European Journal of Economic and Financial Research
ISSN: 2501-9430
ISSN-L: 2501-9430
Available on-line at: http://www.oapub.org/soc

DOI: 10.46827/ejefr.v8i8.1932 Volume 8 | Issue 8 | 2025

GOOD ACADEMIC GOVERNANCE USING ARIMA MODELS: CASE STUDY OF THE PREDICTION OF NEW STUDENTS AT THE FACULTY OF ECONOMIC AND SOCIAL SCIENCES IN TANGIER, MOROCCO

Ghafir Adil[1],
Bennani Anas[2i],
El Hajjaji Soukaina[3],
Tali Abdelhak[4]
[1]Doctoral Student,
Abdelmalek Essaadi University,
Research Team in Management & Dynamics of Organizations,
Morocco
[2]Professor,
Abdelmalek Essaadi University,
Research laboratory Innovative Systems Engineering,
Morocco
[3]Professor,
Abdelmalek Essaadi University,
Research Team in Management & Dynamics of Organizations,
Morocco
[4]Professor,
Abdelmalek Essaadi University,
Information and Communication Technologies Laboratory,
Morocco

Abstract:
Currently, good governance of Moroccan universities plays a crucial role in guiding their strategies to achieve objectives such as efficient resource allocation and improvement of educational performance. Accurate estimates of future student numbers are essential to actively participate in the effective management of human resources, infrastructure and academic programs. In this context, this work aims to predict the number of students enrolled in the fundamental license of the Faculty of Economic and Social Sciences of Tangier, reporting to the Abdelmalek Essaâdi University in Morocco. The choice of the essential degree is explained by its open access character, which guarantees students wishing to continue their studies in this cycle enrollment without restriction. In addition, this institution was selected because of the particularly high number of new students it registers, surpassing other institutions of the university. This study is based on actual

[i] Correspondence: email abennani@uae.ac.ma

data provided by the establishment, which encourages scientific research to ensure good governance in the coming years. This study was carried out using the ARIMA model.

JEL: I23; C53; H75; O21

Keywords: academic governance, decision-making, forecasting, ARIMA, GARCH

1. Introduction

A study by (Gilbert M. Masinading, 2024) treats enrollment forecasting as a key factor for better decision-making in the management and strategic planning of higher education institutions.

For (A. Dela Cruz, 2020) highlighted that student enrolment forecasting is an asset to higher education institutions (HEIs), because the data obtained supports optimal decisions in future planning.

The same vision is shared by (Ward, 2007). According to him, good financial management depends on a good forecast of student enrolments for the following year. In the same context (Ilan, 2005) adds a competitive vision by the university's good position facing competition. This is achieved through efficient management and better allocation of resources with the help of registration forecasts.

This article highlights the importance of forecasting the number of new entrants to a higher education institution in general, and the establishment of the Faculty of Economic and Social Law Sciences in Tangier in particular, as an important factor to ensure good governance in universities.

2. Objectives

The objective of this study is to produce forecasts of annual enrolment in basic studies licenses for the institution studied. It focuses on analyzing the characteristics of the time series of registrations from available data from 22 observations, developing an ARIMA model to make these predictions, and generating accurate projections of future registrations. This forecast will facilitate strategic decision-making to optimize available resources based on planned staffing.

3. Methodology

3.1 Time series model: ARIMA model

In 1976, a two-person team (George E. P. Box, 1976 introduced the ARIMA model. It's designed to study time series characterized by linearity. It combines autoregressive (AR) and moving average (MA) components. This model is recognized for its high accuracy in short-term forecasting (Neusser, 2016). this model (A. Asrirawan, 2022) can also be applied to non-stationary data through the differentiation process.

The ARIMA model (George E. P. Box, 1976) is noted as follows: ARIMA (p,d,q), in which:

$$y_t = \mu + \sum_{i=0}^{p} \varphi_i * y_{t-i} + \sum_{j=0}^{q} \theta_j * \varepsilon_{t-j} + \varepsilon_t$$

Where:
p = The order of the autoregressive model (AR), representing the number of delays in the model.
D = The integration order, which determines the number of differences necessary to make the series stationary.
q = The order of the moving average (MA) model, representing the number of random noise terms in the model.
y_t = The time series at time t.
μ = The constant or mean of the series.
ϕ_i = The AR coefficients.
θ_j = The MA coefficients.
ε_t = The error term at time t, also called white noise.

3.2 The steps of ARIMA modeling

3.2.1 Model identification

Stationarity tests (D. Wulan Sari, 2016) for the mean and variance are calculated at this stage. It's generally evaluated using the Augmented Dickey-Fuller (ADF) test.

The data patterns are identified in the next step to establish the appropriate values for p and q. This involves creating data plots, selecting appropriate transformations, and calculating the Autocorrelation Function (ACF) and Partial Autocorrelation Function (PACF) of the original data. The ACF and PACF calculations are intended to assert the order of differencing required to achieve stationarity. After the transformation, ACF and PACF calculations are performed, as well as differentiation processes, to identify p and q values and determine the trend value d greater than 0 (zero).

3.2.2 Model parameter estimation

The model is estimated (Søren Johansen, 2012) using the least squares method by searching for parameter values that minimize the sum of errors squared (the difference between actual and predicted values).

