ISSN:1533-9211
SEYBOLD REPORT
www.seybold-report.com

FORECASTING THE PUBLIC REVENUE OF THE SOUTHERN REGION USING THE ARIMA APPROACH

ISSN: 1533 - 9211

Esay Solomon Shina*, Debela Geleta Dibaba, Madhu Lincoln*****

*PhD. Scholar at Andhra University and Assistant Professor in Accounting and Finance

** Assistant Professor and Ph.D. Economics Department, Hawassa University, Ethiopia

***Professor, Management Department, Dambi Dollo University, Ethiopia

Abstract

Understanding revenue forecasting practices is essential in assessing budget planning and management processes. Revenue forecasts define the budget envelope and form the basis for effective planning. Many developing countries are unable to generate a sufficient amount of public revenues. Part of this problem is the absence of preparedness in knowing and estimating how much public revenues will be collected and how it will be collected. The main objective of this study is to study the trend of public revenue collection and forecast the amount to be collected in the coming five years. The study used a time series of public revenue data from 2000 EC to 2014 E.C. Data on the total amount of revenue collected is used and computed to understand the trend of public revenue collected. Based on this collected data and using a rigorous forecasting tool, the study estimated the expected revenue to be collected for the next five years. The study also employed Eviews version 12 as a statistical package to carry out the forecasting exercise. The moving average forecasting result did not reasonably fairly capture the real trend, and the difference between the actual and predicted values is much more pronounced. The ARIMA approach has, however, better captured the trend, and the forecast ability condition of the model is satisfied.

CORRESPONDING AUTHOR:

Esay Solomon Shina

KEYWORDS:

Forecasting, Trend, Public, Revenue, ARIMA, Southern Region.

Received: 07 March 2023
Accepted: 30 April 2023
Published: 30 June 2023

TO CITE THIS ARTICLE:

Shina, E. S., Dibaba, D. G., & Lincoln, M. (2023). Forecasting The Public Revenue Of The Southern Region Using The Arima Approach. *Seybold Report Journal*, *18*(04), 205-220. https://seybold-report.com/

INTRODUCTION

Many countries still struggle to collect sufficient revenues to finance their development. Countries collecting less than 15% of GDP in taxes must increase their revenue collection in order to meet the basic needs of citizens and businesses (Danninger et al., 2004)

Collection of revenue based on generating capacity is hampered by various factors; these factories have weak administration systems, lack of compliance of both taxpayers and tax officers, lack of infrastructure, lack of smooth and regular communication between stakeholder sectors, the attitude of taxpayers and perception of tax officers, these leads to corruption, and it highly affects the revenue collection performance of the nation in general and institution in particular.

Understanding revenue forecasting practices is essential in assessing budget planning and management processes. Revenue forecasts define the budget envelope and form the basis for effective medium-term planning. They serve as the principal resource constraint and, if integrated into a top-down budget preparation process approach, facilitate the allocation of expenditures across different uses. Furthermore, transparency of forecasting processes is key in creating accountability in the revenue collection process, as manipulation of forecasts can conceal governance problems (Kaufmann, 2003).

Many developing countries cannot generate sufficient public revenues; part of this problem is the absence of preparedness in knowing and estimating how much public revenues will be collected and how it will be collected.

Objective of the study

The main objective of this study is to study the trend of public revenue collection and forecast the amount to be collected in the coming five years.

Data Type and Source

The study used a time series of public revenue data from 2000 E.C. to 2014 E.C. Data on the total amount of revenue collected is used and computed to understand the trend of public revenue collected. Based on this collected data and rigorous forecasting tools, the study will estimate the expected revenue to be collected for the next five years. The study also employed Eviews version 12 as a statistical package to carry out the forecasting exercise.

Methods and approaches

Forecasting is estimating the magnitude of future time series variables like GDP, tax revenues, government expenses, and future events and providing different results with different assumptions. Top forecasting methods include Qualitative Forecasting (Delphi Method, Market Survey, Executive Opinion, and Sales Force Composite) and Quantitative Forecasting (Time Series and Econometric Methods). The Qualitative methods are based on subjective experiences, intuitions, judgements, personal experiences, and opinions where no statistical and mathematical approaches are involved to carry out qualitative forecasting. On the other hand, Quantitative forecasting Methods employ rigorous statistical and mathematical models.

