonal model is used. It is labeled as SARIMA (p,d,q)(P, D ,Q)m with p, d, q the same as those of ARIMA and P order of the seasonal autoregressive part, D order of the seasonal moving average, Q order of the seasonal moving average and m the period of the seasonal component. A SARIMA (p,d,q)(P, D,Q) model of period m is given by the relation:

$$\sum_j^p \phi_{jm} x_{t-jm} = \mu + \sum_{j=0}^{q} \phi_{jm} \epsilon_{t-jm} \text{ .... (6)}$$

$$\Leftrightarrow \sum_{j=0}^{p} \phi_{jm} L^{jm} x_t = \mu + \sum_{j=0}^{q} \phi_{jm} L^{jm} \epsilon_t \text{ .... (6')}$$

$$\Leftrightarrow \left(1 - \sum_{j=0}^{p} \phi_j L^j\right)\left(1 - \sum_{j=0}^{p} \phi_j L^{jm}\right) y_t = \mu + \left(1 + \sum_{j=0}^{q} \theta_j L^j\right)\left(1 + \sum_{j=0}^{p} \theta_j L^{jm}\right) \epsilon_t \text{ ... (6'')}$$

Box-Jenkins steps (Jenkins et al, 2000).

Box-Jenkins defines 5 steps to analyze a time series:

1. Transformation: this involves analyzing the series to be able to describe it to make it stationary if it is not, but also to obtain a normal distribution of the series. Here d is known;
2. Identification of the parameters p and q: Observation of the ACF and PACF graphs. The values found at this level constitute the upper limits that can be reached for the definition of the ranges of the parameter values;
3. Evaluation of the parameters: Select the candidate model(s) that minimizes one of the AIC, AICc, or BIC criteria;
4. Model diagnosis: Check the normality and non-correlation of the residual terms. If the residual terms follow a normal distribution and are uncorrelated, then the model can be used to make predictions, otherwise, the previous steps are repeated;
5. Prediction: Predict future values, then evaluate the performance of the model by calculating.

RESULTS AND DISCUSSION

Observation and Stationarity of the Series.

Before modeling a time series, it is first necessary to represent it and to test its possible particularities such as the trend and the seasonality. The analysis of the evolution of the tax revenues shows that the variability has a slight growth, but also the Box. cox test informs that $|\lambda|=0.07$ which is close to 0, therefore a logarithmic transformation (Log Receipt) is necessary

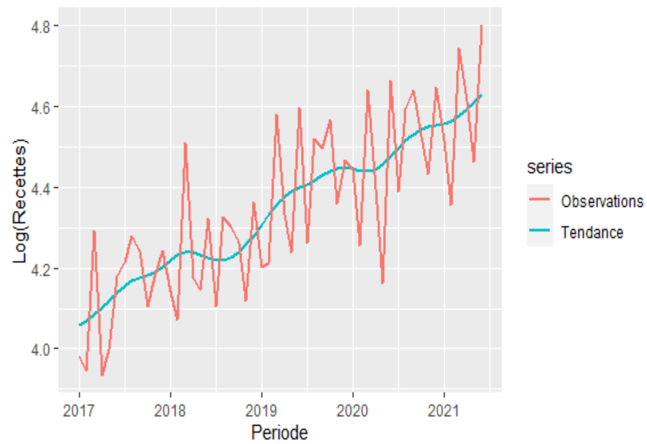

Figure 1: Tendance of Monthly Tax (fbu million)

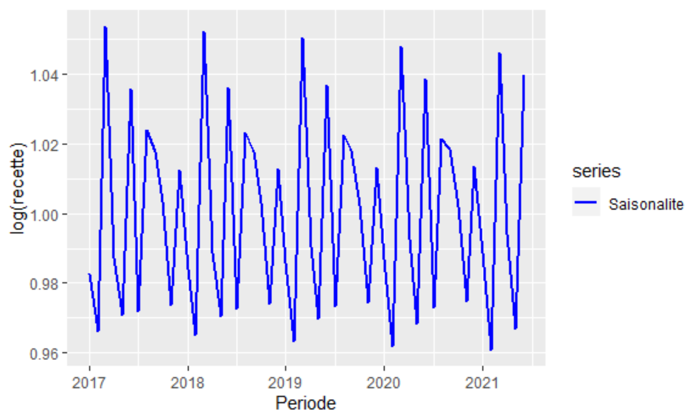

Figure 2: Seasonality of Monthly Tax
Source: Burundian Revenue office.

