Journal of Development Research and Studies, Volume (10)/Issue (01), June 2023, pp–557-571

What beyond massification? Forecasting the Future of Enrollment in Algerian Higher Education Using ARIMA Approach

Salah BOUABDALLAH*

Msila University, Algeria, salah.bouabdallah@univ-msila.dz

Received:25/05/2023 Accepted:05/06/2023 Published:10/06/2023

Abstract:

The demand for higher education in Algeria is expanding rapidly, and accurate forecasting methods are necessary for planning its future. This paper aims to investigate whether data on the student population in higher education can be adjusted to Auto Regressive Integrated Model (ARIMA) for forecasting the future trend until 2030. The author compares three ARIMA models in terms of their fit indices and analysis of residuals. The process of estimation and model selection is conducted using the 'auto.arima' function from the ‘fpp2’ package in the statistical software, 'R'. The data consists of the annual recorded population of students from 1963 to 2022, obtained from relevant national and international bodies. The results show a continuous upward demand for higher education in the upcoming years, slightly exceeding two million in 2030. This finding has direct implications on planning the resources to be deployed for the higher education sector on a national level.

Keywords: ARIMA models; forecasting using time series; tertiary student enrollments; massification in higher education.

JEL Classification Codes: C53; I23.

---

* Coresponding Author

1. Introduction

Higher education in Algeria has been shaped by a steady process of massification. The demand for higher education (HE) in Algeria has been rapidly rising since the 1990's, and the trend has little changed since. Data shows that over the last decade, the number of students enrolled in HE has grown by an average of more than 50,000 students per year. This has led to serious challenges in terms the quality of training and graduate employment among others. To cope with this upward trend, the government had to significantly increase the budget allocated to HE. Over a two-decade period, the government's budget for HE has doubled more than eight times, elevating the share of HE allocation in the total budget to 6.83%, compared to 5.21% at the start of the period (Khouathra, 2019). Since HE is almost entirely funded by the State budget, accurately projecting the student population in the coming years is crucial to meet the demand for tertiary education effectively and efficiently. Failure to meet the needs of this massification process could lead to a decline in the quality of training due to inadequate pedagogical facilities and lack in teaching staff.

Sound forecasting methods need to be developed to help authorities make appropriate strategic decisions and improve planning. Despite its necessity, projecting the future demand for HE is complex. The task is particularly difficult when data is annual; to provide sufficient data points the study has to cover a long period; several decades for instance. The Literature shows that various approaches have been applied for forecasting tertiary enrollments depending on the characteristics of the data. However, there is little research on forecasting future tertiary enrollments in Algeria. To address this gap, this study aims to develop a statistically sound forecasting model to better predict the evolution of the tertiary student population at a national level, namely Auto Regressive Integrated Moving Average method, known as ARIMA model. Specifically, the study seeks to answer the following questions:

Does the data on student population in higher education suit ARIMA modeling, and what adjustments are needed? Furthermore, what are the projected numbers until 2030 based on the adjusted model?

After six decades of independence, we estimate that we do now have the possibility to build a solid model provided that reliable data is gathered for the whole period. To do so, we had to search for the number of the student population in Algeria from different resources. The collected data consists of a time series of 60 annual observations. The data shows a quadratic trend and off course no seasonality. It is a time series of the total number of students enrolled in HE institutions from 1963 to 2022. The data comes from the National Bureau of Statistics (ONS), UNESCO's Institute of Statistics (UIS), the Ministry of HE and Scientific Research, and for the last year, the Algerian Press Agency. Three ARIMA models are presented and compared in this paper in terms of their fit indices and analysis of residuals: ARIMA(2,2,3), ARIMA(0,1,1) with drift, and ARIMA(5,1,0). The final model is used to forecast the evolution of the student population until 2030.The results of this research have direct implications for resource planning to meet the needs of the sector. The model can assist authorities at the national and regional level in making better-informed decisions and policies. The model-building approach could also benefit leaders of HE institutions in forecasting future enrollments for the upcoming academic year and extended time periods. This would provide them with greater efficiency in budgeting and resource allocation.

The paper is organized as follows:

1. We provide a brief overview of the existing trend in the demand for HE, globally and in Algeria,
2. We review the literature on forecasting methods applied in the context of HE, with a special focus on the ARIMA method.
3. We describe the data sources, the methodology used to construct and evaluate the models, and the selection criteria.

4. We present a detailed account of the model construction process, its evaluation, and the selection of the best model.

Overall, we discuss the findings and conclude by providing insights on the model and the model-building approach.

2. Literature review

The terms ‘Higher Education’ and ‘Tertiary Education’ are often used interchangeably to refer to post-secondary education. However, according to (Mohamedbhai, 2014), ‘Tertiary Education’ is more encompassing and is commonly used in statistics provided by international bodies, while ‘Higher Education’ mostly refers to post-secondary education that leads to a degree. In the Algerian context, this discrepancy is irrelevant since practically all post-secondary education falls under the latter category.

2.1 The trend in the demand for higher education:

The demand for HE has been increasing rapidly worldwide due to various factors, such as demographic growth, the need for more specialized skills, and the increase in the number of jobs requiring a degree. According to UNESCO's projections, the global demand for HE will increase to 515 million students by 2030 (UNESCO Institute for Statistics, 2019).