Not all methods would necessarily serve the purpose of forecasting; this study will identify the merits of the quantitative methods of forecasting and use multiple quantitative forecasting methods. The frequently used quantitative budget forecasting tools are the straight-line method, moving average method, Exponential smoothing, trend projection, Simple linear regression and multiple linear regression method.

The study will employ the moving average and ARIMA forecasting model from these.

The ARIMA Model

The study employed the ARIMA model from the set of multiple linear regression models. ARIMA models provide another approach to time series forecasting. Exponential smoothing and ARIMA models are the two most widely used approaches to time series forecasting and provide complementary approaches to the problem. While exponential smoothing models are based on describing the trend and seasonality in the data, ARIMA models aim to describe the autocorrelations in the data.

ARIMA model requires Stationarity being secured. A stationary time series is one whose properties do not depend on the time the series is observed. Thus, time series with trends or with seasonality are not stationary — the trend and seasonality will affect the value of the time series at different times. On the other hand, a white noise series is stationary — it does not matter when you observe it. It should look much the same at any time. The advantage of the ARIMA model is that it will take care of the Stationarity problem due to differencing.

The ARIMA model can be written as:

$$y'_t = c + \phi_1 y'_{t-1} + \phi_2 y'_{t-2} + \cdots + \phi_p y'_{t-p} + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \cdots + \theta_q \epsilon_{t-q} + \epsilon_t$$

Where y'_t is the differenced series. The "predictors" on the right-hand side include both lagged values of y_t and lagged errors. This is an ARIMA(p,d,q)model, where p is the order of the autoregressive part, d is the degree of differencing involved, and q is the order of the moving average part.

The study was under all the necessary processes of Identification, Estimation and Diagnosis of the ARIMA model before the forecast is made.

Result and Discussion

Result of the Moving Average Model

Moving average involves taking previous periods' average—or weighted average— to forecast the future. A moving average is a technical indicator that analysts and researcher may use to determine the direction of a trend. It sums up the data points of a series over a specific period. It divides the total by the number of data points to reach an average. It is called a "moving" average because it is continually recalculated based on the latest data of a given time series, which is extremely useful for forecasting long-term trends.

A moving average of order n can be written as:

$$\hat{T}_t = \frac{1}{n} \sum_{j=1}^{n} y_{t-j}$$

Table 1: predicted Moving average of order 3 and 4

Year	Actual revenue collected (Millions of Birr)	Predicted value MA (3)	Predicted value (MA (4)
2000	233.72	-	-
2001	327.28	-	-
2002	571.92	-	-
2003	670.61	377.64	-

2004	728.73	523.27	450.88
2005	1033.48	657.08	574.63
2006	1386.11	810.94	751.18
2007	1881.23	1049.44	954.73
2008	1990.26	1433.61	1257.39
2009	2502.05	1752.53	1572.77
2010	3172.80	2124.52	1939.91
2011	3340.65	2555.04	2386.59
2012	4365.15	3005.17	2751.44
2013	6298.37	3626.20	3345.16
2014	7002.08	4668.06	4294.24
2015	-	5888.53	5251.56
2016	-	6650.22	5888.53
2017	-	7002.08	6650.22
2018	-	7002.08	7002.08
2019	-	7002.08	7002.08
2000	-	7002.08	7002.08

The table above depicted the moving average forecasting result of revenue collection of the region for order 3 and 4. The study used Secondary time series data of revenue collection of the region for fifteen years (2000-2014 Ethiopian calendars). Based on this, the researcher predicted six years up to the 2000 Ethiopian Calendar. As can be understood from the above table, the moving average forecasting result is not that much credible. It did not reasonably fairly capture the real trend; the difference between the actual and predicted values is much more pronounced. We think we need to resort to a better method of forecasting. The discussion that follows will unveil the alternative methods of forecasting.