At the initial stage, a good estimated model is selected by performing hypothesis tests for the parameters as follows:

- H_0: Non-significant parameters,
- H_1: Significant parameters.

Then, diagnostic tests are performed if the model is considered to be significant.

3.2.3 Diagnostic check

A test of the white noise hypothesis (Lan Luo, 2024) is performed taking into account the delays in the residual model, and using the Ljung-Box test and the unilateral test, where a p-value is greater than 0.05, it is considered an indicator of satisfactory results. White noise tests can be performed using the Ljung-Box test.

The model's effectiveness in forecasting can be measured using evaluation criteria such as RMSE (Root Mean Squared Error) or MAPE (Mean Absolute Percentage Error) (Ahmar, 2020), which assesses the accuracy of the forecasts generated by the model.

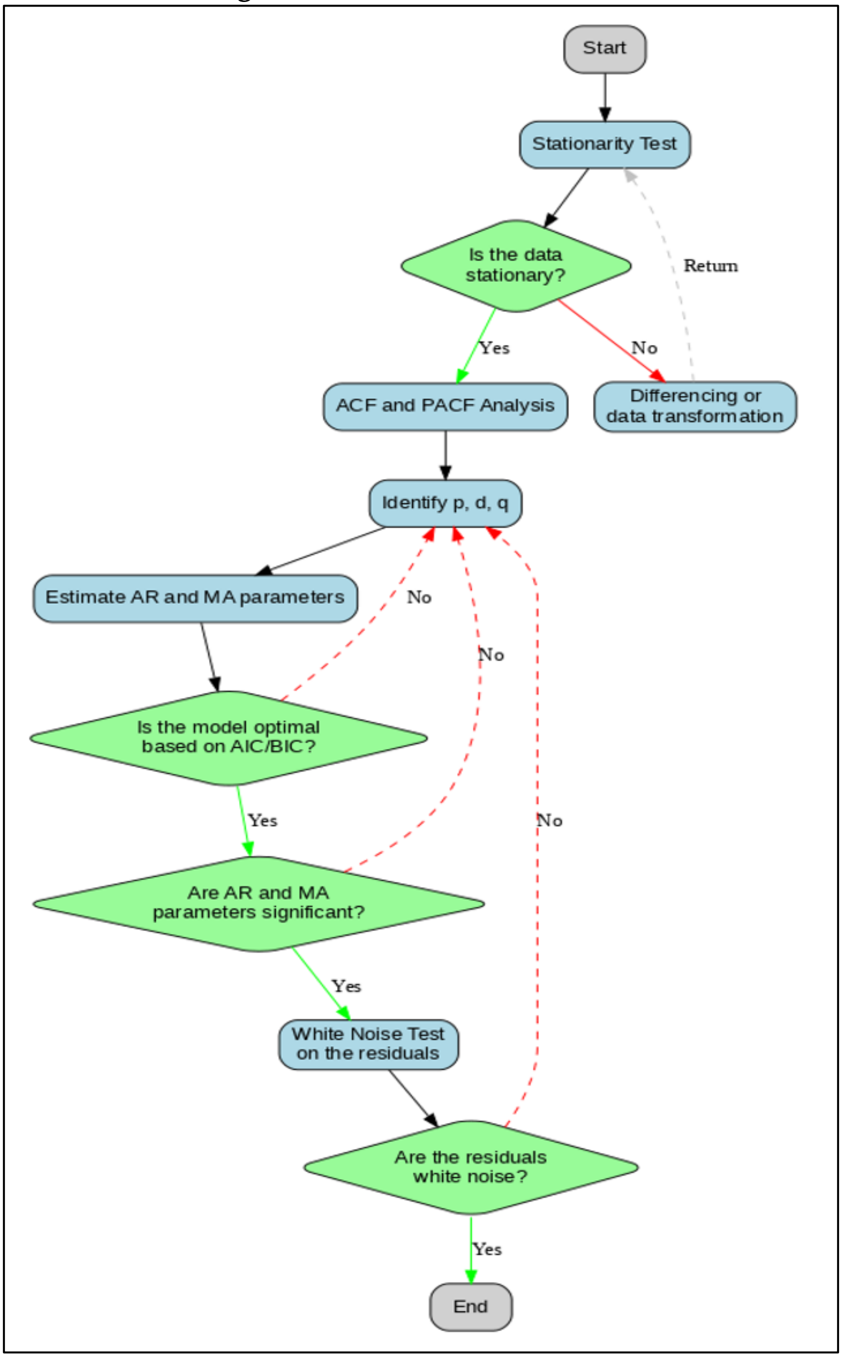
3.2.4 Forecast

The model, having passed diagnostic testing, is used to predict short-term economic growth.

3.2.5 ARIMA Modelling Scheme

The following diagram illustrates the steps of ARIMA modelling, including stationarity testing, parameter identification and residue validation:

Figure 1: ARIMA Model Process



Note: Python source developed by the authors.