Since the early 1900s, the global student population has been on a continuous upward trend, beginning with the United States and spreading to Europe and many other parts of the world, including African countries (Mohamedbhai, 2014). In the ten-year period between 2006 and 2016, the number of students worldwide grew by approximately 50%, from 146 million to over 218 million (Megaud & al., 2019), with projections suggesting that this figure will rise to over 265 million by 2025 (Noui, 2020).

In Algeria, the HE student community has been growing since independence, but the gross enrollment ratio to the relevant age group remained very low in the first three decades after independence. It was in the mid-1990s that the student population began to grow rapidly and has continued to do so until today. According to the Algerian Press Service (2022), Algeria ranked 30th among countries with a student population exceeding one million in 2016, with a student population of 1.3 million at that time. Six years later, the student population has increased to approximately 1.7 million. The annual growth rate reached an average of more than 50 thousand students per year in the last decade.

2-2. Forecasting the trend of enrollments in HE

Quantitative forecasting techniques are often classified into two main categories: curve-fitting methods and causal methods. Causal models attempt to identify the factors that influence enrollment and quantify the contribution of each factor. The model is then used to forecast future enrollment levels, provided that the values of the explanatory variables in that future period can be predicted. Curve-fitting methods, on the other hand, utilize the historical trend and patterns of the time series to predict future enrollment levels.

Forecasting tertiary enrollments accurately can be challenging due to several factors. Firstly, sudden turning points in enrollment patterns can be unpredictable. Secondly, there can be uncertainty about which forecasting method is most appropriate for a particular situation. Lastly, identifying and measuring the various factors that influence tertiary enrollments can be difficult (Chen C. , 2008).

Various methods have been employed to forecast enrollments in HE at both the state and individual institution levels. These methods include the ratio method, subjective judgment, simulation methods, time series analysis, and regression analysis, among others (Chen C. , 2008).

Selecting a particular method depends on the nature and volume of the data available and the degree to which they meet the model assumptions. Time series analysis is regarded as robust and superior to other methods when its assumptions are met and sufficient data is available (Chen, Li, & Hagedorn, 2019). Among these methods, time series methods, ARIMA (Autoregressive Integrated Moving Average) is very popular and has been well documented in the literature. It is a flexible model that fits a large variety of time series trajectories, and it is considered as a sound statistical method and one of the most accurate and powerful approaches for forecasting (Din, 2016).

2-3. ARIMA Modeling

Finding an adequate ARIMA model for the data is a complex process. The researcher first explores the data, mainly through visualization, to identify the series components: trend, seasonality and patterns. He checks the hypothesis of stationarity, either by a formal statistical test, namely ADF: Augmented Dickey-Fuller test, or simply by visual inspection. This can be done by plotting the series and/or the autocorrelation function (ACF) and the partial autocorrelation function (PACF). If the series appears to be stationary (the spikes of ACF and PACF are within the two parallel lines), an ARMA model can be applied in the subsequent step; if not, a transformation is needed, often by differentiation and/or Box-Cox transformation. In this case, an AR(I)MA model is applied.

After obtaining a stationary series; the identification of the model architecture is done, meaning identifying the values of p and q, based on the analysis of the autocorrelation function (ACF) and partial autocorrelation function (PACF) of the transformed data.

The model parameters are then estimated using techniques like the maximum likelihood estimation (MLE).

Then comes the model selection and refinement stage. The model is evaluated through the error variance (sigma squared), the standard error of the estimated model parameters (smaller is better), and the log-likelihood (larger is better). Accuracy measures such as RMSE (Root Mean Squared Error) and MAPE (Mean Absolute Percentage Error) are also calculated. Subsequently, the model is evaluated using diagnostic tests to check for the absence of autocorrelation, assess normality and verify constant variance. The white noise assumption of the residuals is assessed by plotting the residuals against time, examining the ACF and PACF of the residuals, and conducting a portmanteau test, such as the Ljung-Box test. Additionally, the normality assumption of the residuals can be assessed through visual examination such as a histogram or QQ-plot. If the white noise assumption is not met, it suggests the presence of autocorrelation in the residuals, and the model may require further refinement or improvement. If the distribution of the residuals deviates from a normal distribution, this can lead to biased and inconsistent parameter estimates, and tests of the model parameters may not be accurate. If the model does not meet the desired criteria, a process of iteration is initiated by adjusting the model order or considering an alternative model.

After ensuring that the model fits the data well and that the assumption of white noise is satisfied, the researcher may consider comparing the model to other models that may perform well too. Information criteria, such as the Akaike Information Criteria (AIC), Bayesian Information Criteria (BIC), or AICc (AIC corrected), are used for this purpose, as suggested by Hyndman and Khandakar (2008). These criteria balance the goodness of fit of the model with its parsimony in terms of the number of parameters that need to be estimated.

Lastly, the size of the data is of great importance and has to be taken into consideration. The number of observations required by an ARIMA model depends on the complexity of the time series and the chosen model parameters. Usually, a minimum of 50 observations is required, although it is preferable to have more than 100 observations to capture the underlying patterns and trends in the time series and achieve adequate accuracy and precision (Box & Tiao, 1975). Chen (2008) suggests that a minimum of 40 to 50 observations is needed to extract a good ARIMA prediction model. However, this requirement can be challenging to meet with annual observations unless the time series spans a considerable length, as is the case in the present study.

3. Methods and Materials

This paper presents the process of constructing three ARIMA models and selects one of them to forecast the future of student population in HE up to 2030.