Result of the ARIMA Model

As was stated above, before forecasting the public revenue to be collected using the ARIMA model, we have to carry out the necessary processes of Identification, Estimation, and Diagnosis of the ARIMA model.

Identification

Identification addresses the question of Stationarity. A time series process is said to be stationary if its mean and variance are constant over time, and the value of the covariance between the two time periods depends only on the distance or gap or lag between the two time periods and not the actual time at which the covariance is computed. Whether or not a series is stationary should be statistically tested and approved. If a time series is stationary, its mean, variance, and auto covariance (at various lags) remain the same no matter at what point we measure them; that is, they are time-invariant. Such a time series will tend to return to its mean (called mean reversion), and fluctuations around this mean (measured by its variance) will have broadly constant amplitude.

The three common methods to determine the first guess at an ARIMA model should be considered: a time series plot of the data, the ACF, and the PACF.

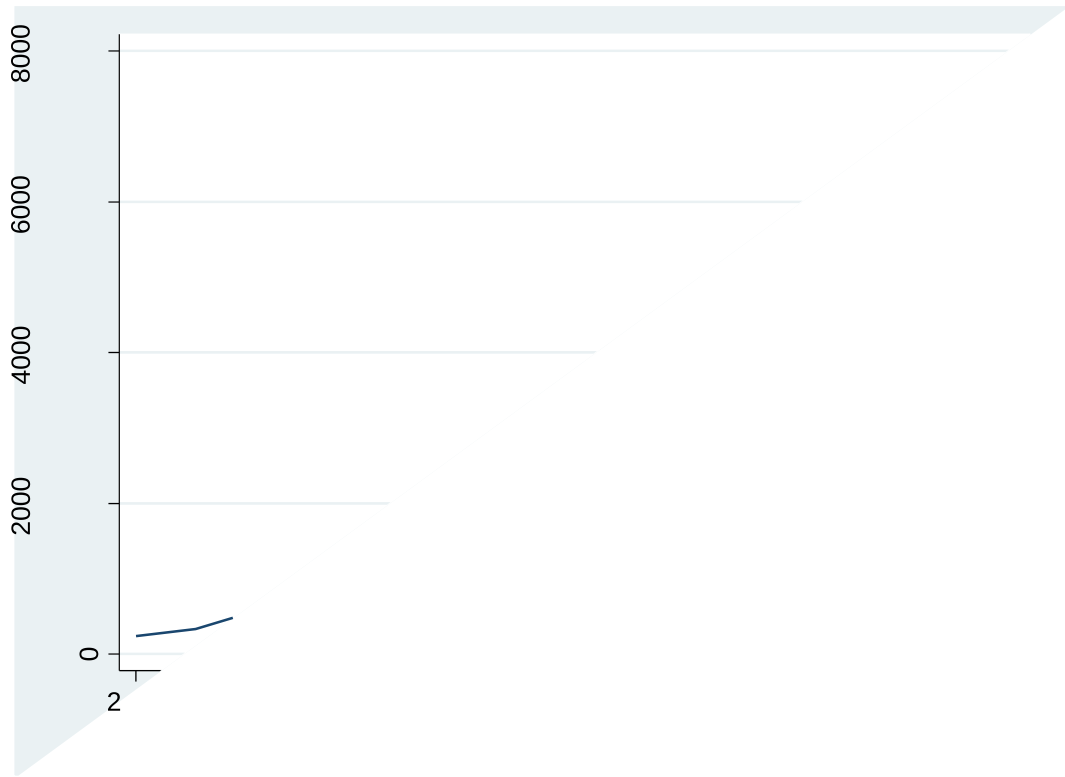
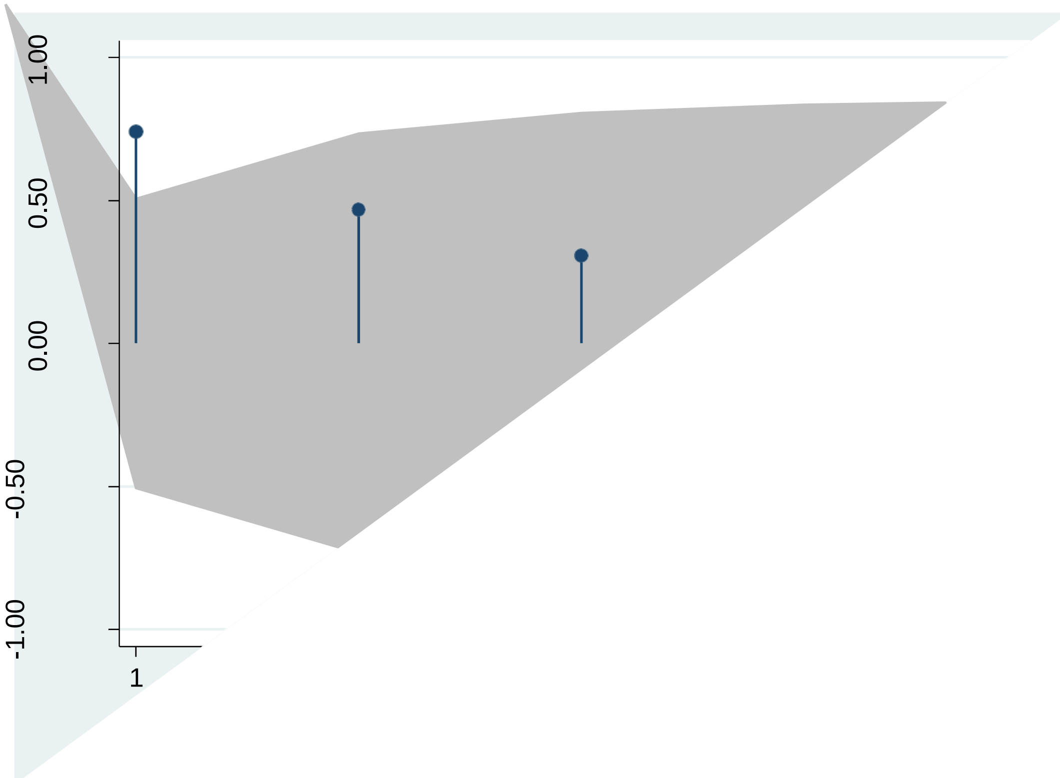


