Modelling and forecasting the number of students enrolled in the College of Administrative Sciences at Kuwait University using Multiplicative SARIMA Model

By: Narges Akbar
Department of information Systems and operations management
College of Business Administration
Kuwait University, Kuwait.
Narges.Akbar@ku.edu.kw

Abstract

This study aims to produce the most adequate time series model for modelling and forecasting the number of students enrolled in the College of Administrative Sciences at Kuwait University using "Box and Jenkins methodology”. As a case study, we collected the data from admission and registration department from the academic year 1995/1996 to the academic year 2020/2021 for all semesters (Fall, Spring and Summer). The SARIMA (0, 1, 2) (0,1,3) model, which successfully passed all the diagnostic tests and checks, has been used in forecasting the next **two** academic years.

keyword

Forecasting, Stationarity, invertibility, autocorrelation function (ACF), partial autocorrelation function (PACF), multiplicative SARIMA model.

1. Introduction

The main aims of the current research are to obtain the most adequate model among the multiplicative SARIMA (p, d, q) (P, D, Q) models, and study the reality of the quantitative change in the series of students enrolled in the College of Administrative Sciences at Kuwait University from 1995/1996 to 2020/2021. In addition, the research aims to forecast the values for the number of students enrolled in the College of Administrative Sciences at Kuwait University using "Box and Jenkins methodology" for the coming quarters. For more about the methodology of Box and Jenkins the reader is referred to Abraham, B. and Ledolter (2005), Box, G.E.P., Jenkins, G.M. (1970), Box et. al (2016), Bowerman, B. L. and O'Connell, R. T. (1993), Chatfield, C. (2019), Harvey, A.C. (1993), Liu, L. M. (2009), Shaarawy, S. M (2005) and Shaarawy et. al (2014). It should be noted that all tables, graphs, estimates, and forecasts were made using MINITAB Version 21.

The data on the numbers of students enrolled in the College of Administrative Sciences at Kuwait University includes 78 observations from the academic year 1995/1996 to the academic year 2020/2021, as shown in Table (1).

2. Series Preliminary Inspection

The first step in the analysis of time series is to plot the historical curve of the series, which shows the pattern in which the number of students develops during the period under study, to identify the basic features of the data under study such as general trend, dispersion, stationarity, autocorrelation, and outlier values. Figure (1) presents the time curve of the student series, which shows a general trend of increase over the period under study, which means that the series is not stationary in the arithmetic mean. The increase in the arithmetic average reflects the effects of some main factors, such as the transfer of the Faculty of Commerce, Economics and political science to the Faculty of Administrative Sciences in 2005, and other factors such as the increase in the size of society and the rise in its standard of living, which are logical changes due to population growth and development in all aspects of life and an increase in awareness of the labor market's need for specializations. These factors work together or separately to increase the level of the series. It is also clear from the careful examination of Figure (1) that expressing the general trend using one of the well- known mathematical functions is not appropriate due to the existence of a clear positive autocorrelation between the observations of the series. The evidence for existence of such positive autocorrelation is that if we visualize a straight line or a curve from the second degree mediating the data and one of the observations is located above the line or the curve, the following observation tends to locate above the line and vice versa, which loses the least squares estimates of their ideal properties. Thus, one has to use a stochastic process such as multiplicative SARIMA process to model and forecast the series being studied.