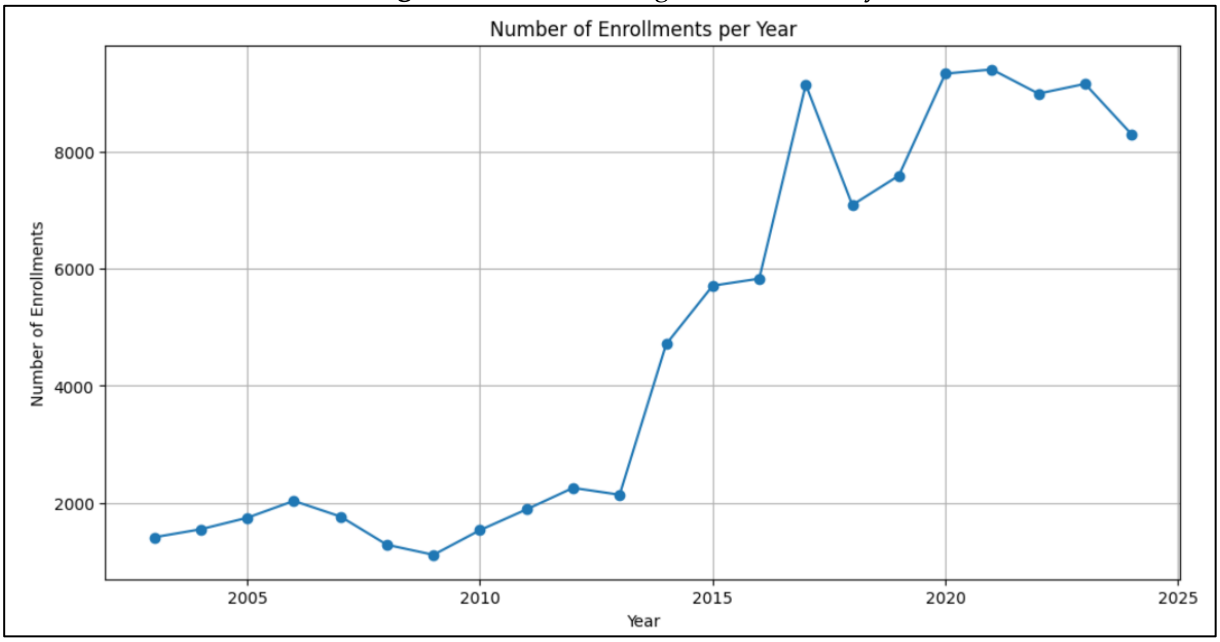
4. Analysis Results

4.1 Data Presentation

The data in this study reflect the actual number of new students enrolled annually at the institution over 22 years, from the academic year 2003-2004 to 2024-2025.

The chart below shows the trend in annual enrolment. There is an overall upward trend, with a sharp increase from 2010 and sustained growth until 2020. This period is followed by stabilisation or slight variations around a high level. Before 2010, the series remained relatively stable with a modest number of registrations, around 2000, before experiencing a strong upward trend.

Figure 1: New data registered annually

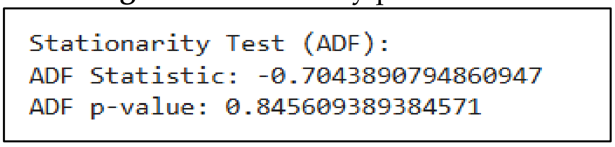


Note: Python source developed by the authors.

4.2 Augmented Dickey-Fuller (ADF) Stationarity Test

According to a preliminary study, the initial series is not stationary. The following figure shows parameters in this direction.

Figure 3: Stationarity parameters


```
Stationarity Test (ADF):
ADF Statistic: -0.7043890794860947
ADF p-value: 0.845609389384571
```

Note: Python source developed by the authors.

4.2.1 Interpretation

ADF test assumptions:

- H_0 (null hypothesis): The time series has a unit root, so it is not stationary.
- H_1 (alternative hypothesis): The time series is stationary.

4.2.2 Significance threshold

If the p-value is below the typical threshold of 0.05, H_0 is rejected in favor of H_1.

4.2.3 Result

In this case study, the p-value of 0.846 is much higher than 0.05. That means you cannot reject H_0. In other words, the time series is not stationary.

Following the first differentiation with ADF, we find the following results:

Figure 2: Parameters after first differentiation

```
=== Stationarity Test (ADF) ===
ADF Statistic: -5.4200813439813285
p-value: 3.07331142402529e-06
```

Note: Python source developed by the authors.

After first-order differentiation, the p-value is very low, indicating that the differentiated series is stationary. In other words, taking the difference between one year and the previous one, we get a trend-free process, which is the prerequisite for using an ARIMA model.

4.3 Stationarity test (KPSS)

The KPSS test, which starts from a zero assumption of stationarity gives a statistic of KPSS equal to 0.1286 and a p-value of 0.1 after differentiation. The statistic obtained is much lower than the critical values (at 10%, the threshold is 0.347).

Figure 3: Stationarity test parameters KPSS

```
=== Stationarity Test (KPSS) ===
KPSS Statistic: 0.12864104830611905
p-value: 0.1
Critical Values: {'10%': 0.347, '5%': 0.463, '2.5%': 0.574, '1%': 0.739}
The differenced series is stationary (according to KPSS).
```

Note: Python source developed by the authors.

