DOI: 10.32602/jafas.2023.033

JOURNAL OF ACCOUNTING, FINANCE AND AUDITING STUDIES

http://www.jafas.org

Modeling and Forecasting Unemployment Rate in Tanzania: An ARIMA Approach

Peter E. Tengaa[a] **Yohana M. Maiga**[b] **Amos M. Mwasota**[c]

[a] College of Business Education, Lecturer - Department of Mathematics and ICT, tengaason@gmail.com
[b] Tanzania Institute of Accountancy, Lecturer - Department of Mathematics and Statistics, yohanamaiga2016@gmail.com
[c] College of Business Education, Lecturer - Department of Mathematics and ICT, mwasotea@yahoo.com

Keywords
Box-Jenkins, Unemployment rate, ARIMA, Forecasting.

Jel Classification
L32.

Paper Type
Research Article

Received
07.05.2023

Revised
12.06.2023

Accepted
28.06.2023

Abstract

Purpose: This study aims to develop a reliable forecasting approach for Tanzania's unemployment rate and provide policymakers with an effective tool for decision-making. Unemployment forecasting is vital for informed policymaking, particularly in countries like Tanzania.

Methodology: This study employs a quantitative research design and adopts Box Jenkin's methodology and the ARIMA (AutoRegressive Integrated Moving Average) model for unemployment forecasting in Tanzania. The entire available dataset for the specified period is utilized, employing a non-probability sampling technique. Diagnostic tests, including ACF (AutoCorrelation Function), PACF (Partial AutoCorrelation Function), and unit root analysis, are conducted to guide the optimal model selection. Differencing addresses non-stationarity in the time series data by removing trend and seasonality effects. The optimal model selection is based on criteria such as AIC (Akaike Information Criterion), Schwartz, and Hannan-Quinn.

Findings: The study finds that the ARIMA (3,1,4) model demonstrates superior performance in forecasting the unemployment rate in Tanzania. Diagnostic checks validate the adequacy of the model, revealing white noise residuals. The forecasts indicate a consistent downward trend in unemployment rates over the next nine years, suggesting potential labour market improvements in Tanzania. These findings enhance our understanding of Tanzania's unemployment dynamics and provide valuable insights for policymakers.

Originality/Value: The study lies in its application of Box Jenkin's methodology and the ARIMA model to unemployment forecasting in Tanzania. By utilizing the entire available dataset and employing diagnostic tests for model selection, the study enhances the reliability of the forecasting approach. The study offers policymakers an informed decision-making tool by providing accurate forecasts and capturing underlying trends.

Introduction

Forecasting is a valuable technique that utilizes past data to predict future trends in various phenomena (Didiharyono and Syukri, 2020). It plays a crucial role in our daily lives by providing insights into future events. Among the macroeconomic indicators that can be forecasted, the unemployment rate stands out.

Unemployment refers to the situation where individuals who are actively seeking employment are unable to find suitable jobs, not due to their inability to work. The unemployment rate quantifies the number of people in this category (Desaling Germay, 2016). High unemployment rates indicate a struggling economy, while low rates signify an improved standard of living and overall economic well-being.

There are several causes of unemployment, including structural, frictional, classical, voluntary, and cyclical factors. Structural unemployment occurs when individuals lack the necessary skills for available jobs, while frictional unemployment arises from delays in finding new employment. Voluntary unemployment occurs when individuals choose not to work due to inadequate wages compared to the job requirements. Cyclical unemployment, on the other hand, emerges from a decrease in labor demand caused by economic factors. These factors collectively contribute to an excess of labor supply over demand. Unemployment is a globally debated issue that affects individuals and nations alike, and finding solutions remains a challenge worldwide (Didiharyono and Syukri, 2020).

In a study conducted by Dritsaki (2016) on the employment market in the Czech Republic, Box-Jenkin's method was employed to model the unemployment rate from 2004 to 2012. The study obtained possible time series models, such as SARIMA (1, 1, 0) (1, 1, 0)12, and used them to forecast the unemployment rate for the Czech Republic. The impact of unemployment extends from individuals to society and the entire nation. Therefore, reduced spending leads to decreased contributions to the economy in terms of goods and services produced, highlighting the adverse effects of unemployment on all productive sectors.

In Tanzania, unemployment represents a significant challenge to national development across various sectors. Over the years, job creation in Tanzania has not kept pace with the growth of the labor force. The unemployment problem in Tanzania can be traced back to the 1970s, when the country faced an economic crisis (Jelena et al., 2017). This study aims to forecast Tanzania's unemployment rates using time series data from 1991 to 2021, analyzing the trends in unemployment over this period.

Unemployment has persistently posed a major challenge in Tanzania, despite government efforts to create job opportunities. The high unemployment rate has adversely affected the country's economic growth and social development. Therefore, accurate modeling and forecasting of unemployment rates are essential to assist policymakers in making informed decisions.

Consequently, this study seeks to model and forecast the unemployment rate in Tanzania from 1991 to 2021. By employing time series analysis, the study aims to identify the key drivers of unemployment and develop a model that accurately captures its dynamics. Furthermore, the study will generate forecasts of the unemployment rate for the next nine years, providing valuable policy formulation and implementation insights.