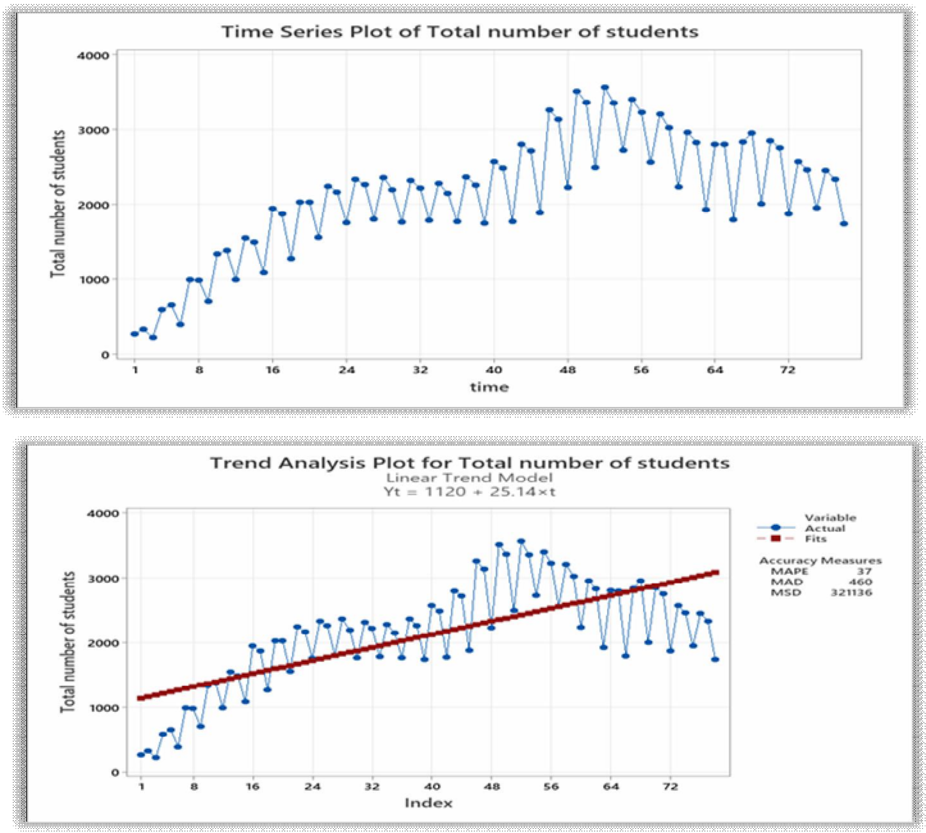
Table (1): Number of students enrolled in the College of Administrative Sciences at Kuwait University from 1995/1996 to 2020/2021

Year Semester	1995-1996	1996-1997	1997-1998	1998-1999	1999-2000	2000-2001	2001-2002	2002-2003	2003-2004
Fall	269	591	994	1341	1546	1944	2024	2242	2331
Spring	327	659	986	1384	1492	1870	2025	2164	2264
Summer	221	392	706	995	1091	1275	1560	1756	1803

Year Semester	2013-2014	2014-2015	2016-2017	2017-2018	2018-2019	2019-2020	2020-2021
Fall	3399	3203	2957	2804	2837	2851	2569
Spring	3224	3019	2829	2801	2954	2756	2457
Summer	2556	2233	1924	1790	2003	1869	1946

Year Semester	2004-2005	2005-2006	2006-2007	2007-2008	2008-2009	2009-2010	2010-2011	2011-2012	2012-2013
Fall	2360	2316	2277	2365	2570	2800	3254	3507	3561
Spring	2193	2218	2148	2258	2485	2719	3132	3361	3352
Summer	1765	1784	1767	1743	1772	1883	2221	2492	2728

Figure (1): time series plot for y(t) number of students enrolled in the College of administrative Sciences at Kuwait University from 1995/1996 to 2020/2021

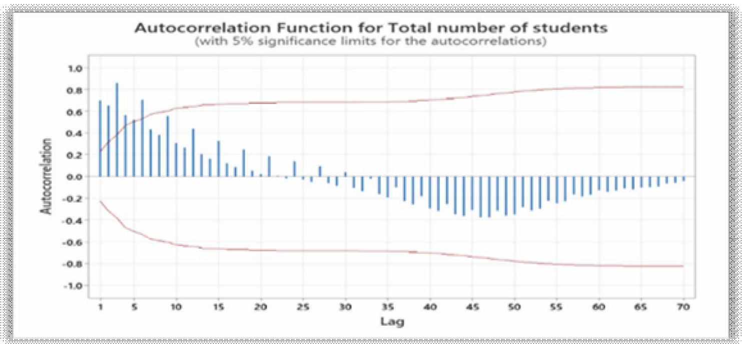


With more careful examination of Figure (1), it can be observed that the scattering data does not change around the level of the series, and thus the series seems to be stationary in the variance. In addition, it appears from the same figure that the series does not contain unusual observations. It can be said that the initial examination of the series showed no stationarity in the mean of the series. To verify this, the estimated autocorrelation function of the series was calculated and

Table (2): Autocorrelation function of the original series y(t)