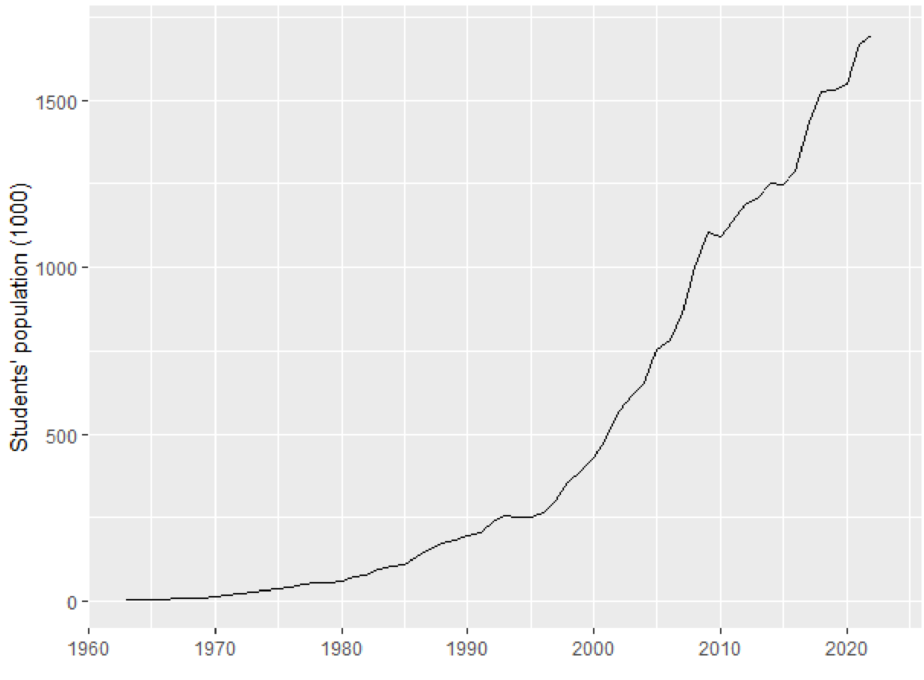
3.1. Data Source

The data used in this study consists of 60 annual observations of the number of students enrolled in HE institutions in Algeria. Nearly all HE institutions in Algeria are state-owned, with few private institutions representing a tiny percentage of the sector. The data covers the period from fall 1962 to fall 2022 and was obtained from relevant national and international institutions, including the National Bureau of Statistics (ONS), the Ministry of HE and Scientific Research, the UNESCO Institute of Statistics (UIS), and, for the last year, the Algerian Press Service (APS).

3-2. Data analysis

Traditionally, forecasting with an ARIMA model can be a laborious process. ARIMA class of models is quite complex and demands some expertise to identify, estimate, and evaluate the model before forecasting. However, there are now several statistical software packages available that can assist researchers by automatically selecting the best model based on a set of criteria. For instance, the 'auto.arima' function in the 'forecast' package of the statistical 'R' software automatically selects the optimal ARIMA model for a given time series based on information criteria. These software packages save time and effort and provide more accurate and reliable results. However, it is important to note that these automated procedures still require careful scrutiny and critical evaluation of the results by the researcher to ensure the validity of the model and the accuracy of the forecast.

Fig.1. The Evolution of Tertiary Education Enrollments in Algeria Since the Independence



Source: Author, based on data collected from: National Bureau of Statistics (ONS), the Ministry of HE and Scientific Research, the UNESCO Institute of Statistics (UIS), and the Algerian Press Service (APS) for the last year

Our analysis consisted of three phases. In Phase 1, we visually inspected the time series to identify its components and patterns. As expected for annual data, no seasonality was observed; however, the plot clearly showed that the series was non-stationary, characterized by a non-linear trend.

In Phase 2, we developed several ARIMA models using the ‘auto.arima’ function from the ‘fpp2’ package of ‘R’ statistical software. This function employs a variation of the algorithm developed by Hyndman & Khandakar (2008), which estimates and tests a large set of models and compares them using unit root tests, Akaike Information Criteria Corrected (AICc) values, and Maximum Likelihood Estimation (MLE). The function then returns the best model based on these criteria. The default values of the function’s arguments are optimized for quick model estimation. As recommended by Hyndman & Khandakar (2008), we modified some of these arguments to enable the function to consider more models and adjusted other arguments to account for the absence of seasonality, as this was confirmed during Phase 1.

The first ARIMA model, ARIMA(2,2,3), was found to be unsatisfactory. We then re-ran the function, modifying the ‘lambda’ argument from its default value of ‘FALSE’ to ‘auto’ in order to introduce a Box-Cox transformation to take into account the existence of a quadratic trend in the series. The resulting model produced by the function was an ARIMA(0,1,1) model with drift. Although this model was an improvement, it still had significant autocorrelation and a skewed histogram of residuals. To address these issues, we re-ran the function again with a modification that excluded models with drift during the search for the best model. This time, the resulting model (ARIMA(5,1,0)) was satisfactory.

The performance of each suggested model was evaluated based on the significance of the parameter estimates, error variance, log-likelihood value, error measures, and information criteria. Additionally, we examined the residuals of each model using the ‘checkresidual’ function in R software. This function provided us with the results of the Ljung-Box test, residuals against time plot, the ACF correlogram (Autocorrelation Function), and the histogram of the residuals with normal curve.

In Phase 3, we applied the final model, ARIMA(5,1,0), to forecast the future of HE enrollments for the coming eight years.