Research Methodology

The study employed a quantitative and empirical research design, utilizing unemployment data sourced from secondary sources, including the World Bank and the National Bureau of Statistics (NBS) Tanzania, spanning the period from 1991 to 2021. The analysis involved applying the Autoregressive Integrated Moving Average (ARIMA) stochastic model, introduced initially by Box-Jenkins (1976). The ARIMA (p, d, q) model incorporates the Autoregressive (AR) component, capturing the relationship between present and past values, along with a random component and a Moving Average (MA) component that reflects the influence of past residuals.

The data analysis process involved several key steps. Initially, the data underwent pre-processing, which included checking for missing values, outliers, and errors. Data cleaning and transformation techniques were applied to ensure the data's suitability for analysis. The identification of appropriate ARIMA models was carried out by analyzing the autocorrelation and partial autocorrelation functions. The estimation of ARIMA model parameters was performed using maximum likelihood estimation.

The evaluation of ARIMA models was based on a thorough residual analysis, encompassing assessing the residuals' normality, stationarity, and autocorrelation. The selection of the optimal ARIMA model was determined based on criteria such as the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC). Subsequently, the selected ARIMA model was employed to generate forecasts for the unemployment rate in Tanzania for the next nine years (Dritsakis and Klazoglou, 2018; Reinsel et al., 1994).

The ARIMA process can be defined as a combination of autoregressive, differencing, and moving average components, which collectively enable the modeling and forecasting of time-series data. Applying ARIMA modeling to the unemployment data in this study allowed for capturing the relationships and patterns within the data to provide valuable insights into future unemployment trends. The ARIMA process is defined as:

$$\varphi(B)(\Delta^d y_t - \mu) = \theta(B)e_{t,}$$

where

y_t=represents unemployment rate,

$\mu =$mean of $\Delta^d y_{t,}$

$\varphi(B) = 1 - \varphi B - \cdots - \varphi_p Bp,$

$\varphi(B) = 1 - \varphi B - \cdots - \varphi_q Bq,$

$\varphi(B) =$ represents the autoregressive (AR) component,

$\Delta^d y_t - \mu =$ represents the differenced time series,

$\theta(B) =$ represents the moving average (MA) component,

$e_t =$ represents the error term.

p, q and d denote the autoregressive, moving average and differenced order parameter of the process, respectively.

Δ and B denote the difference and backward shift operators respectively, The model estimation involves three distinct steps: identification, parameter estimation, and diagnostic checking.

Findings and Discussion

The future unemployment rates for Tanzania were modeled using the Autoregressive Integrated Moving Average (ARIMA) time series model. The modeling process followed the well-established Box-Jenkins Methodology, which incorporates both autoregressive (AR) and moving average (MA) components (Box and Jenkins, 1970). The results of the modeling process are presented as follows:

1.1 Descriptive Statistics

Descriptive statistics were employed to analyze and summarize the unemployment data of Tanzania from 1991 to 2021. Measures such as mean, median, standard deviation, and percentiles were used to understand central tendencies, dispersion, and distributional properties of unemployment rates. Graphical representations, like line charts and histograms, visualized unemployment's distribution, and temporal variations. These descriptive statistics provided insights into average rates, the spread of values, and overall patterns. Figure 1 presents detailed information on the descriptive analysis, laying the groundwork for subsequent analysis and modeling.