Lag	ACF	T	LBQ	Lag	ACF	T	LBQ
1	0.699217	6.18	39.62	36	-0.101811	-0.29	370.95
2	0.655154	4.11	74.86	37	-0.227624	-0.66	378.83
3	0.860775	4.51	136.51	38	-0.257574	-0.74	389.18
4	0.566611	2.41	163.58	39	-0.182878	-0.52	394.53
5	0.519434	2.06	186.64	40	-0.292391	-0.83	408.57
6	0.706061	2.66	229.85	41	-0.315770	-0.89	425.39
7	0.429689	1.49	246.08	42	-0.253481	-0.71	436.53
8	0.381867	1.29	259.07	43	-0.346080	-0.96	457.88
9	0.557382	1.84	287.17	44	-0.362347	-0.99	481.98
10	0.305720	0.97	295.75	45	-0.307368	-0.83	499.84
11	0.265322	0.83	302.30	46	-0.374070	-1.00	527.13
12	0.437014	1.36	320.36	47	-0.374038	-0.99	555.29
13	0.206105	0.62	324.44	48	-0.313836	-0.82	575.77
14	0.162938	0.49	327.03	49	-0.358812	-0.93	603.48
15	0.326541	0.98	337.59	50	-0.347644	-0.89	630.41
16	0.121470	0.36	339.07	51	-0.281944	-0.71	648.78
17	0.087158	0.26	339.85	52	-0.311724	-0.79	672.10
18	0.248392	0.74	346.27	53	-0.295233	-0.74	693.86
19	0.053603	0.16	346.57	54	-0.224361	-0.56	706.95
20	0.022714	0.07	346.63	55	-0.247106	-0.61	723.51
21	0.186423	0.55	350.43	56	-0.229299	-0.56	738.43
22	0.005938	0.02	350.43	57	-0.168567	-0.41	746.87
23	-0.020968	-0.06	350.48	58	-0.185360	-0.45	757.59
24	0.138617	0.41	352.70	59	-0.169877	-0.41	767.07
25	-0.031012	-0.09	352.82	60	-0.128824	-0.31	772.82
26	-0.053585	-0.16	353.16	61	-0.143580	-0.35	780.39
27	0.094774	0.28	354.26	62	-0.132282	-0.32	787.21
28	-0.064284	-0.19	354.78	63	-0.111937	-0.27	792.42
29	-0.089270	-0.26	355.79	64	-0.119027	-0.29	798.74
30	0.040295	0.12	356.00	65	-0.103548	-0.25	803.89
31	-0.109564	-0.32	357.60	66	-0.098670	-0.24	808.95
32	-0.135676	-0.40	360.09	67	-0.094565	-0.23	814.02
33	-0.024101	-0.07	360.17	68	-0.068623	-0.17	816.96
34	-0.163214	-0.47	363.95	69	-0.061103	-0.15	819.55
35	-0.193883	-0.56	369.41	70	-0.046007	-0.11	821.20

Figure (2): Autocorrelation function of the original series y(t)



From figure (2), it is noticed that the estimated autocorrelation function dies down slowly to zero. This indicates that the original series of the number of students, which will be denoted by the symbol y(t) is not stationary. Therefore, we had to transform the original series to stationary one by taking the suitable differences.

Regular trend (non-seasonal pattern):

It is observed from Figure (2) that the behavior of the autocorrelation coefficients for the non-seasonal pattern die down slowly, and then one must take the first difference for the non-seasonal pattern, i.e., we take d=1.

Seasonal trend (seasonal pattern):

Regarding the behavior of the seasonal pattern, we notice from Figure (2) that the value of the autocorrelation coefficient for the time unit 3 is $r_3 = 0.86$ and the value of the corresponding T statistic equals 4.51. Then the coefficients decrease at the time units 4 and 5, then they started to increase again at the time unit 6. This pattern continues at the seasonal lags 9, 12... Etc. with observing that the coefficients at these lags die down slowly. This proposes to take the seasonal difference D=1.

Thus, we consider the following transformation:

$$z(t) = \Delta \, \Delta_3 \; y(t) = (1 - \mathrm{B})\,(1 - \mathrm{B}^3)\; y(t)$$

Figure (3), Figure (4) and Figure (5) show the time series plot, autocorrelation function and partial autocorrelation function of the new series z_t respectively.