4. Results

In this section, we present the three models that resulted from our analysis and explain why and how we progressed through them to arrive at the final model. We then provide the forecasting results generated by the final model for the next eight years, until 2030.

4-1. Model 1: ARIMA (2,2,3)

The first model was obtained by running the 'auto.arima' function on the original data without transformation. We set the optional arguments 'stepwise = FALSE' and 'approximation = FALSE' to allow the function to take more time in the process of model building and selecting. We also set the optional argument 'seasonal = FALSE' to indicate that the series does not contain seasonality, which simplifies the calculation process (appendix 1).

The model summary is shown in Table 1.

Table 1. Model 1: ARIMA(2, 2, 3)

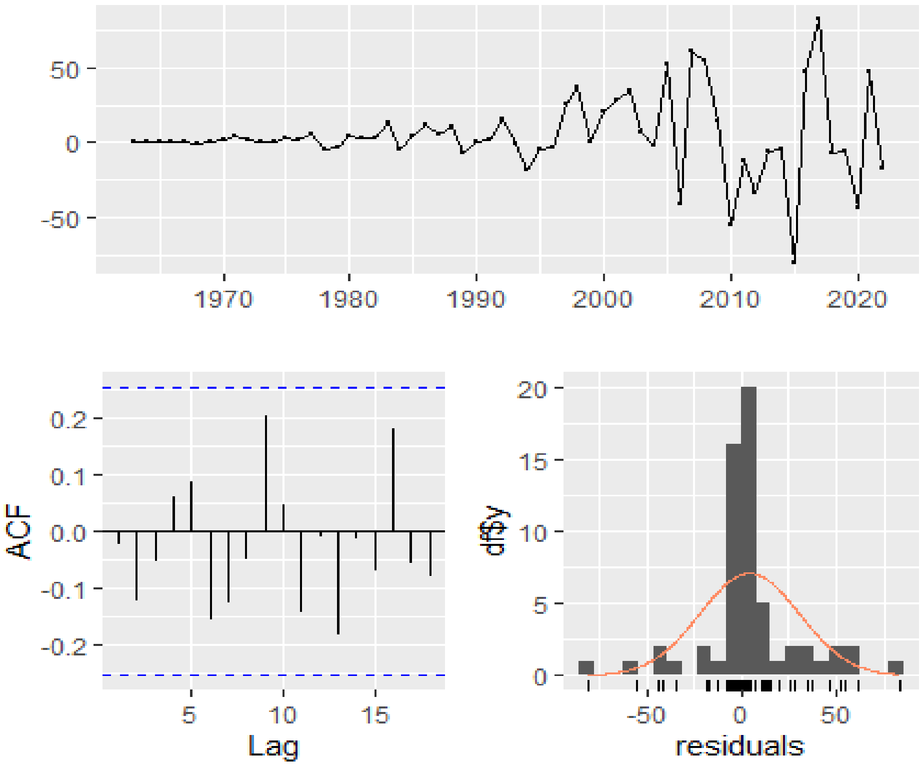
	ar1	ar2	ma1	ma2	ma3
Coefficients	-0.8135	-0.8764	0.2717	0.0254	-0.1762
s.e.	0.1087	0.1177	0.1175	0.1608	0.1218
Sigma² = 788.9; log-likelihood = -274.53 Information criteria(*): AIC = 561.05; AICc = 562.7; BIC = 573.41 Training set error measures(**): RMSE = 26.39784; MAE = 16.0465; MAPE = 4.539017					

(*) AIC: Akaike Information Criterion, AICc: Corrected AIC; BIC: Bayesian Information Criterion; (**) RMSE: Root Mean Squared Error, MAE: Mean Average Error, MAPE: Mean Average Percentage Error.

Source: Author, based on R software output.

As Table 1 shows, the function suggested a model of ARIMA(2,2,3). The parameter estimates are significant, except for one (ma2). The negative log-likelihood is not a good indicator of fitness to the data. The values of AIC, AICc, and BIC, as well as Sigma squared and error measures, are to be compared to values from other models. Next, we conducted a residuals analysis to verify the model assumptions - using the 'checkresidual' function in R software (appendix 2). The absence of autocorrelation was initially assessed using the Ljung-Box test with 10 lags. The test statistic Q* had a value of 8.0504, with 5 degrees of freedom, and a p-value of 0.1535. This indicates no evidence of autocorrelation among the residuals. A visual investigation is provided below.

Fig.2. Residuals from ARIMA(2, 2, 3)



The ACF plot in Figure N° 2 shows no significant autocorrelation, with all spikes within the 95% confidence interval. However, the plot of residuals against time - in the same figure - shows a growing volatility of the errors in the second half of the period, which indicates heteroscedasticity in the residuals. The reason for this is likely the quadratic trend in the data. On the other hand, the histogram of the residuals shows a poor fit to the normal distribution. For these reasons we decided to look for another model.

4-2. Model 2: ARIMA(0,1,1) with drift

As model 1 was unsatisfactory, we introduced a Box-Cox transformation to address the issue of heteroscedasticity and the non-normality of the residuals - by setting the argument 'lambda' to 'auto' in the 'auto.arima' function (appendix 3). This modification helps to take into account the quadratic trend existing in the series. The resulting model is ARIMA(0,1,1) with drift, which is shown in Table 2. It shows that a Box-Cox transformation was applied, with lambda set -automatically by the algorithm - to 0.21667. The drift term (or constant term) allows for a non-zero mean in the time series in the long run after removing the trend and seasonality.