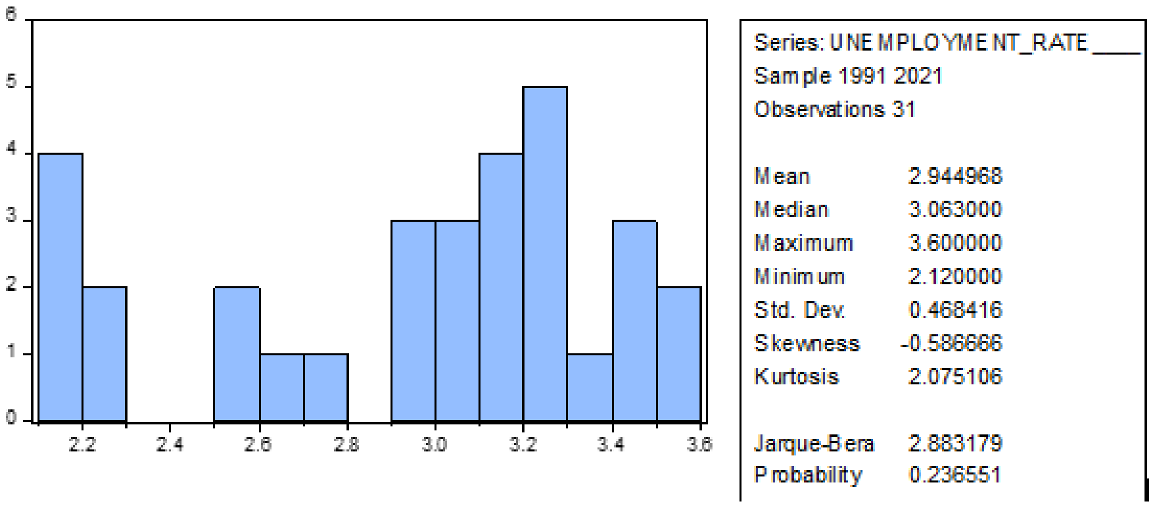


Figure 1: Descriptive statistics for Tanzania's Unemployment rate 1991–2021

In Figure 1, the study utilized central tendency and dispersion measures to analyze the unemployment rate data. The central tendency measures, such as the mean and median, provide insights into the average and middle values of the unemployment rate data set. The mean unemployment rate of 2.94 indicates that, on average, the unemployment rate over the study period was 2.94. The median unemployment rate of 3.06 represents the middle value, with half of the data falling below this value and half above. These measures help us understand the central value around which the unemployment rate data is clustered, aiding in identifying trends, patterns and changes over time.

Measures of dispersion, such as the standard deviation, assess the spread of the data around the mean unemployment rate. With a standard deviation of 0.45, the data is tightly clustered around the mean, indicating less variability. A small standard deviation suggests that the unemployment rate remains close to the mean. Understanding measures of dispersion is crucial for assessing data reliability and making accurate forecasts and predictions.

The kurtosis measure examines the shape of the distribution compared to the normal distribution. A positive kurtosis value of 2.07 indicates a relatively peaked distribution, with the unemployment rate data more concentrated around the mean and fewer extreme values. Conversely, a negative kurtosis value would suggest a more spread-out distribution with more extreme values.

Skewness measures the asymmetry of the distribution. A positive skewness value indicates a right-skewed distribution, with a long tail on the right-hand side, suggesting a higher risk of very high unemployment rates. Conversely, a negative skewness value of -0.58 indicates a left-skewed distribution, with a long tail on the left-hand side, suggesting a higher risk of low unemployment rates.

In summary, the measures of central tendency and dispersion provide valuable insights into the characteristics of the unemployment rate data. The analysis of kurtosis and skewness helps understand the distribution shape and potential risks associated with extreme unemployment rate values. These measures contribute to informed modeling and forecasting of the unemployment rate in Tanzania.

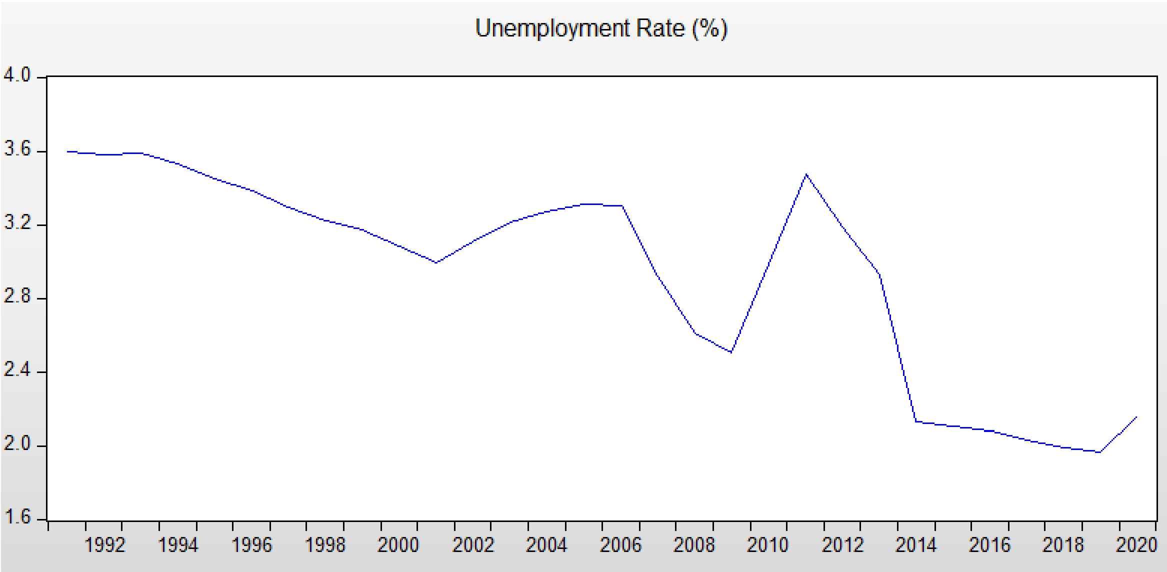


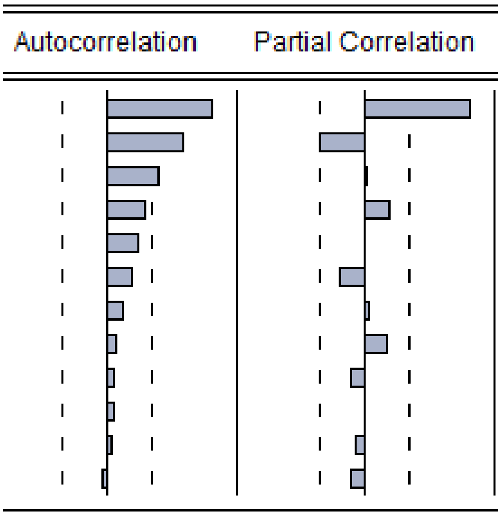
Figure 2: Trend of Tanzania's Unemployment rate from the period 1991–2021

Figure 2 shows the plot of the Unemployment rate data from 1991 to 2021, and the graphs depict the significant decreases in the unemployment rate in Tanzania.