Figure 1: Annual revenue collected over the last 15 years

One of the simplest approaches of inspecting the presence of a unit root in a time series analysis is graphical depiction. Before pursuing formal tests, it is always advisable to plot the time series under such a plot because it gives us an initial clue about the likely nature of the time series. This is presented in the figure above. The above graphical depiction clearly tells us that there is a problem of Stationarity as revenue varies with the variation in time. This perhaps suggests that time series, i.e. revenue collected, is not stationary. Such an intuitive feeling is, however, only the starting point of more formal tests of Stationarity, and thus, it is not conclusive evidence.

The above result can be further evidenced by the Correlogram test. The test also validated our suspicion of a unit root telling us that there is the weak problem of Stationarity.



Formal test

Stationarity is an important property of time series data that indicates that the statistical properties of the data do not change over time. It is essential for various time series analysis techniques, including forecasting and modelling. The tests we employed to conduct a unit root analysis are the the Phillips-Person test and the augmented dickey- fuller test for a unit root.

The Stata result of Phillips-Perron and Augmented dickey-fuller test of Stationarity is depicted in the table below :

Test type	Test statistics	1% Critical value	5% Critical value	10% Critical value		Decision
	0.698	-4.380	-3.600	-3.240		
	Revenue	Coefficient	Std. Err.	t-value	P-value	

Phillips-Perron test for unit root						The null cannot be rejected
	Revenue L1	1.015149	.1727081	5.88	0.000	
	Trend	76.55564	72.54782	1.06	0.314	
	Constant	-121.5539	273.6762	-0.44	0.666	
	$H_0 : there\ is\ a\ unit\ root\ (the\ series\ is\ non-stationary)$ $MacKinnon\ approximate\ p-value\ for\ Z(t) =\ 0.9970$					
Test type	Test statistics	1% Critical value	5% Critical value	10% Critical value		Decision
Dickey-Fuller test for unit root	0.088	-4.380	-3.600	-3.240		The null cannot be rejected
	Revenue	Coefficient	Std. Err.	t-value	P-value	
	Revenue L1	.0151488	.1727081	0.09	0.932	
	Trend	76.55564	72.54782	1.06	0.314	
	Constant	-121.5539	273.6762	-0.44	0.666	
	$H_o : there\ is\ a\ unit\ root\ (the\ series\ is\ non-stationary)$ $MacKinnon\ approximate\ p-value\ for\ Z(t) =\ 0.9950$					