Both model coefficients are statistically significant, which indicates good precision. The log-likelihood is positive, meaning the model is more likely to have generated the observed data than the previous one. The negative information criteria indicate an improved trad-off between goodness of fit and complexity. This model also exhibits slightly better precision as evidenced by its lower error measures compared to model 1.

Table 2. Model 2: ARIMA(0, 1, 1) with drift

	ma1	**drift**
Coefficients	-0.8135	-0.8764
s.e.	0.1087	0.1177
Box-Cox transformation: lambda = 0.2166734 sigma² = 0.02937: log-likelihood = 21.25 Information criteria (*): AIC = -36.5 AICc = -36.06 BIC = -30.27 Training set error measures (**): RMSE = 30.71228; MAE = 18.5018; MAPE = 4.351855		

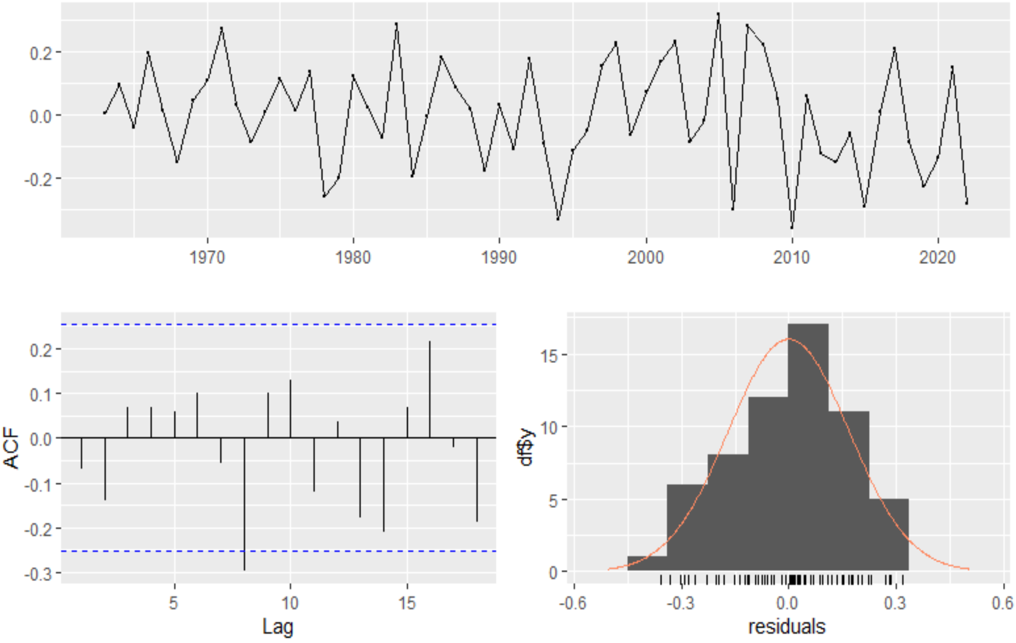
(*) AIC: Akaike Information Criterion, AICc: Corrected AIC; BIC: Bayesian Information Criterion; (**) RMSE: Root Mean Squared Error, MAE: Mean Average Error, MAPE: Mean Average Percentage Error.

Source: Author, based on R software output.

We then analyzed the residuals of the model to check its assumptions. We conducted an Ljung-Box test with 10 lags, which yielded a Q* statistic value of 11.693, with 8 degrees of freedom (df), and a p-value of 0.1654, indicating no significant evidence of autocorrelation.

Figure 3 displays the residuals plot against time, Auto Correlation Function (ACF) and the residuals histogramme (R code in appendix 4).

Fig.3. Residuals from ARIMA(0, 1, 1) with drift



The residuals plot against time indicates that the error term fluctuates around zero with approximately equal variance and no discernible outliers or patterns. Despite these good results, the ACF plot reveals significant autocorrelation at lag 8, and the residuals histogram exhibits a left-skewness. These findings suggest that the current model may not have fully captured certain residual structures, possibly due to the non-linear trend in the data. Consequently, we decided to explore an alternative model without drift -by utilizing the 'auto.arima' function with the 'allowdrift' argument set to 'FALSE'. The the resulting model is presented below.

4-3. Model 3. ARIMA (5,1,0)

This alternative model (ARIMA(5,1,0)) demonstrates performance similar to the previous model in terms of information criteria, error measures, and log-likelihood. However, it includes more parameters, potentially reducing its parsimony as reflected in slightly higher values for the information criteria. The summary of the ARIMA(5,1,0) model is displayed in Table 5 (R code in appendix 5).

Table 3. Model 3. ARIMA(5, 1, 0)

	ar1	**ar2**	**ar3**	**ar4**	**ar5**
Coefficients	0.5141	-0.2511	0.3743	0.0347	0.2669
s.e.	0.1272	0.1456	0.1432	0.1466	0.1282
Box Cox transformation: lambda = 0.2166734 Sigma² = 0.0324; log-likelihood = 18.97 Information criteria (*): AIC = -25.94; AICc = -24.32; BIC = -13.47 Training set error measures (**): RMSE = 30.47038; MAE = 18.28784; MAPE = 4.369409					

(*) AIC: Akaike Information Criterion, AICc: Corrected AIC; BIC: Bayesian Information Criterion; (**) RMSE: Root Mean Squared Error, MAE: Mean Average Error, MAPE: Mean Average Percentage Error.