3.2 Identification Step of the Model

The identification step involves utilizing techniques to determine the values of p, q, and d in the ARIMA modeling process. These values are determined by analyzing the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). In an ARIMA (p, q, d) process, the theoretical PACF exhibits non-zero partial autocorrelations at lags 1, 2,..., p, and zero partial autocorrelations at all other lags. Similarly, the theoretical ACF displays non-zero autocorrelations at lags 1, 2,...,q, and zero autocorrelations at all other lags.

By examining the non-zero lags in the PACF and ACF samples, we can tentatively identify the p and q parameter values. To assess the stationarity of the time series, both the ACF and PACF can be used, along with unit root tests like the Augmented Dickey-Fuller (ADF) test or other relevant tests. If the time series is non-stationary, differencing is applied to make it stationary. The number of differencing steps required determines the value for the d parameter (Suleman and Sarpong, 2012).



	AC	PAC	Q-Stat	Prob
1	0.849	0.849	23.872	0.000
2	0.620	-0.362	37.057	0.000
3	0.412	0.009	43.090	0.000
4	0.301	0.194	46.444	0.000
5	0.256	0.001	48.966	0.000
6	0.194	-0.190	50.465	0.000
7	0.117	0.028	51.039	0.000
8	0.079	0.173	51.314	0.000
9	0.062	-0.114	51.489	0.000
10	0.065	0.001	51.691	0.000
11	0.033	-0.078	51.747	0.000
12	-0.039	-0.099	51.829	0.000

Figure 3: Autocorrelation and Partial autocorrelation at level

Based on the correlogram analysis, two potential models emerge. The first model is an Autoregressive (AR) model of order 1, indicated by the geometric decay observed in the Autocorrelation Function (ACF) and the significant spike at lag 1 in the Partial Autocorrelation Function (PACF), as shown in Figure 3. This suggests a relationship between the present value and the immediate past value.

The second potential model is an Autoregressive (AR) model of order 2, characterized by the geometric decay observed in the ACF and a significant spike at lag 2 in the PACF, as depicted in Figure 3. This indicates a relationship between the present value and the values at lags 1 and 2.

The correlogram analysis helps identify these potential models by examining the decay patterns in the ACF and the significant spikes in the PACF. Further analysis and comparison of these models would be necessary to determine the most appropriate one for modeling the data accurately.

Table 1: Unit root test for Tanzania's unemployment rate

Ho: UNEMPLOYMENT RATE has a unit root Exogenous: Constant

		t-Statistic	Prob.*
ADF test statistic		−1.754095	0.3944
Test critical values:	1% level	−3.689194	

5% level −2.971853

10% level −2.625121

Dependent Variable: D(UNEMPLOYMENT RATE) Method: Least Squares

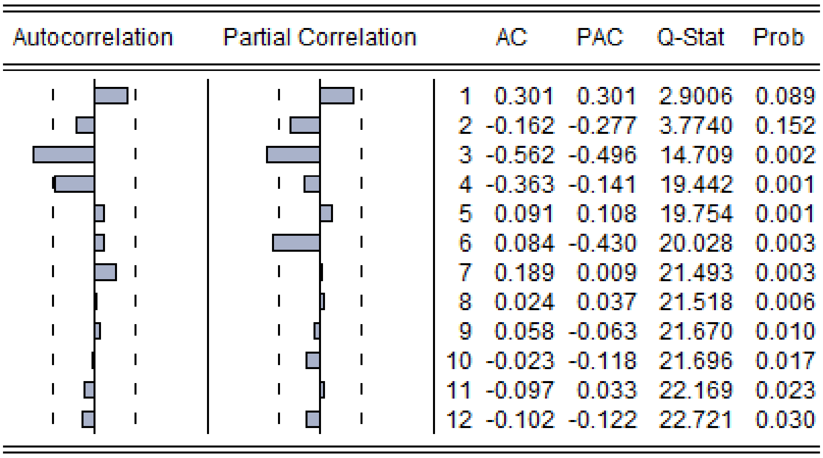
Variable	Coefficient	Std. Error	t-Statistic	Prob.
UNEMPLOYMENT RATE (-1)	−0.147942	0.084341	−1.754095	0.0917
D(UNEMPLOYMENT RATE(-1))	0.418315	0.192911	2.168438	0.0398
C	0.409002	0.255390	1.601476	0.1218

Table 1 presents the unit root test results using the Augmented Dickey-Fuller (ADF) test. The p-value of 0.3944 exceeds the significance level of 0.05, indicating that the time series data is non-stationary. In non-stationary time series, the mean and variance change over time, making them challenging to model and analyze.