In the graph above of the evolution of the logarithm of revenues, the study reveals a positive trend in the data, with the amplitude of the variations increasing slightly over time. There is also a quarterly seasonality, with significant peaks indicating a large increase in revenue in each third month. Also, revenues in the first two quarters of the year show a relatively large increase compared to the last two. The study performs panel unit-root tests using the techniques by Litterman (Litterman and Supel, 1983) designated as hypothesis H_0: the time-series data are not accepted.

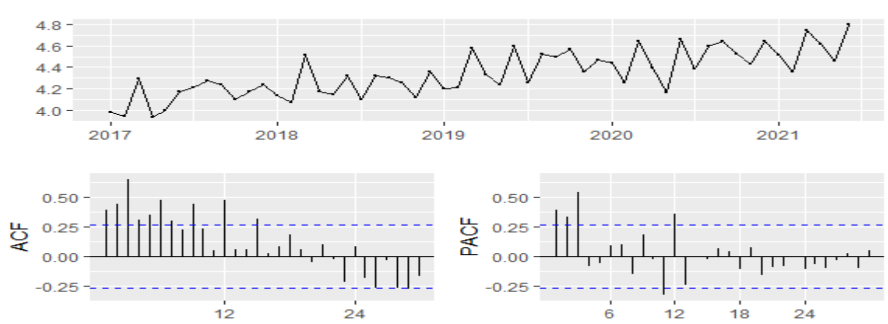

Figure 3: Display Correlogram ACF and partial correlogram PACF
Source: Researcher's computations

The correlogram obtained from the data shows that there is a quarterly seasonality: particularly significant lags are observed at orders 3, 6, 9, 12, and 15. Differentiation is therefore necessary. After the seasonal differentiation, as can be seen in the graph below, the seasonality has been eliminated and so has the trend, so regular differentiation will not be necessary. The parameters of the seasonal component are identified by analysis of the correlogram (ACF) and the partial correlogram (PACF)

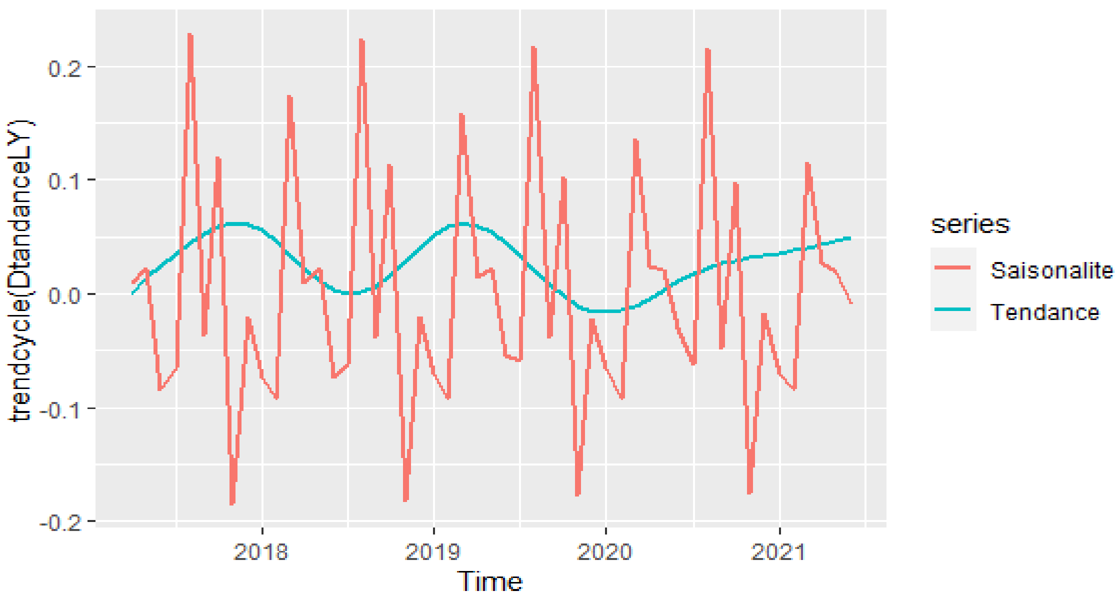

Figure 4: Display both tendance and seasonality of monthly data collection (in fbu million)
Source: Researcher's computations