Source: Author, based on R software output.

This model can be written as follows:

$$\hat{y}_t = c + 0.5141y'_{t-1} - 0.2511y'_{t-2} + 0.3743y'_{t-3} + 0.0347y'_{t-4} + 0.2669y'_{t-5} + e_t$$

Where:

$\hat{y}_t$: the differenced and transformed expected value of the student population at time t.

$y'_{t-1}; y'_{t-2}; \dots; y'_{t-5}$: the differenced and transformed values of the student population at time t-1; t-2, …; t-5.

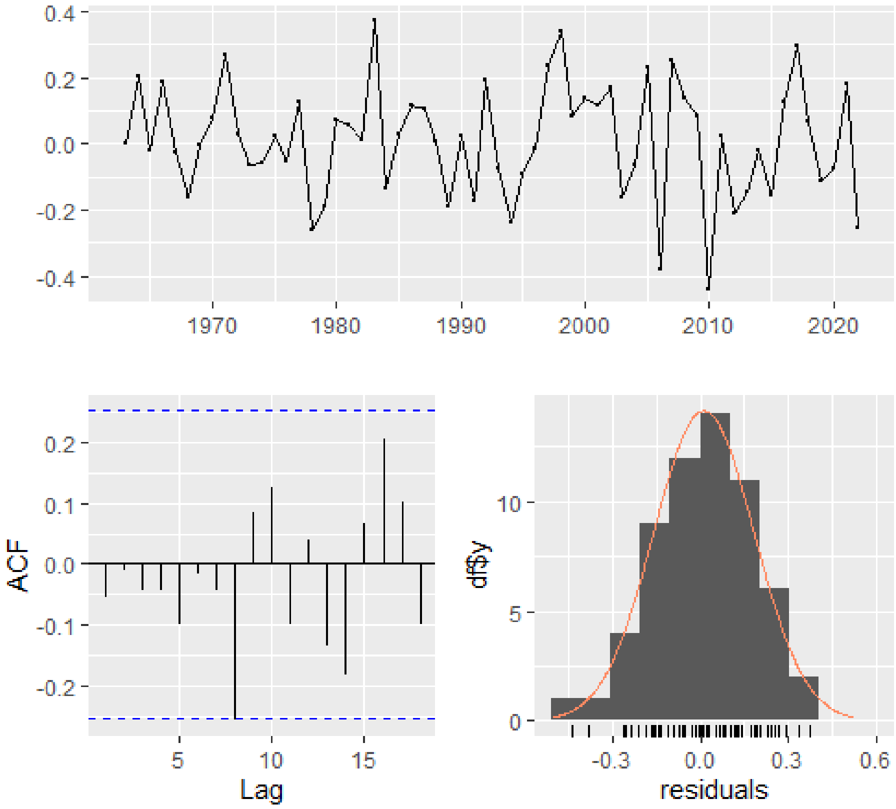
e_t: is the error term.

The residuals of this model were examined using the same diagnostic tools as before (R code in appendix 6), and the results are presented in the following section.

The Ljung-Box test with 10 lags yielded a Q* statistic of 7.7466, with 5 degrees of freedom and a corresponding p-value of 0.1708, which is greater than 0.05. These findings suggest no significant evidence of autocorrelation in the residuals. Therefore, the null hypothesis of white noise can be upheld. This result is further supported by the ACF plot, which reveals no individually statistically significant autocorrelations. Additionally, the plot of residuals against time shows no visible patterns or outliers, with the residuals fluctuating around a mean of zero and displaying stable variance. The histogram of the residuals indicates a good fit to the normal distribution, with no noticeable skewness. These results suggest that the residuals exhibit randomness and adhere to a normal distribution.

Overall, the ARIMA(5,1,0) model fulfills all the necessary assumptions and demonstrates a high level of fitness to the data. While the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC) values are slightly higher compared to the previous model, this model is preferred due to its perfect adherence to the assumptions of white noise and normality. Therefore, it can be considered a suitable candidate for forecasting purposes.

Fig. 4. Residuals from ARIMA(5,1,0)



According to the ARIMA(5,1,0) model, the student population is projected to slightly exceed two million by the year 2030. This prediction aligns well with the previous trend in the demand for higher education in Algeria, making it a plausible estimate. Table 4 displays the points forecasts for the targeted period along with the corresponding 80% and 90% confidence intervals (code in appendix 7).

Table 4. Forecast using model 3: ARIMA(5, 1, 0) until 2030.

Year	Point Forecast	Low 80	High 80	Low 95	High 95
2023	1711.564	1634.293	1791.669	1594.509	1835.246
2024	1764.225	1622.708	1915.233	1551.484	1999.166
2025	1807.623	1620.925	2010.760	1528.410	2125.296
2026	1860.108	1623.856	2122.472	1508.699	2272.618
2027	1905.734	1612.634	2239.010	1472.493	2433.000
2028	1939.650	1582.585	2356.941	1415.690	2604.599
2029	1983.306	1556.166	2497.257	1361.353	2808.558
2030	2029.935	1532.623	2645.586	1311.258	3025.860

The forecasted values, along with the historical data of the student population, are displayed in Figure 5. (R code in appendix 8).