To address the issue of non-stationarity, differencing is applied to the data. Differencing involves subtracting the current observation from the previous observation to eliminate the trend component. This transformation helps make the time series stationary by ensuring a constant mean over time.

By differencing the data, we aim to remove any temporal dependence and make the statistical properties of the time series constant. Stationary time series are more amenable to modeling and analysis as their characteristics remain consistent over time.

Therefore, this study performs differencing on the non-stationary time series data to ensure its stationarity and facilitate further modeling and analysis.



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.301	0.301	2.9006	0.089
		2	-0.162	-0.277	3.7740	0.152
		3	-0.562	-0.496	14.709	0.002
		4	-0.363	-0.141	19.442	0.001
		5	0.091	0.108	19.754	0.001
		6	0.084	-0.430	20.028	0.003
		7	0.189	0.009	21.493	0.003
		8	0.024	0.037	21.518	0.006
		9	0.058	-0.063	21.670	0.010
		10	-0.023	-0.118	21.696	0.017
		11	-0.097	0.033	22.169	0.023
		12	-0.102	-0.122	22.721	0.030

Figure 4: First difference in Tanzania's unemployment rate

Table 2: First difference in Tanzania's unemployment rate

H_0: D(UNEMPLOYMENT RATE) has a unit root Exogenous: Constant

t-Statistic Prob.*

ADF test statistic −4.998362 0.0004

Test critical values: 1% level −3.711457

5% level −2.981038

10% level −2.629906

Dependent Variable: D (UNEMPLOYMENT RATE)

Method: Least Squares

Variable	**Coefficient**	**Std. Error**	**t-Statistic**	**Prob.**
D(UNEMPLOYMENT RATE(-1))	−1.354150	0.270919	−4.998362	0.0001
D(UNEMPLOYMENT RATE(-1),2)	0.604384	0.212417	2.845267	0.0094
D(UNEMPLOYMENT RATE(-2),2)	0.517612	0.188281	2.749147	0.0117
C	−0.073567	0.044272	−1.661698	0.1108

Based on the findings in Table 2, the p-value obtained from the unit root test is 0.0004, which is smaller than the predetermined significance level of 0.05. This indicates that the time series process becomes stationary after applying the first difference. By differencing the data, the mean and variance of the time series are stabilized, making it suitable for further analysis. Furthermore, Figure 4 provides insights into the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF). These plots help identify potential models and determine the best model for forecasting Tanzania's unemployment rate. By examining the ACF and PACF in Figure 4, we can observe the decay in the ACF and significant spikes in the PACF at specific lags. These patterns suggest the presence of autocorrelation in the time series data. Potential models can be deduced using the information from the ACF and PACF plots. The best model can be determined by considering various criteria such as model fit, goodness-of-fit measures, and diagnostic tests.

Identifying the best model is crucial for accurately forecasting Tanzania's unemployment rate, as it ensures that the chosen model captures the underlying patterns and dynamics of the data.

3.3 Estimation of the Model

The second step involves estimating the parameters of the selected tentative models. The model selection criterion utilized in this study is based on evaluating the significance of the Auto Regressive Moving Average (ARMA) components and comparing criteria such as the Akaike information criterion (AIC), Schwarz criterion, and Hannan-Quinn criterion. Smaller values of these criteria indicate a better fit of the model.

Following this process, two potential models were identified, and important conditions were compared for ARIMA (3,1,3) and ARIMA (3,1,4) models. The model selection results are presented in Table 3.

Table 3: ARIMA (3,1,3) Model Selection

Variable	Coefficient	Std. Error	T - statistic	Prob
C	-0.052060	0.011442	-4.59922	0.0001
AR (3)	-0.038526	0.203297	-0.189505	0.8512

MA (3)	-1.00000	2380.977	-0.000420	0.9997
Criterion				
AIC	-0.261416			
Schwarz criterion	-0.03282			
Hannan Quin crit	-0.130251			

Table 4: ARIMA (3,1,4) Model Selection

Variable	**Coefficient**	**Std. Error**	**T - statistic**	**Prob**
C	-0.052621	0.014775	-3.5614	0.0015
AR (3)	-0.517334	0.134613	-3.843122	0.0007
MA (4)	-0.425262	0.142645	-2.981258	0.0063
Criterion				
AIC	-0.301074			
Schwarz criterion	-0.112482			
Hannan Quin crit	-0.242010			

Based on the information presented in Table 4, it is evident that the ARIMA (3,1,4) model exhibits smaller values for various criteria such as AIC, Schwarz criterion, and Hannan Quinn criterion compared to the ARIMA (3,1,3) model presented in Table 3. These criteria indicate model fit, with smaller values indicating a better fit to the data.

Additionally, the AR, MA, and constant terms in the ARIMA (3,1,4) model have significant p-values, further supporting its suitability for forecasting. A significant p-value suggests that the coefficients associated with these terms are statistically different from zero, indicating their importance in capturing the underlying patterns and dynamics of the data.

Considering the favorable values of the model selection criteria and the significance of the model's terms, the ARIMA (3,1,4) model emerges as a viable option for forecasting the unemployment rate in Tanzania. This model is expected to provide reliable and accurate forecasts based on its ability to capture the relationships between past and present values and the influence of past residuals.