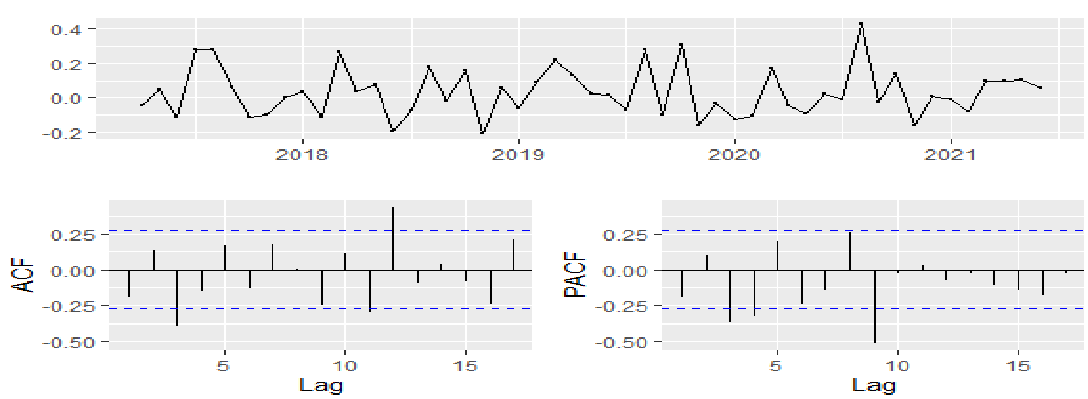

Figure 5: Display lags component on both correlogram and partial correlogram
Source: Researcher's computations

On the correlogram, the study shows that the lags of orders 3 and 12 are significant, so the coefficient of the seasonal component MA could be equal to 1 or 2. On the partial correlogram, the study reveals that the lags of orders 3 and 9 are significant, so the seasonal autoregressive coefficient could be equal to 1 or 2. After seasonal differentiation, according to the Dickey-Fuller test (p-value < 0.01) and the Phillips Perron test (p-value < 0.01), the survey concludes that the series is indeed stationary with a p-value below 1%.The competitive models are as followed:
a. ARIMA (0,0,0) (1,1,1) [4], b. ARIMA (0,0,0) (1,1,2) [4], c. ARIMA (0,0,0) (2,1,2) [4] and d. ARIMA (0,0,0) (2,1,1) [4]

Model Selection and Validation

The choice of the model is made according to the Akaike information criterion (AIC), the rule requires retaining the one that minimizes the AIC. The ARIMA (0,0,0) (2,1,2) [4] and ARIMA (0,0,0) (2,1,1) [4] models have an AIC of 359.44 and 359.43 respectively, which is smaller than the other models. Both models can model our series. We retain the ARIMA (0,0,0) (2,1,1) model [4] which has the least parameter compared to ARIMA (0,0,0) (2,1,2) model [4].

Table 1: Display ARIMA (0,0,0) (2,1,1) [4] with drift componen

Coefficients							
	Sar1	**Sar2**	**Sma1**	**Drift**			
	-0.9581	-0.8866	0.0845	0.8084			
s.e	0.0827	0.0532	0.2266	0.1032			
sigma^2 = 53.27		log-likelihood = -174.71					
AIC=359.43 AICc=360.79 BIC=368.99							
Training set errors measures							
	ME	**RMSE**	**MAE**	**MPE**	**MAPE**	**MASE**	**ACF1**
Training ste	0.1101	6.7362	4.9573	-0.6887	6.3053	0.4713	0.1224522
Z test of Coefficients							
	Estimate	**Std.Error**	**z value**	**Pr(>|z|)**			
Sar1	-0.958099	0.082704	-11.5847	<2.2e-16 ***			
Sar2	0.084534	0.226592	0.3731	0.7091			
Drift	0.808387	0.103237	7.8304	4.864e-15 ***			
Significant.Codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1							

Source: Researcher's computations

Once the model has been chosen, its validation is carried out by analyzing the estimation residuals to check whether it presents certain trends or irregularities. It is therefore analyzed to see if it meets the white noise, normality, and non-autocorrelation hypothesis.