Fig.5. Forecast from ARIMA(5,1,0) until 2030

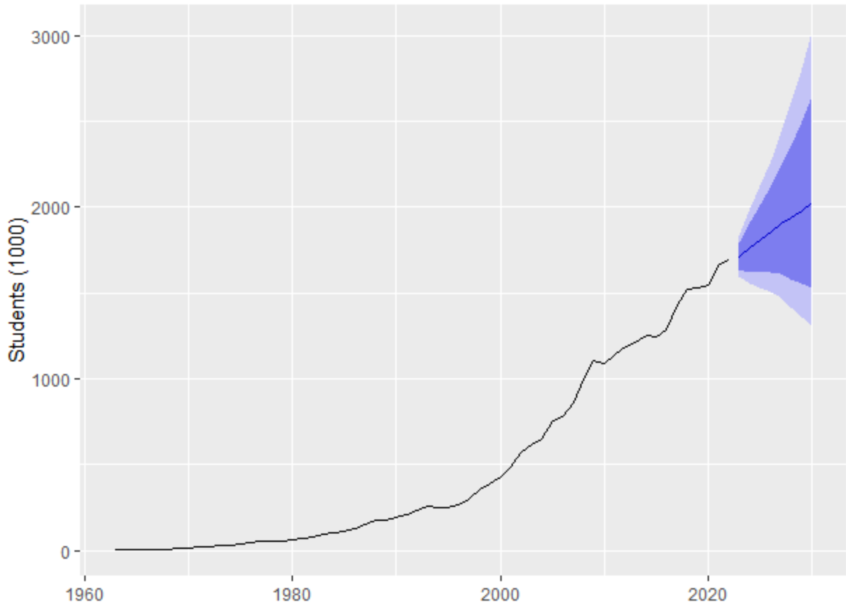


Figure 5 visually represents the predictions described in Table 4. The model shows a consistent upward trend, reminiscent of the pattern observed over the past decade. According to the model, the projected number of students in higher education is approximately 2 million. However, it's important to note that the 80% confidence interval suggests a range of approximately 1.5 million to 2.6 million. This interval accounts for the uncertainty associated with the forecast.

5. Discussion

During the process of selecting the best ARIMA model, we encountered two models that had to be rejected in the diagnostic stage for various reasons. The first model, ARIMA(2,2,3), exhibited no significant autocorrelation but did not provide a good fit for normality. Additionally, it presented serious concerns regarding heteroskedasticity and had a low value of log-likelihood.

On the other hand, the second model, ARIMA(0,1,1) with drift, performed well in most aspects except for normality and a remaining autocorrelation at lag 8.

Although normality is not a strict requirement for ARIMA modeling, it is still desirable feature. When the residuals are normally distributed, it allows for hypothesis testing and the construction of confidence interval for the model parameters. It also required by some diagnostic tests that asses the adequacy of the model. Similarly, while heteroskedasticity is of secondary importance compared to other assumptions in time series regression (Wooldridge, 2015), it can affect the results of hypothesis tests and the confidence intervals around the estimators. However, it does not introduce bias to the estimators of the model's coefficients nor compromise their consistency (Wooldridge, 2015, pp. 622-624).

The final model, ARIMA(5,1,0), demonstrated good performance in terms of fitting the data and satisfying the model's assumptions. According to the model, the number of students in higher education will continue to follow its upward trend, reaching approximately 2.03 million by 2030 (Figure 5). This estimate comes with an 80% confidence interval ranging from 1.53 to 2.65 million (Table 4).

Notably, this projection persists despite the high unemployment rate among graduates which has surpassed that of individuals with no degree (Bouabdallah, 2023). It raises an intriguing question: How do youth respond to shrinking job opportunities for graduates in terms of their investment in education? Investigating the relationship between graduate unemployment and the demand for HE, along with exploring the effect of other potential factors such as the declining social value of a degree and HE institutions prestige, would provide valuable insights into this matter.

In many developed countries, it is well-known that the demand for higher education tends to decrease when the economy provides better opportunities in the labor market. Conversely, during periods of high unemployment and increased competition in the labor market, the demand for higher education typically rises. This is because individuals perceive a higher education degree as a means to enhance their competitiveness and attractiveness in the job market. However, this pattern may not necessarily hold true in developing countries. The situation in Algeria, in particular, may differ due to the unique characteristics of higher education in the country. Notably, in Algeria, all expenses related to higher education, including tuition fees and accommodation, are covered by the state. This has led to a significant increase in the massification of higher education, resulting in a larger number of graduates. Consequently, unemployment among graduates has become a major concern, surpassing the unemployment rate among individuals without a degree.

Finally, as the process of identification, evaluation and selection of the ARIMA models has been conducted using R software by means of ‘auto.arima’ function, it is important to highlight some insights about this function. The function is a valuable tool as it reduces the tedious work involved in selecting the most appropriate model. However, the function has its limitations. While it may provide sometimes an adequate model at the first attempt, in other cases, it may require modifying the function's default arguments. For example, as in the case presented in this paper, the function does not use log or Box-Cox transformation by default. If necessary, the researcher can specify these transformations by modifying the appropriate arguments. Additionally, the function does not test for homoscedasticity or normality during the model selection process, requiring this analysis to be conducted separately afterward. Similarly to any automated process, there is a risk of the function overlooking the best model, necessitating the researcher’s thorough understanding of the algorithm to select an appropriate model (Hyndman & Khandakar, 2008). Overall, while the 'auto.arima' function is very helpful, it should be used with caution and the results should be carefully evaluated.

6. Conclusion

This study aimed to investigate the suitability of ARIMA modeling for predicting the future student population in higher education and the necessary adjustments. Additionally, it aimed to project the demand for higher education until 2030.