By choosing the ARIMA (3,1,4) model, we can leverage its forecasting capabilities to make informed predictions of unemployment rates in Tanzania.

3.4 Diagnostic Checking

After estimating the model, it is important to conduct diagnostic checks to ensure its adequacy in representing the series. These checks involve analyzing the residuals to determine if they exhibit randomness and normal distribution. The Anderson-Darling test for normality was employed for this purpose. Additionally, a plot of the residuals against the fitted values was used to assess if the residuals are randomly scattered. To further assess the model's adequacy, the Ljung-Box Q statistics were utilized. This statistic provides an overall evaluation of the model by testing the null hypothesis that the residuals follow a white noise process.

Another crucial step in the analysis is to examine the residual diagnostics to ensure a stable univariate process. This involves verifying if the residuals satisfy the conditions of covariance stationary and invertibility. The test statistic for this assessment is given by:

$$\emptyset_n = m(m+2)\sum_{i=1}^{m}(m-i)^{-1}S_i^2 \approx X_{n-s,}^2$$

where

S_i^2 = the residuals autocorrelation at lag i,

$m =$ the number of residuals,

$n =$ the number of time lags included in the test.

When the p-value associated with the $\emptyset$ is large the model is considered adequate, else the whole estimation process has to start again in order to get the most adequate model.

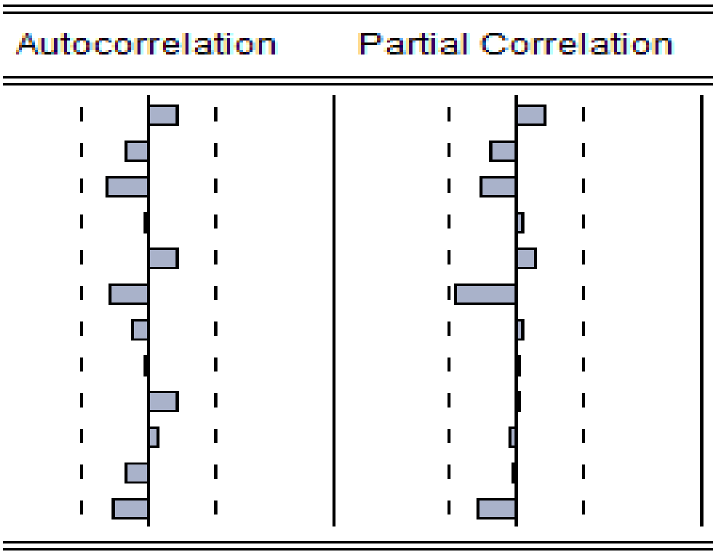


Figure 5: Residual diagnostic test for Tanzania Unemployment rate

Based on the diagnostic test results, it can be observed that the values of both the autocorrelation function (ACF) and partial autocorrelation function (PACF) lie within the confidence bounds. This suggests that the residuals exhibit characteristics of white noise, indicating that the model has effectively captured any remaining patterns or correlations.

Furthermore, the p-values associated with the diagnostic tests in Figure 5 are more significant than the significance level of 0.05. This leads to the acceptance of the Null Hypothesis (H0), which states that no significant correlation or pattern is remaining in the residuals. Therefore, we can conclude that the residuals are random and independent, supporting the model's validity.

The fact that the ACF and PACF values are within the confidence bounds and the p-values are greater than 0.05 provides evidence that the ARIMA model adequately captures the underlying patterns and correlations in the data. This strengthens our confidence in the model's ability to generate reliable and accurate forecasts.

Overall, the diagnostic test results indicate that the model's residuals exhibit the characteristics of white noise, signifying that the model is a suitable representation of the data and that any remaining patterns have been accounted for.