Table 2 : test for Stationarity

As can be understood form the above table the null hypothesis which say that the series is not stationary, cannot be rejected. Phillips-Perron and Dickey-fuller's formal test result confirmed the informal tests, and we conclude that the variable public revenue has a unit root. This means that the series has a systematic pattern that is unpredictable and thus, it has to be differenced before any forecasting is tried to be made.

Diagnosis

Having tested for the presence of a unit root, the next step that should follow is choosing the specific ARIMA model that may cater to our need of forecasting. The table below has outlined the competing ARIMA models. While diagnosing, the researchers perused the question of how well a given ARIMA model fits the data based on the Akanke information criterion (AIC), Bayesian Information Criterion(BIC), Maximum likelihood, Sigma S.Q. (estimate of the error variance) and the number of significant variables.

Decision criteria	**ARIMA(1,1,1) A**	**ARIMA(1,1,2) B**	**ARIMA(2,1,1) C**	**Decision Rule**	**Decision**
Significant variables	3/3	4/4	¾	The more the better	A AND B
AIC	1863.342	1864.059	1865.189	The lesser the better	A
BIC	1868.409	1870.815	1871.944	The lesser the better	A
Maximum likelihood	-928.671	-928.0297	-928.5943	The higher the better	B
Sigma S.Q. (estimate of the error variance)	2.89e+09	2.18e+09	2.89e+09	The smaller the better	B
CONCLUSION					A

Table 3: The ARIMA model selection criteria

Form the above table Model A or ARIMA(1,1,1) is the best model for it has all its variables significant, it has the least criterion (AIC), Bayesian Information Criterion(BIC), compared with the rest of them, concerning maximum likelihood and Sigma S.Q. (estimate of the error variance) the other competing models exceed it. But since ARIMA(1,1,1) satisfies most of the criteria, it is the model based on which we will do the forecasting.

Residuals Diagnosis

Residual diagnosis in the ARIMA model refers to Checking for the residuals of the ARIMA (1, 1, and 1) model is white noise. The residuals are the differences between the fitted model and the data. In the ARIMA model, if the residuals are white noise, if we can be sure that we have a good fit for the data on an excellent forecasting method will yield residuals with white noise. Suppose they are not white noise (i.e., they are not normal, do not have zero mean or are serially auto-correlated). In that case, your model is not fully adequate.

For us to test the white noise hypothesis, we used The Lung – Box Q statistics . The test examines the autocorrelations of the residuals. If the autocorrelations are very small, we conclude that the model does not exhibit a significant lack of fit. The null hypothesis of the Ljung-Box test is that the autocorrelations (for the chosen lags) in the population from which the sample is taken are all zero, or the residuals are independently distributed, meaning the residuals are white noise. The Ljung-Box test statistic (X-squared) gets more prominent as the sample auto-correlations of the residuals get larger (see its definition), & its p-value is the probability of getting a value as large as or larger than that observed under the null hypothesis that the true innovations are independent.

As can be seen from Appendix I, there no value that crosses the lines of autocorrelation and partial autocorrelation result besides we have a very high p-value (0.294) and thus we cannot reject the null. This implies that the residuals of ARIMA (1, 1,1) are white noise, and thus our model is a good fit.

Checking the stability Condition of our ARIMA (1, 1,1) model

After estimating the ARIMA model it is also worthwhile to check the stability condition of our model in addition to the necessity that the residuals are white noise. This is another important aspect of the model's ability to forecast. It is the restriction on the smoothing parameters. This is related to the model's stability and forecast ability conditions, defined by Hyndman et al. (2008). The stability implies that the weights for observations in a dynamic model decay over time. This guarantees that the newer observations will have higher weights than the older ones; thus, the impact of the older information on forecasts slowly disappears with the increase of the sample size.