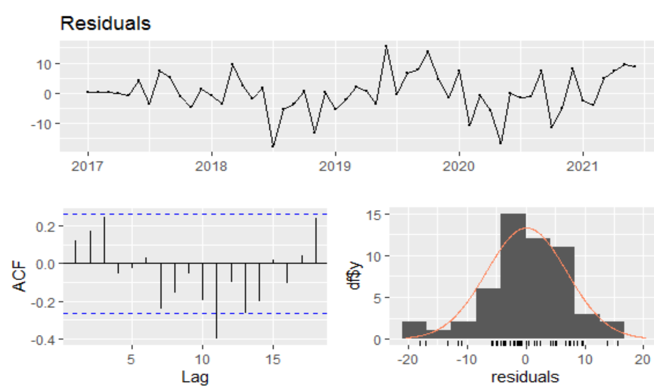

Figure 6: Display residue stability test
Source: Researcher's computations

Table 2: Model conditions for forecasting

#Ljung-Boxtest	
Data: Residuals from	ARIMA (0,0,0) (2,1,1) [4]
X-squared = 0.85554,	df = 1, p-value = 0.355
#Shapiro-Wilk normality test	
Data: Residuals from	ARIMA (0,0,0) (2,1,1) [4]
W = 0.96487, p-value = 0.1139	

Source: Researcher's computations

The residuals resulting from the model are normally distributed and are not auto-correlated as the Ljung-Box and Shapiro tests are significant. Thus, the model is good and can be used for predicting future values of total revenue

Calculation of forecasts

The revenue forecast made is from July 2021 to December 2022. The red part of the graph below shows the period over which we estimate future revenue values. The future revenue forecasts are calculated with a 95% confidence interval.

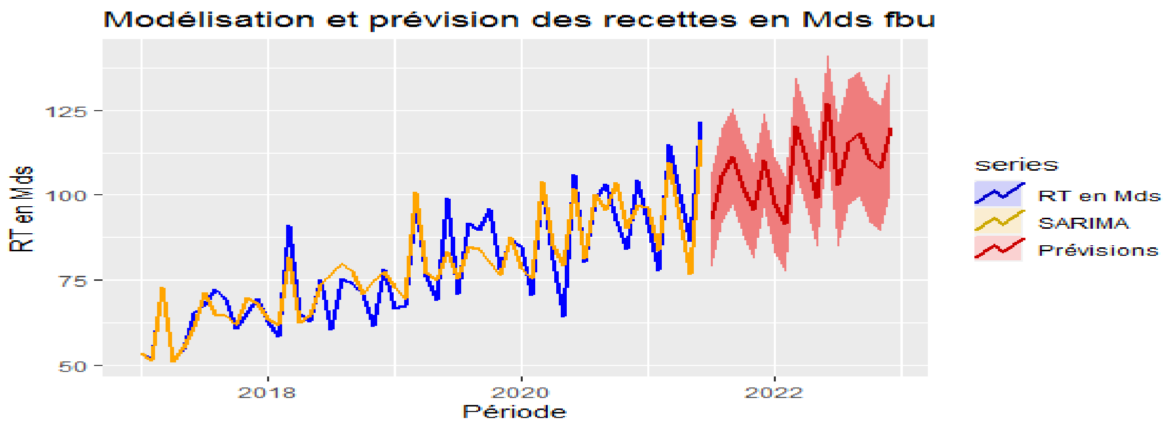

Figure 7: Display forecasting values
Source: Researcher's computation

Table 3: Forecasting tax (in million of Burundian Francs)

Month	Forecasting	Minimum Revenue	Maximum Revenue
July 2021	92.906	78.924	106.889
August 2021	105.684	91.701	119.666
September 2021	111.425	97.442	125.407
October 2021	102.238	88.256	116.221
November 2021	95.513	81.530	109.497
Décember 2021	110.337	96.353	124.320
January 2022	97.237	83.254	111.220
February 2022	91.545	77.562	105.528
March 2022	120.307	106.042	134.572
Avril 2022	110.581	96.316	124.847
May 2022	99.062	84.796	113.327
June 2022	126.928	112.663	141.194
July 2022	103.110	84.724	121.495
August 2022	115.378	96.993	133.764
September 2022	118.115	99.730	136.500
October 2022	110.645	92.260	129.030
November 2022	107.880	89.495	126.266
Décember 2022	119.471	101.086	137.857

Source: Researcher's computations

Table 3 shows the projected revenues for the second half of 2021 and the year 2022. It can be seen that revenues in April 2022 will range from 96.3 billion Burundian francs (fbu) to 124.8 billion Burundian francs. In May, revenues will decrease to 99.06 billion fbu and increase again in June to 126.9 billion fbu. This result is supported by the findings of other studies (Thompson and Gates, 2007; Boyd and Dadayan, 2014) which can be separated into two; those which suggest the substitution of more progressive taxes with a less progressive tax to reduce revenue volatility and improve revenue forecasting and those which emphasize on diversification of taxes, than the substitution of taxes, as a means to reduce revenue volatility (Crain, 2003)