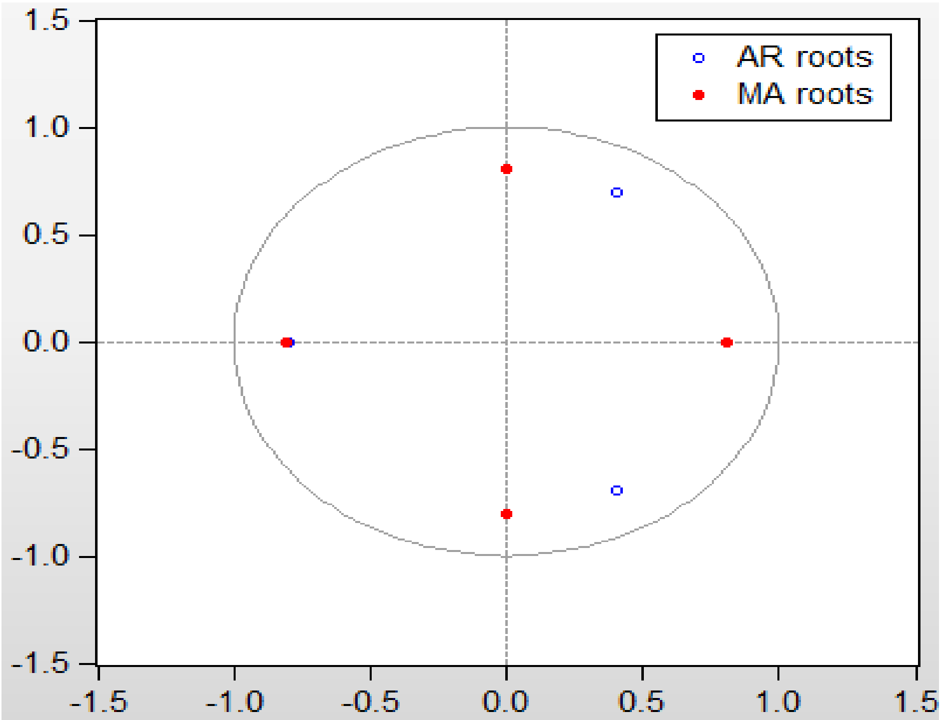


Figure 6: Inverse roots of AR/MA Polynomials

Figure 6 presents the inverse roots of the autoregressive (AR) and moving average (MA) polynomials. The roots of the AR polynomial are all contained within the unit circle, indicating covariance stationarity. Similarly, the roots of the MA polynomial are also situated inside the unit circle. These observations satisfy the criteria for a suitable model to generate accurate forecasts. However, if these criteria still need to be met, it is necessary to repeat the estimation process.

3.5 Forecasting

Accurate forecasting of the unemployment rate is essential for policymakers in Tanzania. This study utilizes the ARIMA approach to develop a reliable forecasting model. By analyzing historical data and employing diagnostic tests, the study aims to identify the best-fitting ARIMA model. The forecasts obtained will provide valuable insights into future unemployment trends and assist policymakers in formulating effective labor market policies.

Table 5: Forecasted unemployment rates in Tanzania 2022–2030

Year	Forecasted value
2022	2.0631
2023	1.9395

2024	1.8134
2025	1.7461
2026	1.7302
2027	1.7157
2028	1.6706
2029	1.5990
2030	1.5267

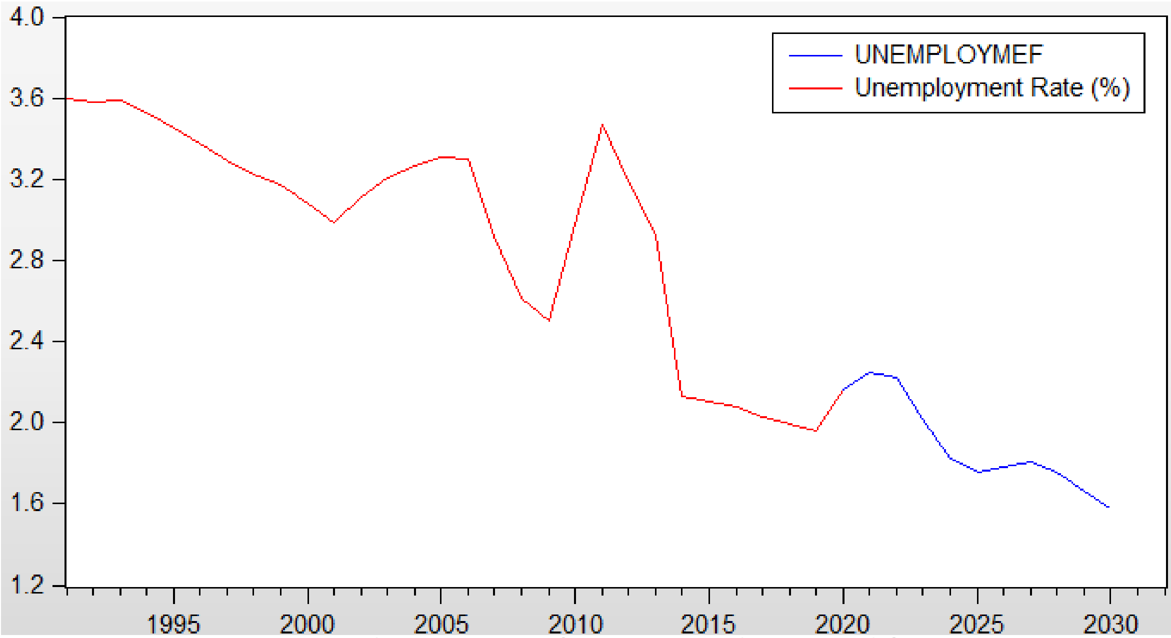


Figure 7: Tanzania Unemployment rate forecast trend in period from 2022–2030

Figure 7 displays the unemployment rate in Tanzania, encompassing actual recorded values spanning from 1991 to 2021, along with projected values spanning from 2022 to 2030 provided by the World Bank (WB). The forecasted values reveal a noticeable rise in the unemployment rate for 2022 and 2023, followed by a decline in 2027 and 2028. Table 5 presents the projected values for the unemployment rate in Tanzania. It is worth noting that Tanzania's unemployment rate demonstrates a downward trend and does not exhibit distinct seasonal or cyclical patterns.

Discussions

The discussion section focuses on interpreting and analyzing the results obtained from the study. In this case, we discussed various aspects of modeling and forecasting the unemployment rate in Tanzania.