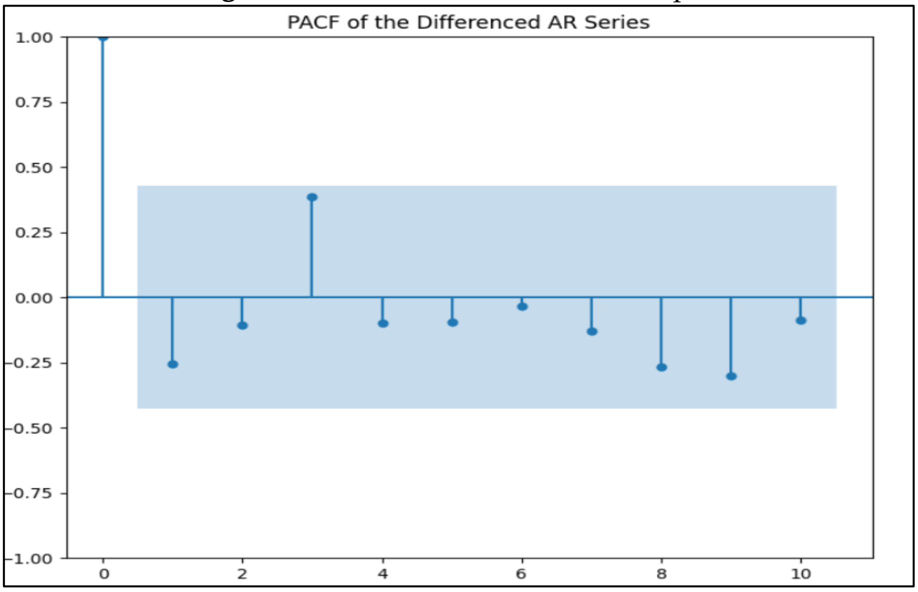
According to these observations, we do not reject the stationarity hypothesis. This test is used in addition to the ADF test, thus confirming that the differentiated series is beautiful and well-stationary.

After a differentiation of order 1, the ADF and KPSS tests show that the series no longer shows a systematic trend and can be considered stationary. This meets the requirements for adjusting an ARIMA model that requires data to be stationary to provide reliable forecasts.

4.4 Outline of ACF and PACF

After differentiation, we find: On the PACF as shown in Figure 6 below, a very important peak at lag 1, and the other lags don't seem to come out significantly from the confidence zone (apart perhaps from a lag 4, but much less marked).

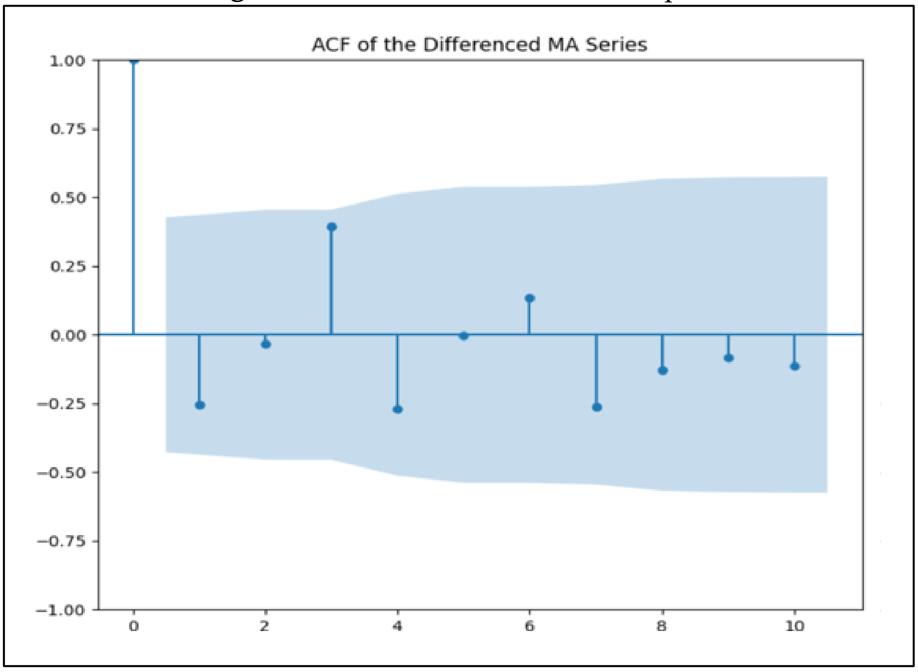
Figure 6: Differentiated series PACF plot



Note: Python source developed by the authors.

Taking into consideration Figure 7 of the ACF, which indicates a first significant positive lag, the other lags are not clearly significant. We can deduce that there is a slight peak at lag 4, but not necessarily above the confidence zone if you are strict.

Figure 4: Differentiated series ACF plot



Note: Python source developed by the authors.

The fact of having a strong peak on the PACF at lag 1, combined with the absence of a clean cut on the ACF, is a typical characteristic of an AR(1) model applied to an already differentiated series. In other words, the differentiated series seems to be modeled by an AR(1). Since the original series was differentiated once to achieve stationarity, this corresponds to an ARIMA(1,1,0) model applied to the initial series.

4.5 Validation of the Models

In order to validate our ARIMA(1,1,0) model, we use the Information Criteria (AIC), (BIC), and Statistical Tests on Residues.

The Akaike Information Criterion (AIC) (Gerda Claeskens, 2008), the Bayesian Information Criterion (BIC) (Arindrajit Pal, 2015) are essential tools for selecting ARIMA models, especially when there are many candidates. These criteria balance the complexity and fit of the model, ensuring that the chosen model is both lean and efficient in capturing the underlying data models.