Based on our findings, the answer to the first question is affirmative. The ARIMA model, specifically ARIMA(5,1,0), demonstrated good performance in terms of data fitness and adherence to the model's assumptions. The inclusion of a Box-Cox transformation was necessary to address the nonlinear trend, successfully addressing heteroskedasticity issues and deviations from normality observed in other models.

Regarding the second question, the proposed model predicts that massification in higher education is expected to continue along its current trajectory. The number of tertiary students in Algeria is projected to reach approximately 2.03 million by 2030, with an 80% confidence interval ranging from 1.53 to 2.65 million.

These findings raise significant concerns regarding the quality of education, graduate employability, and the government's capacity to meet the growing demands of this sector in the coming years. The sustainability of the current funding model reliant almost solely on the state budget is becoming a pressing issue.

In light of these challenges, it is crucial for the government to explore alternative sources of funding for higher education in Algeria. While the current funding model heavily relies on the state budget, it may not be sustainable in the long run. By involving families, the industry, and the broader economy in the financing of higher education, a more diversified and robust funding system can be established. Additionally, it is essential to prioritize the quality of instruction and the employability of graduates. A holistic approach that combines sustainable funding and a focus on quality will contribute to the development of a flourishing and resilient higher education system in Algeria.

In conclusion, our study sheds light on the future of enrollment in Algerian higher education using the ARIMA approach. The model employed in our analysis demonstrated good performance in terms of accurately capturing the underlying patterns and trends in the data, thereby enhancing its reliability as a predictive tool. The results underscore the importance of proactive measures to ensure sustainable and thriving higher education system in Algeria.

References

Algerian Press Service. (2022, January 4). *Santé, science, technilogie*. Retrieved 5 22, 2023, from APS: www.aps.dz/sante-science-technologie/128538

Bouabdallah, S. (2023). The Impact of Higher Education Massification in Algeria on Quality and Graduate Employment. *Economic Studies, 17*(2).

Box, G. E., & Tiao, G. C. (1975). Intervention analysis with applications to economic and environmental problems. *Journal of the American Statistical Association*, 70-79.

Chen, C. (2008). An integrated enrollment forecast model. *IR Applications, 15.*

Chen, Y. A., Li, R., & Hagedorn, L. S. (2019). Undergraduate International Student Enrollment Forecasting Model: An Application of Time Series Analysis. *Journal of International Students, 9*(1), 242-261.

Din, M. (2016). ARIMA by Box Jenkins Methodology for Estimation and Forecasting Models in Higher Education. *Athens ATINER'S Conference.* Athen: Athens Institute for Education and Research.

Hyndman, R. J., & Khandakar, Y. (2008). Automatic time series forecasting: The forecast package for R., 27(1). *Journal of Statistical Software, 27*(1), 1-22. doi:10.18637/jss.v027.i03

Khouathra, S. (2019). Public Expenditure on Higher Education in Algeria from 2000 to 2018, Reality and Challenges. *Journal of Strategy and Development, 9*(3), 75-95.

Megaud, M., & al. (2019). *La mobilité internationale des étudiants, S'organiser pour les défis à venir.* Retrieved from www.ccompte.fr

Mohamedbhai, G. (2014). Massification in Higher Education institutions in Africa, Causes, Quansequences, and Responses. *International Journal of Higher Education, 1*(1), 59-83. doi:https://doi.org/10.6017/ijahe.v1i1.5644

Noui, R. (2020). Higher Education Between Massification and Quality. *Higher Education Evaluation and Development, 14*(2), 93-103. Retrieved July 16, 2022, from https://doi.org/10.1108/HEED-04-2020-0008

UNESCO Institute for Statistics. (2019, Septembre). *Etudiants inscrits à l'étranger par région d'accueil.* Retrieved November 13, 2021, from data.uis.unesco.org: https://www.campusfrance.org/system/files/medias/documents/2019-12/algerie_fr.pdf

Appendices:

Appendix 1.: R code used to produce Model 1 and its summary.

```
> library (fpp2)
> arima_fit <- auto.arima(data, stepwise = FALSE, approximation =
FALSE, seasonal = FALSE)

>summary(arima_fit)
```

Appendix 2.: R code used to produce Model 1 output for residual analysis.

```
> checkresiduals(arima_fit)
```

Appendix 3.: R code used to produce Model 2 and its summary.

```
> arima_fit2 <- auto.arima(data, stepwise = FALSE, approximation =
 FALSE, seasonal = FALSE, lambda = "auto")

> summary(arima_fit2)
```

Appendix 4.: R code used to produce Model 2 output for residual analysis.

```
> checkresiduals(arima_fit2)
```

Appendix 5.: R code used to produce Model 3 and its summary.

```
> arima_fit3 <- auto.arima(data, stepwise = FALSE,approximation =
FALSE, seasonal = FALSE, lambda = "auto", allowdrift = FALSE)

> summary(arima_fit3)
```

Appendix 6. R code used to produce Model 3 output for residual analysis.

```
> checkresiduals(arima_fit3)
```

Appendix 7. R code used to apply Model 3 to forecast for the next 8 years.

```
>forecast3 <- forecast(arima_fit3, h = 8)
```

Appendix 8. R code used to display a plot for the forecast obtained by Model 3 for the next 8 years.

```
autoplot(forecast3, ylab = "data (1000)", xlab = "")
```