We used a graphical approach to verify that all eigenvalues of the autoregressive polynomial lie inside the unit circle. As evidenced by Appendix II, the A.R. roots and the M.A. roots lie inside the circle. We can conclude that the ARIMA process is covariance stationary and invertible. Thus the forecast ability condition of the model is satisfied.

Forecasting

Now that all the necessary conditions are satisfied, we resort to the question of forecasting revenue collection of the southern region for the next six years until 2020 E.C. The Eviews results of the forecast are appended in Appendix III and IV.

The table below summarizes the revenue to be collected as forecasted by our ARIMA (1,1,1) model.

Year	**2015**	**2016**	**2017**	**2018**	**2019**	**2020**
Revenue to be collected (millions of Birr)	**7767.04531**	**8230.66916**	**8683.1560**	**9137.02301**	**9590.72022**	**10004.4388**

Table 4: Forecasted Revenues to be collected from 2015 to 2020 EC

From the table above it is expected that the region will collect **7,767,045,310** Birr in 2015 E.C.**,** **8,230,669,160** Birr in 2016 E.C., **8,683,156,000** Birr in 2017 E.C., **9,137,023,010** in 2018 E.C., **9,590,720,220** Birr in 2019 E.C. and **10,004,438,800** Birr in 2020 E.C.

Implication

The study, as can be understood from the results of the moving average method and the ARIMA model, has unveiled the possible expected public revenue of the region. This can be a good springboard from which many planning and preparations can be undertaken to ensure proper revenue collection and get ready for alternative action where fiscal deficit can be duly rectified.

COMPETING INTERESTS

The authors have no competing interests to declare.

Author's Affiliation

Esay Solomon Shina*, Debela Geleta Dibaba*Madhu Lincoln*

*PhD. Scholar at Andhra University and Assistant Professor in Accounting and Finance.

*PhD. and Assistant Professor in Economics.

*Professor, Management Department, Dambi Dollo University, Ethiopia

HOW TO CITE THIS ARTICLE:

Shina, E. S., Dibaba, D. G., & Lincoln, M. (2023). Forecasting The Public Revenue Of The Southern Region Using The Arima Approach. *Seybold Report Journal*, *18*(04), 205-220. https://seybold-report.com/

References

Hyndman, R.J., Koehler, A.B., Ord, J.K., Snyder, R.D., 2008. Forecasting with Exponential Smoothing. Springer Berlin Heidelberg.

Alt, Robert, 1993, "Revenue Forecasting and Estimation—How It's Done, State By State," State Tax Notes, Vol. 4 (May), pp. 1038–51.

Danninger, Stephan, Annette Kyobe, and Marco Cangiano, 2004, "The Political Economy of Revenue-Forecasting Experience from Low-Income Countries," forthcoming IMF Working Paper (Washington: International Monetary Fund).

Golosov, Mikhail, and John King, 2002, "Tax Revenue Forecasts in IMF-Supported Programs," IMF Working Paper 02/236 (Washington: International Monetary Fund).

Kaufmann, Daniel, Aart Kraay, and Massimo Mastruzzi, 2003, "Governance matters III:governance indicators for 1996–2002", World Bank Policy Research Working Paper 3106 (Washington: World Bank).