Figure (3): time series plot for Z(t)

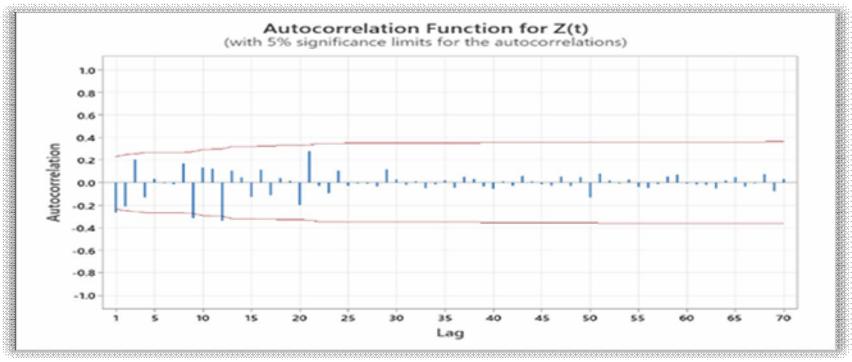


Figure (4): Autocorrelation function for z(t)

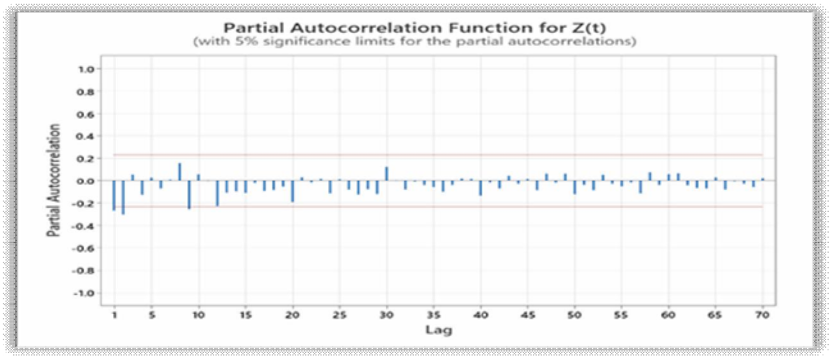
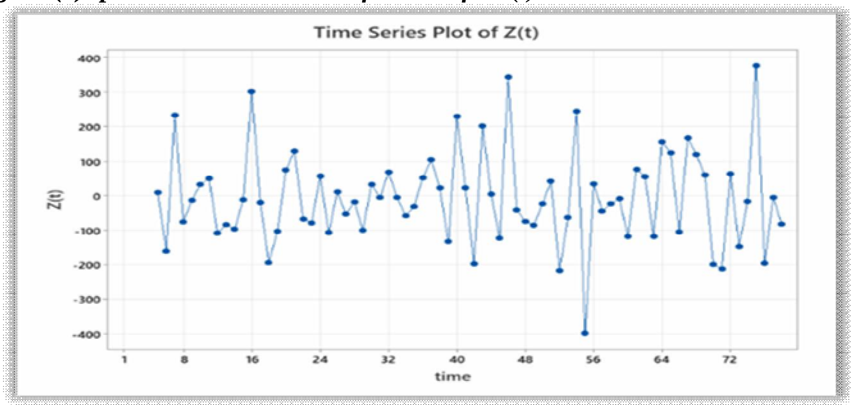


***Figure (5): partial Autocorrelation function for** z(t)*



3. Model Tentative identification

It is usual in time series analysis to have more than one model that is initially suitable for analyzing the data under study in the identification stage because the main objective of this stage is to narrow the range of models that can be selected for further study. Our data is one of the series that can have more than one suitable model. Examining the behavior of the autocorrelation and partial autocorrelation functions of the transformed series z(t), I found myself in front of the following four different models that can tentatively model the data: SARIMA (2,1, 0) (4,1,0), SARIMA (2,1, 0) (2,1,2), SARIMA (0,1,2) (2,1,2), and SARIMA (0,1,2) (0,1,4). These models deserved more study and diagnostic checking tests. Each of the first two models has been used to fit the data and all the diagnostic checking tests have been done, and it has been found that all tests satisfy all conditions and assumptions about the error term except the normality assumption. This means that none of these two models can be used to make statistical inference about the coefficients and forecasts, and they be removed from the comparison.