Figure 5: Four parameters for ARIMA model validation

```
=== Searching for the Best ARIMA Model with Multiple Criteria ===
Trying to fit ARIMA(0, 1, 0) with trend...
AIC=358.3479448419022, AICc=358.9795237892706, BIC=360.43698971734904, Log-Likelihood=-177.1739724209511
Trying to fit ARIMA(0, 1, 1) with trend...
AIC=359.0668191899503, AICc=360.4001525232836, BIC=362.20038650312057, Log-Likelihood=-176.53340959497515
Trying to fit ARIMA(1, 1, 0) with trend...
AIC=359.0581027894244, AICc=360.39143612275774, BIC=362.1916701025947, Log-Likelihood=-176.5290513947122
Trying to fit ARIMA(1, 1, 1) with trend...
AIC=360.9754670615385, AICc=363.3284082380091, BIC=365.1535568124322, Log-Likelihood=-176.48773353076925

=== Best ARIMA Model (0, 1, 0) ===
AIC=358.3479448419022, AICc=358.9795237892706, BIC=360.43698971734904, Log-Likelihood=-177.1739724209511
```

Note: Python source developed by the authors.

4.5.1 AIC (Akaike Information Criterion)

From the above figure, it is noted that the lowest AIC value (358.35) corresponds to the ARIMA model (0,1,0). So this model offers a good balance between the complexity of the model and the quality of fit.

4.5.2 AICc (Corrected AIC)

From Figure 8, it is noted that the lowest value of the AICc criterion (358.98) corresponds to the same ARIMA model (0,1,0) mentioned to reinforce the choice of model.

4.5.3 BIC (Bayesian Information Criterion)

From the figure above, it is noted that the value of the lowest BIC criterion (360.44) corresponds to the same model ARIMA (0,1,0) mentioned to reinforce the choice of model.

4.5.4 Log-Likelihood

From the above figure, it is noted that the lowest value of the Log-Likelihood criterion (-177.17) corresponds to the same ARIMA model (0,1,0) mentioned to reinforce the choice of model.

Based on these four criteria, the best model is ARIMA (0,1,0), which is demonstrated in the last two lines of Figure 8.

The mathematical formula of our module is as follows:

$$Y_t = Y_{t-1} + \varepsilon_t$$

- Y_t = The value of the time series at time t.
- Y_{t-1} = The value of the time series at time t- 1.
- ε_t = The error term at time t, which generally follows a normal distribution with an average of zero and a constant variance (white noise).

4.5.5 Interpretation of the Drift

The drift parameter, as shown in the following figure, is equal to 23.7487. It shows a linear trend with an average increase of 23.75 per period. It means a monotonic evolution of the time series.

Figure 6: Drift parameters and residue variance

	coef	std err	z	P>|z|	[0.025	0.975]
drift	23.7487	17.875	1.329	0.184	-11.287	58.784
sigma2	1.246e+06	3.06e+05	4.077	0.000	6.47e+05	1.85e+06

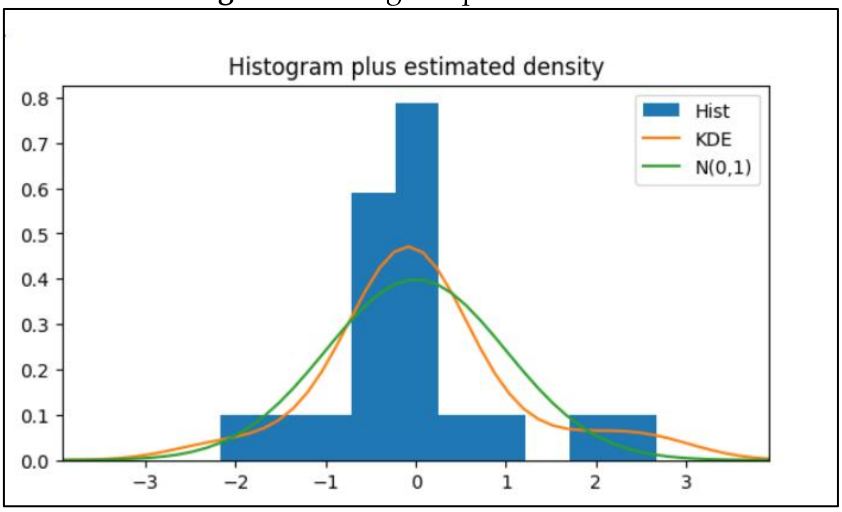
Note: Python source developed by the authors.

However, the p-value associated with drift is 0.184, greater than 0.05, which means that the trend is not significantly different from zero.

4.5.6 Residue histogram and density estimation (KDE)

The analysis of the residue histogram in conjunction with the density estimation (KDE), is essential to evaluate the distribution of the deviations between the observed values and the values predicted by the ARIMA model.