From the above results, it is clear that Burundian revenues evolve according to the ARIMA model (0,0,0) (2,1,1) [4]:

$$y_t = \mu + \phi_1 y_{t-4} + \phi_2 y_{t-8} + \theta_1 \epsilon_{t-4} + \epsilon_t \text{............} (7)$$

Conclusion

This study analyzed and forecasted tax revenue for Burundi Tax Authority using dynamic models (ARIMA and combined forecasts) and volatility models. The study proceeded by testing the trend and the seasonality which constitute possible particularities. The analysis of the evolution of the receipts shows that the variability has a slight growth, but also the Box. cox test informs us that |λ|=0.07 close to 0 leads to the necessity of a logarithmic transformation (Log Receipt). A positive trend in the data, with the amplitude of the variations increasing slightly over time, is observed and a quarterly seasonality is also detected. Indeed, significant peaks are indicating a large increase in revenue in each third month. Also, revenues in the first two quarters of the year show a relatively large increase compared to the last two. From these results, we see that a seasonal differentiation is necessary. There is a quarterly seasonality, particularly at the lags, which are particularly significant at orders 3, 6, 9, 12, and 15. A differentiation is therefore necessary. After eliminating the trend and seasonality, we identified the parameters of the seasonal component by analysis of the correlogram (ACF) and the partial correlogram (PACF). Several ARIMA models such as ARIMA (0,0,0) (2,1,2) [4] and ARIMA (0,0,0) (2,1,1) [4] have a smaller AIC of 359.44 and 359.43 respectively compared to the other models. Both models were retained to model our series. We retain the ARIMA (0,0,0) (2,1,1) model [4] which has the least parameter compared to the ARIMA (0,0,0) (2,1,2) model [4] but also the residuals resulting from the model are normally distributed and are not auto-correlated as the Ljung-Box and Shapiro tests are significant. Thus, the model is good and can be used for the prediction of future values of the total revenue under study.

References

Akaike H (1974). A New Look at the Statistical Model Identification. IEEE Transactions on Automatic Control 19(6): 716-723.

Auerbach AJ (1995). Tax Projections and the Budget: Lessons from the 1980’s. American Economic Review 85(2): 165-169.

Boyd DJ, Dadayan L (2014). Revenue Forecasting: State Tax Revenue Forecasting Accuracy. Rockefeller Institute of Government Technical Report.

Bunn DW (1985). Statistical Efficiency in the Linear Combination of Forecasts. International Journal of Forecasting 1(2):151–163.

Crain WN (2003). Volatile States: Institutions, Policy, and the Performance of American State Economies. Michigan Publishing University of Michigan Press 2(1):73-81.

Fjeldstad OH, Jensen SK, Paulo FM (2014). Poor Revenue Forecasting: A Major Challenge for Sound Fiscal Policy in Angola. Angola Brief 4(1):1-4.

Garrett TA, Leatherman JC (2000). An Introduction to State and Local Public Finance. Wholbk, Regional Research Institute, West Virginia University, number 20, Winter.

Jenkins GP, Kuo CY, Shukla GP (2000). Tax Analysis and Revenue Forecasting: Issues and Techniques. Harvard Institute for International Development, Harvard University.

Johnston J, DiNardo J (1997). Econometric Methods. Cambridge University Press16(1):139–142.

Leal T, Perez JJ, Tujula M, Vidal JP (2007). Fiscal Forecasting Lessons from the Literature and Challenges. Fiscal Studies Journal of Applied Public Economics 29(3):347-386.

Litterman RB, Supel TM. (1983). Using Vector Autoregressions to Measure the Uncertainty in Minnesota’s Revenue Forecasts. Federal Reserve Bank of Minneapolis Quarterly Review (spring): 10-22.

Nandi BK, Chaudhury M, Hasan GQ (2014). Univariate Time Series Forecasting: A Study of Monthly Tax Revenue of Bangladesh. EWUCRT Working Paper No 9. East West University, Dhaka, Bangladesh.

Tanzania Revenue Authority TRA (2013). Fourth Corporate Plan (2013/2014 – 2017/2018). TRA, January.

Thompson F, Gates BL (2007). Betting on the Future with a Cloudy Crystal Ball? How Financial Theory Can Improve Revenue Forecasting and Budgets in the States.Public Administration Review 67(5): 825-836.

Voorhees WR (2004). More Is Better: Consensual Forecasting and State Revenue Forecast Error. International Journal of Public Administration 27(8-9): 651-71.