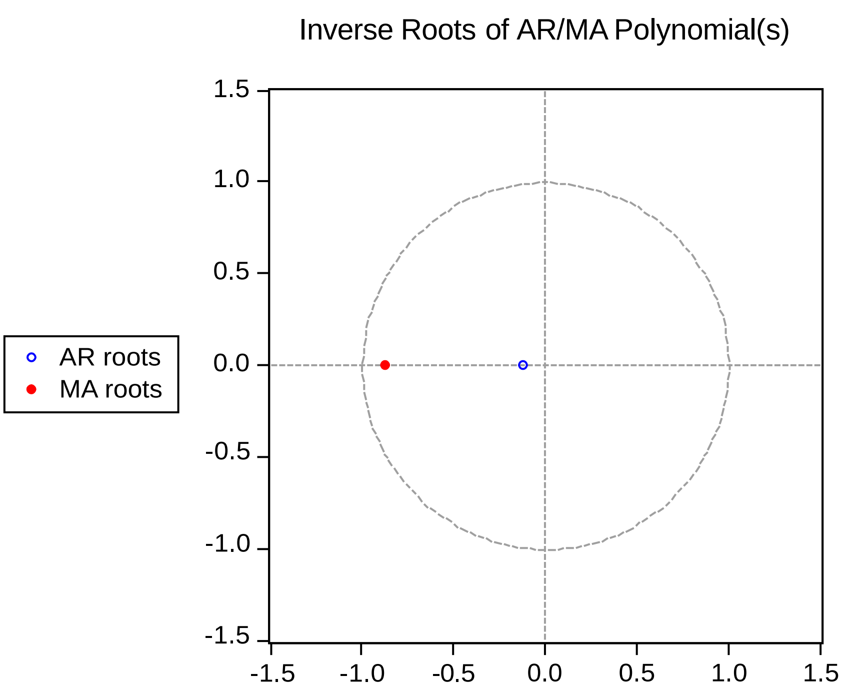
Lienert, Ian C., and Feridoun Sarraf, 2001, "Systemic Weaknesses of Budget Management in Anglophone Africa," IMF Working Paper 01/211 (Washington: International

Appendix I: Hypothesis test for White noise

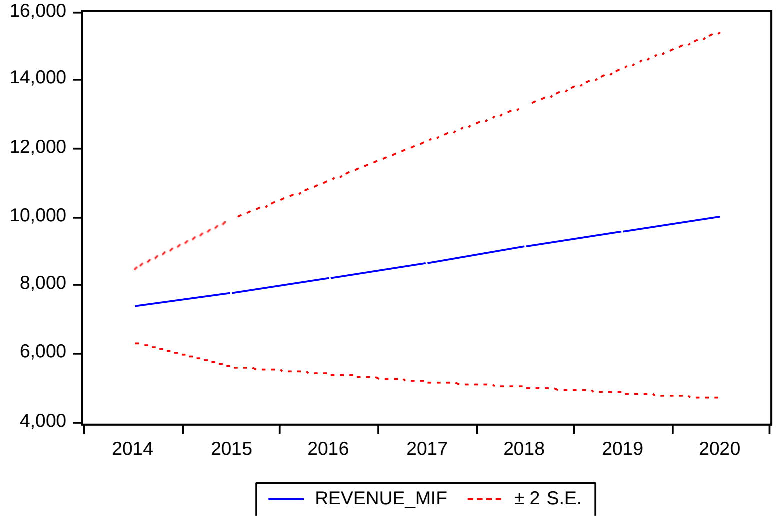
Date: 03/17/23 Time: 18:03
Sample (adjusted): 2001 2014
Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.021	0.021	0.0075	
		2	-0.163	-0.163	0.5033	
		3	0.171	0.184	1.1019	0.294
		4	0.062	0.023	1.1889	0.552
		5	-0.175	-0.128	1.9488	0.583
		6	0.047	0.048	2.0101	0.734
		7	0.009	-0.061	2.0129	0.847
		8	-0.088	-0.026	2.3037	0.890
		9	-0.157	-0.172	3.4036	0.845
		10	-0.103	-0.138	4.0032	0.857
		11	-0.058	-0.069	4.2529	0.894
		12	-0.122	-0.131	5.9102	0.823

Appendix II: Covariance Stationarity test (A.R. roots and M.A. roots)



Appendix III: For caste result



Forecast: REVENUE_MIF	
Actual: REVENUE_MILLIONS_	
Forecast sample: 2014 2020	
Included observations: 7	
Root Mean Squared Error	390.9329
Mean Absolute Error	390.9329
Mean Abs. Percent Error	5.583093
Theil Inequality Coef.	0.027157
Bias Proportion	1.000000
Variance Proportion	NA
Covariance Proportion	NA
Theil U2 Coefficient	NA
Symmetric MAPE	5.431471

Appendix IV: Forested graph