Figure 7: Histogram plus estimated



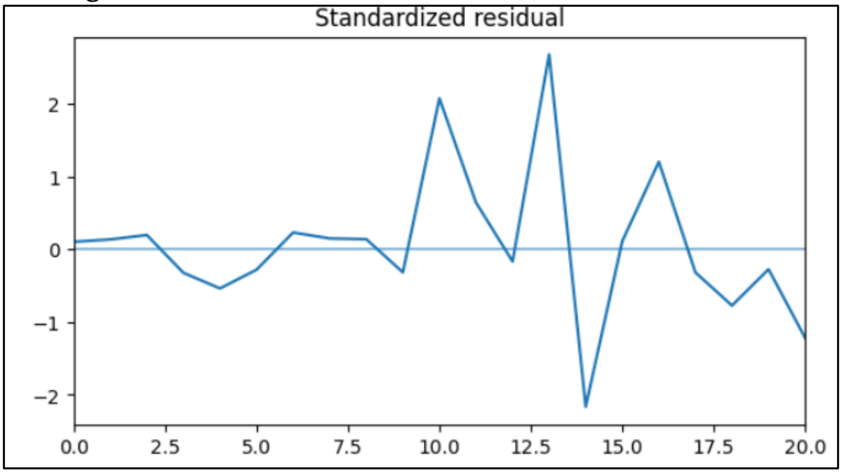
Note: Python source developed by the authors.

The graph compares the histogram of the residuals, which are the deviations between observed values and the values predicted by the model at an estimated density (Kernel Density Estimation: KDE). It has a standard normal distribution N(0,1). Visually, the residuals appear to be generally centered around 0, which is a good sign.

However, the shape of the distribution of residues does not perfectly correspond to the curve of the normal law. The residues tend to cluster closer to the mean (central value), which is normally observed in a standard normal distribution.

The graph of the residuals over time below shows fluctuations around 0. There is no marked trend or visible cyclical pattern that would remain unmodelled. The variations appear irregular, which is usually the objective sought: Residues close to a white noise.

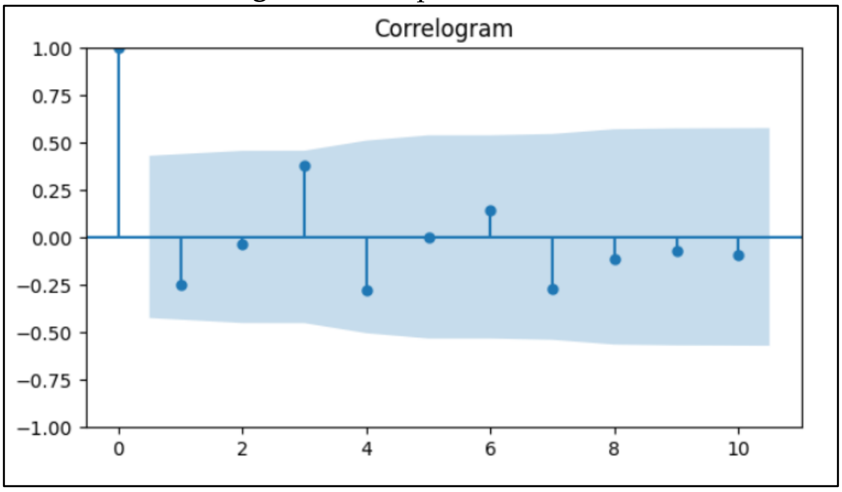
Figure 8: Residues of the best ARIMA model with trend



Note: Python source developed by the authors.

The ACF (Autocorrelation Function) graph or correlogram shown below shows that there is no clear pattern of autocorrelation in the residues. The points represented are generally within the confidence zone (in blue), suggesting that no specific lag has significant autocorrelation. This is a good indicator that there is no remaining sequential information not explained by the model.

Figure 9: ACF plot of residues



Note: Python source developed by the authors.

4.5.7 Ljung-Box test

The Ljung-Box test is a statistical test used to check for the absence of autocorrelation in the residuals of a time series model. In other words, the test is used to check whether the model's residuals behave like white noise.

Figure 10: Ljung-Box test

```
=== Ljung-Box Test ===
      lb_stat  lb_pvalue
10  11.537785   0.317181
The residuals do not show significant autocorrelation.
```

Note: Python source developed by the authors.

The figure above shows that the p-value (0.317) is greater than 0.05. This confirms the absence of significant autocorrelation of residues. The ARIMA(0,1,0) model has therefore captured the dependency in the series, leaving a residue that behaves like a white noise.

4.5.8 Variance of residues

From Figure 9, the variance of the residues is relatively high, which may suggest that the model does not fully explain the data. However, the associated p-value is close to 0.45, indicating that this variance is significant.

Although the variance is high, a significant p-value (close to zero) suggests that the residuals follow a normal distribution and can be considered as white noise, which is essential for the validity of the ARIMA model.

The following figure shows two key parameters for this study.

Figure 11: Performance metrics results

```
MAPE : 20.46%
MASE : 1.0012
```

Note: Python source developed by the authors.

The MAPE (Mean Absolute Percentage Error) parameter is equal to 20.46. This means that, on average, the forecasts of the ARIMA model deviate by 20.46% from the actual values. A MAPE below 10% is generally considered excellent, between 10% and 20% acceptable, and above 20% is relatively imprecise. In our case, 20.46% indicated that the model is acceptable but can be improved.

The resulting MASE (Mean Absolute Scaled Error) metric is 1.0012. If a MASE value is close to 1, the ARIMA model is about as good as the naive prediction method (which predicts that each observation equals the previous one). If the MASE is equal to 1, this means that the ARIMA model is less accurate than the naive method.