Firstly, we highlighted the reliability of the developed model in accurately predicting the unemployment rate in Tanzania. By applying the Autoregressive Integrated Moving Average (ARIMA) time series model, we were able to capture the underlying patterns and trends in the data and make reliable forecasts. This finding aligns with related studies conducted in Greece by Dritsaki (2016) , Meyer and Tasci (2015); Samuel et al. (2021) in the USA, and Nigeria, which also utilized time series analysis and ARIMA models to forecast unemployment rates with satisfactory results.

However, it is important to acknowledge the limitations of the study. One limitation is that the model considered only the unemployment rate and did not incorporate other factors that may affect unemployment rates, such as economic growth, the impact of COVID-19, inflation, and population growth. Integrating these variables into the model could provide a more comprehensive understanding of the dynamics influencing unemployment rates in Tanzania.

Another area for improvement is the time frame of the data used in the study, which only extended up to 2021. To enhance the accuracy and relevance of the model, it would be beneficial to update it with more recent data. This would give researchers and policymakers more up-to-date information for decision-making and forecasting.

Despite these limitations, the study contributes to the existing literature on modeling and forecasting unemployment rates in Tanzania. It emphasizes the usefulness of time series analysis as a reliable method for predicting unemployment rates and understanding their patterns over time. By employing ARIMA models and conducting diagnostic tests, we were able to assess the adequacy of the model and validate its forecasting capabilities.

Conclusions

Unemployment is pervasive in many countries, influenced by factors such as population growth and economic performance. Tanzania faces similar challenges, and this study has presented a reliable model for forecasting the country's unemployment rate. The application of the Box Jenkin's methodology allowed for the identification of the best model by analyzing the Autocorrelation Function (ACF), Partial Autocorrelation Function (PACF), and conducting a unit root test.

The initial analysis revealed that the time series data was non-stationary, indicating the need for differentiation to eliminate trend and seasonality effects. After performing the first difference, two potential models were identified. The best model selection was based on criteria such as Akaike Information Criterion (AIC), Schwartz Criterion, and Hannan-Quinn Criterion. The findings indicated that the ARIMA (3,1,4) model exhibited superior performance, demonstrating lower values of these criteria.

Furthermore, diagnostic checks confirmed that the model's residuals exhibited white noise characteristics, validating its appropriateness for forecasting. The forecasted unemployment rates for the next nine years showed a consistent downward trend, suggesting a potential improvement in Tanzania's labor market conditions.

In conclusion, this study provides valuable insights into Tanzania's unemployment rate and offers a robust model that can assist policymakers in making informed decisions regarding labor market policies. However, it is essential to acknowledge the study’s limitations, such as excluding other influential factors like economic growth, the impact of COVID-19, inflation, and population growth. Future research should consider incorporating these variables to develop a more comprehensive model for forecasting unemployment rates in Tanzania.

Overall, the findings of this study contribute to the understanding of unemployment dynamics in Tanzania and provide a useful tool for policymakers to assess and address labour market challenges effectively. By utilizing the ARIMA (3,1,4) model, policymakers can make data-driven decisions and implement targeted interventions to alleviate unemployment and foster sustainable economic growth.

Conflict of Interest

The authors declare no conflicts of interest regarding the publication of this manuscript.

References

DESALING GERMAY, M. (2016). *Modeling and forecasting the unemployment rate in sweden using various econometric measures*.

DIDIHARYONO, D. AND SYUKRI, M. (2020). *Forecasting with arima model in anticipating open unemployment rates in south sulawesi*. Int. J. Sci. Technol. Res, 9(3):3838–3841.

DRITSAKI, C. (2016). *Forecast of sarima models: An application to unemployment rates of Greece.* American Journal of Applied Mathematics and Statistics, 4(5):136–148.

DRITSAKIS, N. AND KLAZOGLOU, P. (2018). *Forecasting unemployment rates in the USA using box- Jenkins methodology.* International Journal of Economics and Financial Issues, 8(1):9.

JELENA, M., IVANA, I., AND ZORANA, K. (2017). *Modeling the unemployment rate at the EU level by using box-Jenkins methodology.* KnE Social Sciences, pages 1–13.

MEYER, B. AND TASCI, M. (2015). *Lessons for forecasting unemployment in the United States*: use flow rates, mind the trend.

REINSEL, G. ET AL. (1994). *Time series analysis: forecasting and control*. Journal of Marketing Research, 14(2):5561569.

SAMUEL, A. N., SAMUEL, B. S., JOHN, D. I., MAGDALENE, P., AND KAJURU, Y. J. (2021). *Modeling Nigeria's unemployment rates: Box-Jenkins approach*. KASU Journal of Mathematical Science, 2(1):28–35.

SULEMAN, N., & SARPONG, S. (2012). *Forecasting milled rice production in Ghana using the Box-Jenkins approach.* International Journal of Agricultural Management and Development, 2(2), 79-84.

WORLD BANK (2020). *World Development Indicators*. Retrieved from https://databank.worldbank.org/reports.aspx?source=world-development-indicators