Then the fourth model SARIMA (0, 1, 2)(0, 1, 4) has been used to fit the data and it has been found that the fourth seasonal coefficient Θ_4 does not significantly different from zero and was deleted, then we ended up with the model SARIMA (0, 1, 2)(0, 1, 3) as alternative model. This model and the third model SARIMA (0,1,2) (2,1,2) have been used to fit the data and it has been found the each of them satisfy all assumptions and conditions about the error term. Therefore, we had to propose another criterion to choose one of them. To do that, we deleted the last six values from the data and used each model to forecast them and we got the following results:

Table (3): Comparing the Point Forecasts

Period	Actual	Forecasts of SARIMA (0,1,2) (0,1,3)	Forecasts of SARIMA (0,1,2) (2,1,2)	error percentage % of SARIMA (0,1,3) (0,1,3)	error percentage % of SARIMA (0,1,2) (0,1,2)
73	2569	2751.03	2916.54	7.1	13.5
74	2457	2534.20	2799.03	3.1	13.9
75	1946	1775.01	1938.43	8.8	0.4
76	2450	2539.74	3048.52	3.7	24.4
77	2333	2227.78	2881.71	4.5	23.5
78	1739	1495.62	2036.20	14	17.1
Average				6.7	15.5

Table (4): Comparing the 95 Percent Confidence Interval of Forecasts

		SARIMA (0,1,2) (0,1,3)		SARIMA (0,1,2) (2,1,2)	
Period	**Actual**	**Lower bound**	**Upper bound**	**Lower bound**	**Upper bound**
73	2569	2545.44	2956.63	2701.84	3131.23
74	2457	2275.60	2792.80	2535.10	3062.97
75	1946	1499.95	2050.07	1662.46	2214.40
76	2450	2107.02	2972.45	2642.00	3455.04
77	2333	1713.97	2741.60	2411.44	3351.98
78	1739	943.07	2048.17	1538.67	2533.73

From the last two tables, we conclude the following:

1. The error percentage for each future observation computed by the model SARIMA (0,1,2) (2,1,2) is bigger than the corresponding error percentage, computed by the model SARIMA (0,1,2) (0,1,3), for all observations except one value.

2. The average of error percentages of the future observations computed for the model SARIMA (0, 1, 2) (0, 1, 3) is 6.7, while the corresponding average for the other model is 15.5.
3. The average of error percentages of the future observations computed for the model SARIMA (0,1,2) (2,1,2) is 131% more than the corresponding average computed by the SARIMA model (0,1,2) (0, 1,3).
4. All the confidence intervals of model SARIMA (0,1,2) (0,1,3) contain the actual observations, while only two confidence intervals of the model SARIMA model (0,1,2) (2,1,2) contain the actual observations.

From the above analysis and conclusions, one may be convinced to select the SARIMA (0, 1, 2) (0, 1, 3) model to be the most adequate one to model and forecast the data being analyzed.

4. Model Estimation

Table (5) gives the estimates of the model parameters and their standard errors of the identified model SARIMA (0, 1, 2) (0, 1, 3)

Table (5): Estimation of the SARIMA (0, 1, 2) (0, 1, 3) Model

Type	Coef	SE coef	T	P
MA 1	0.2619	0.1201	2.18	0.033
MA 2	0.2983	0.1206	2.47	0.016
SMA 3	-0.1902	0.1074	-1.77	0.081
SMA 6	-0.0648	0.1053	-0.62	0.541
SMA 9	0.7304	0.1063	6.87	0.000

Differencing: 1 regular, 1 seasonal of order 3
Number of observations: Original series 78, after differencing 74
Residuals: SS = 801000 (back forecasts excluded)
MS = 11609 DF = 69

5. Diagnostic Checking

Once the appropriate tentative model is identified and its parameters are estimated, the adequacy of the model's theoretical assumptions to the observed time series data must be carefully examined in order to improve, develop or keep the model as it is if the theoretical hypotheses are appropriate. This stage is one of the most important and critical stages of modern time series analysis, which always requires hard effort from the researcher to be assured of the suitability of the identified model and then the possibility of using it in future prediction. Four main tests were conducted to evaluate the model, including invertibility analysis, residual analysis, the possibility of removing some parameters (underfitting) from the model, and the possibility of adding some parameters (overfitting) to the model. The check for stationarity analysis is not relevant since the identified SARIMA (0, 1, 2) (0, 1, 3) model is always stationary regardless of the parameter values. For more details about the diagnostic checking, one may see Box et al. (2016). Below we present the results of these tests in some detail.