4.5.9 The summary

Modeling annual registrations with only 22 observations presents challenges, particularly due to the small sample size, the initial non-stationarity of the time series and the need to select a model capable of capturing underlying dynamics while maintaining robust interpretability.

The analysis revealed that the ARIMA(0,1,0) model is the most suitable for these data, according to the information criteria AIC, BIC and AICc. This model, although simple, effectively captures the trend by first-order differentiation and inclusion of a drift. However, several points need to be discussed.

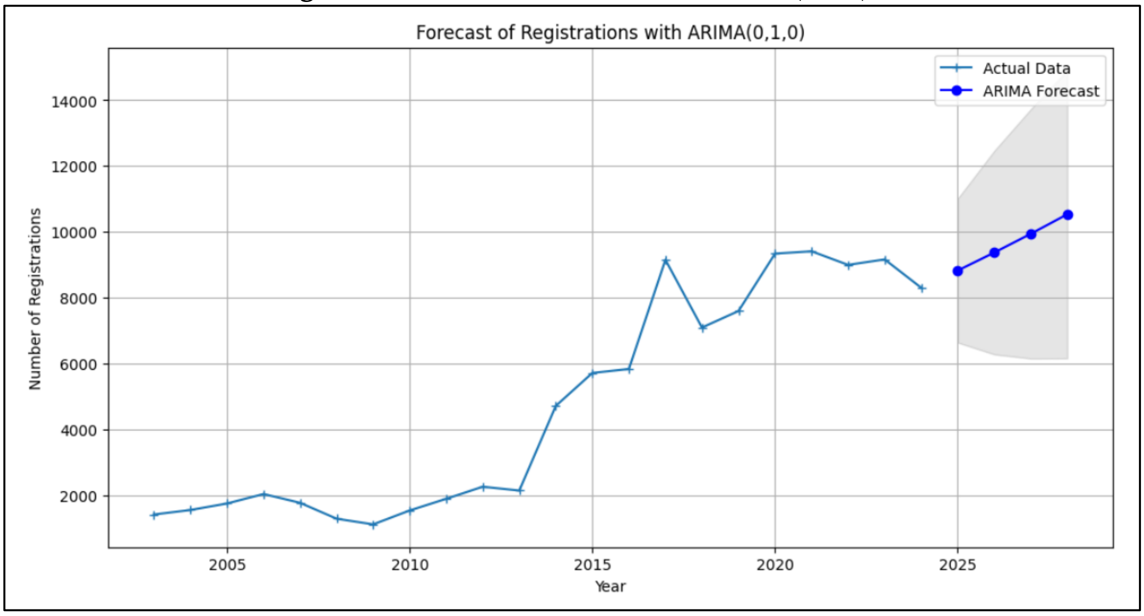
The MAPE of 20.46% indicates acceptable but not ideal forecasts, reflecting limitations in capturing complex variations. The residue tests (Ljung-Box, Jarque-Bera, ACF) show that the model detects the dependency structure of the series well and that the residue behaves like white noise, thus validating the model.

However, the p-value of drift, while significant, shows uncertainty about its exact role, which could limit confidence in long-term forecasts.

Given the limited amount of data available (22 observations), it's difficult to use machine learning algorithms (Goodfellow, 2016). According to (Pereira, 2006), in our study, the implementation of complex models like GARCH will not be beneficial when there is a small sample; the results may be biased. In this case, volatility estimates may not be accurate because the model lacks sufficient data to reliably detect changes in volatility over time. In addition, the lack of more detailed data, such as monthly or by sector information, limits the identification of seasonal trends or specific behaviors. Despite these limitations, the ARIMA(0,1,0) model remains relevant and allows reliable short- and medium-term forecasts.

The following figure shows the data forecast according to the ARIMA model (0.1.0) under study.

Figure 12: Forecast with ARIMA model (0,1,0)



Note: Python source developed by the authors.

The short-term forecasts are shown in the following figure. For the academic year 2025-2026, the number of new students enrolled in the Bachelor's degree will be 10,533. In 2026-2027, the number will increase by 600 new students. Furthermore, in 2027-2028, the institution will be asked to welcome 11792 new students enrolled in the degree of fundamental licensed studies.

Figure 13: Short-term forecast of new enrolments per year

25	10533.379342
26	11150.845968
27	11792.061311

Note: Python source developed by the authors.

5. Conclusion and Perspective

The results of the forecasts obtained in the next few years allow us to have a vision of what will happen in the near future. These results may have a positive influence on decision-making. The decision-makers at this institution will be able to develop short- and medium-term projections.

At the level of university governance, the establishment council, as the decision-making body will be able to question the reception capacity in the rooms and lecture halls for studies. The number of forecasts continues to increase, which will enable the group management of students. Indeed, the addition of groups will influence the volume of teaching timetables, and consequently, the recruitment of new teachers will be at stake. Informed decision-making in this direction will ensure, from year to year, an optimal educational supervision rate, thus ensuring the quality of teaching and student monitoring. The council is led to monitor this growth in the number of students through an evaluation of administrative management, which raises the question of the sufficiency of administrative staff to effectively carry out daily tasks.

Financial governance is also involved in this study. It is the main actor for such decisions. Following this study, a sufficient budget allocation for the next academic year is needed. The draft finance laws of the next few years are designed to take into account these forecasts for the recruitment of human resources relating to administrative and teaching staff. The construction of new premises for administration and teachers, as well as classrooms, is essential. In some cases, the board of education may propose to the government the possibility of building a new space if the capacity is limited and the institution will not be able to have extensions.

This study provides a forecast of the overall number of new students enrolled annually at the institution we are researching.

This study may also consider the gender approach to new students, the degrees offered by the institution, student choices, annual graduates, dropouts, as well as other factors related to the infrastructure of the institution.

These studies will be the objective of future work to provide decision-makers with statistics that can contribute positively to good university governance in Morocco.

Conflict of Interest Statement

The authors declare no conflicts of interest.

About the Authors

Adil Ghafir is a PhD student in university governance with a strong interest in artificial intelligence and decision-making. His research focuses on optimizing academic decision-making processes and designing intelligent assistants. He also has knowledge in developing AI agents and multi-agent systems, aiming to enhance efficiency and automation in academic and institutional management.

Anas Bennani is a professor and researcher at Abdelmalek Essaâdi University, Morocco specializing in data analysis, decision support systems, and public sector digital policies. Passionate about AI, NTIC, and their legal frameworks, he researches semantic web technologies and their applications. He is committed to advancing digital transformation and higher education in Morocco.

Abdelhak Tali is a professor and researcher at Abdelmalek Essaâdi University, Morocco. He specializes in artificial intelligence, data science, and blockchain technologies. He coordinates the Master’s program in Data Science for the Digital Economy and Finance. His research focuses on machine learning, biometric identification, and intelligent systems, with expertise in deep learning and big data analytics.

Soukaina El Hajjaji is a professor and researcher at the Faculty of Legal, Economic, and Social Sciences in Tangier, Morocco. She specializes in project and quality management, governance, and leadership, with a focus on university governance, decision-making processes, and institutional performance. Her work aims to improve strategic planning, organizational efficiency, and higher education policies.

References

A. Asrirawan, S. U. (2022). Univariate Time Series Modeling Approach for Quarterly Prediction of Indonesia's Economic Growth Post-COVID-19 Vaccination. *Jambura Journal of Mathematics*, 86–103. doi:10.34312/jjom.v4i1.11717

A. Dela Cruz, M. B. (2020). Higher Education Institution (HEI) enrollment forecasting using data. *International Journal of Advanced Trends in Computer Science and Engineering*, 2060-2064. Retrieved from https://doi.org/10.30534/ijatcse/2020/179922020

Ahmar, A. S. (2020). *Calcul d'erreur de prévision avec erreur quadratique moyenne (MSE) et erreur absolue moyenne en pourcentage (MAPE).* Indonesia: Yayasan Ahmar Cendekia Indonesia. Retrieved from https://doi.org/10.35877/454RI.jinav303

Arindrajit Pal, J. P. (2015, 03). Path length prediction in MANET under AODV routing: Comparative analysis of ARIMA and MLP model. *Egyptian Informatics Journal, 16*(1), 103-111. Retrieved from https://doi.org/10.1016/j.eij.2015.01.001

D. Wulan Sari, R. G. (2016). Parameter estimation of an ARIMA model for river flow forecasting using least squares and goal programming. *Jurnal EKSPONENSIAL*. Retrieved from http://jurnal.fmipa.unmul.ac.id/index.php/exponensial/article/view/57

George E. P. Box, G. M. (1976). *Time Series Analysis: Forecasting and Control.* Revised Edition, San Francisco: Holden Day. Retrieved from https://doi.org/10.1111/jtsa.12194

Gerda Claeskens, N. L. (2008). *Model Selection and Model Averaging.* Cambridge: Press, Cambridge University. Retrieved from https://doi.org/10.1017/CBO9780511790485.003

Gilbert M. Masinading, S. E. (2024, October 24). Forecasting the semestral enrollment of DOrSU curricular programs. *Advances and Applications in Statistics*, 1579 -1592. Retrieved from https://doi.org/10.17654/0972361724080

Goodfellow, I. B. (2016). *Deep Learning.* Cambridge, MA: MIT Press.

Ilan, A. (2005). Forecasting University Enrollment: An Historical Case of a College of Business in Northeast United States of America. *Journal of College Teaching & Learning, 2*(4). doi:10.19030/tlc.v2i4.1803

Lan Luo. (2024, 09 20). Statistical model validation through white noise hypothesis testing in regression analysis and ARIMA models. *Theoretical and Natural Science, 42*, 99-104. Retrieved from https://doi.org/10.54254/2753-8818/42/20240672

Neusser, K. (2016). *Time series econometrics.* Springer. Retrieved from https://link.springer.com/book/10.1007/978-3-319-32862-1

Pereira, S. H. (2006). Propriétés des estimations GARCH sur petits échantillons et persistance. (Routledge, Ed.) *European Journal of Finance*, 473-494. Retrieved from https://doi.org/10.1080/13518470500039436

Søren Johansen, M. R. (2012, 11 8). The Selection of ARIMA Models with or Without Regressors. (U. o. Discussion, Ed.) *Social Science Research Network,* 12-17. Retrieved from http://dx.doi.org/10.2139/ssrn.2175553

Ward, J. (2007). Forecasting enrollment to achieve institutional goals. *College University Journals*, 41-46.