5.1 Analysis of invertibility:

The non-seasonal estimates should satisfy the following invertibility conditions:

1. $\theta_1 + \theta_2 < 1$
2. $\theta_2 - \theta_1 < 1$
3. $|\theta_2| < 1$

It is clear that

1. $\theta_1 + \theta_2 = 0.2619 + 0.2983 = 0.5602 < 1$
2. $\theta_2 - \theta_1 = 0.2983 - 0.2619 = 0.0369 < 1$
3. $|\theta_2| = 0.2983 > 1$

These mean that the three invertibility conditions of the non-seasonal estimates are satisfied.

In addition, the seasonal estimates should satisfy the following invertibility conditions:

1. $\Theta_1 + \Theta_2 + \Theta_3 < 1$
2. $-\Theta_1 + \Theta_2 - \Theta_3 < 1$
3. $|\Theta_3| < 1$
4. $|\Theta_2 + \Theta_1\Theta_3| < |1 - \Theta_3^2|$

It is clear that

1. $\Theta_1 + \Theta_2 + \Theta_3 = -0.1902 - 0.0648 + 0.7304 = 0.4754 < 1$
2. $-\Theta_1 + \Theta_2 - \Theta_3 = 0.1902 - 0.0648 - 0.7304 = -0.605 < 1$
3. $|\Theta_3| = 0.7304 < 1$
4. $|\Theta_2 + \Theta_1\Theta_3| = |-0.0648 - (0.1902)(0.7304)| = 0.2037 < |1 - \Theta_3^2| = 0.4665$

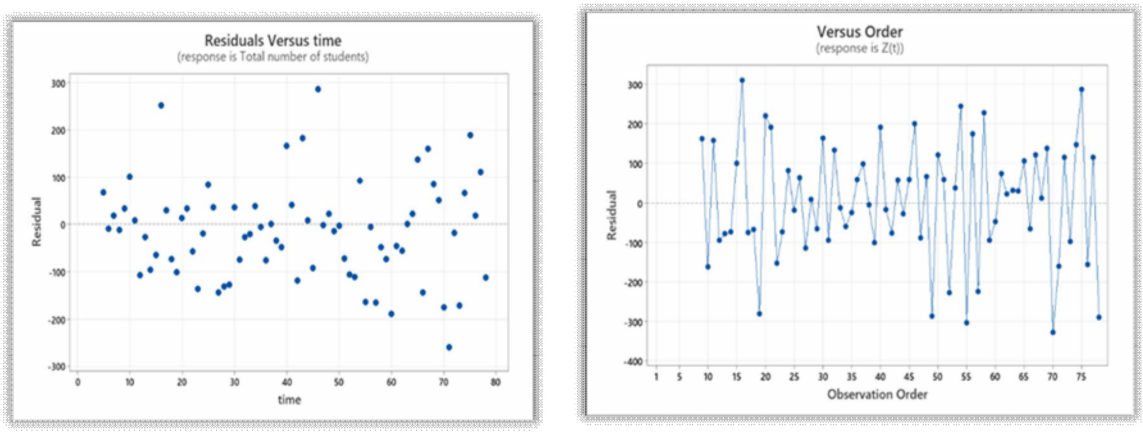
These mean that the four invertibility conditions of the seasonal estimates are satisfied. Hence, the identified model is invertible. One is referred to Shaarawy (2005) for the invertibility (conditions).

5.2 Residual analysis

a. Residual versus time:

Figure (6) appears to be devoid of all the regular patterns and moves that could be used to improve the model since the data randomly oscillating around the zero line

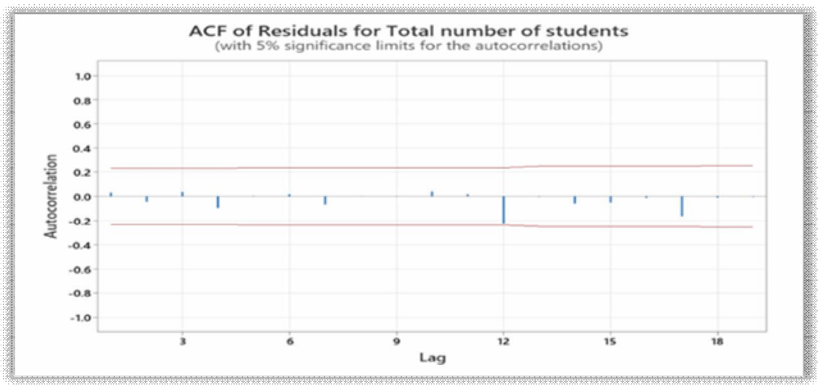
Figure (6): residuals versus time



b. Autocorrelation function (ACF):

Figure (7) gives the autocorrelation (ACF) function of the residuals. It is easy to see that each coefficient of the autocorrelations of residuals falls within the confidence interval for large samples, meaning that the shape of the autocorrelation function of residuals has no spikes and this is another good indication that the errors $\varepsilon(t)$ represent purely random errors.

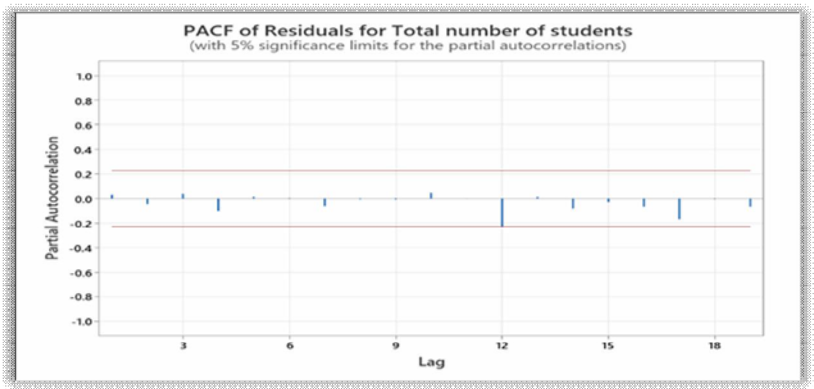
Figure (7): ACF of residuals for total number of students



c. Partial Autocorrelation function (PACF):

Figure (8) gives the partial autocorrelation (PACF) function of the residuals. It is clear that each coefficient falls within the confidence interval for large samples, meaning that the shape of the partial autocorrelation function of residuals has no spikes and this is another good indication that the errors $\varepsilon(t)$ represent purely random errors.

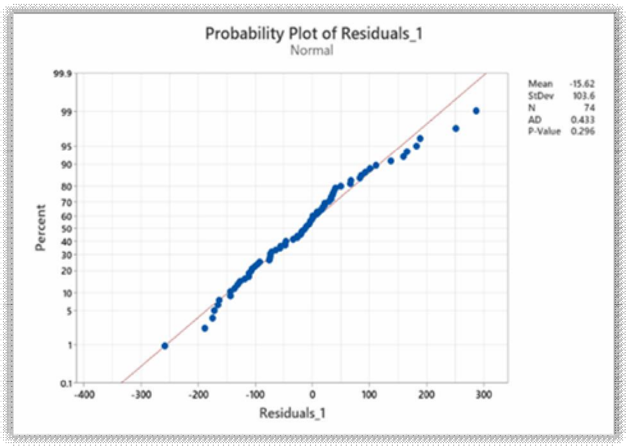
Figure (8): PACF of residuals for total number of students

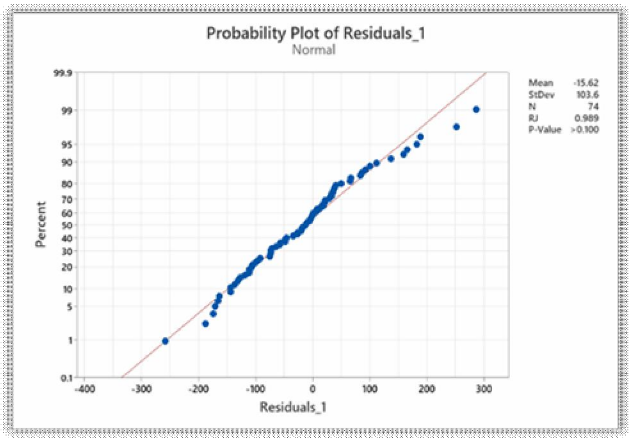


d. Normal probability plot:

Figure (9) represents the normal plot of the residual, which appears to be adequately fitted by straight line. However, the p-value of Anderson –darling statistic was 0.298 and p- value of Ryann –Joiner statistic (or W-statistic) was more than 10 %. This means we cannot reject the normality assumption, which is consistent with the theoretical assumption; hence, the statistical inferences are valid.

Figure (9): the normal plot of the residual

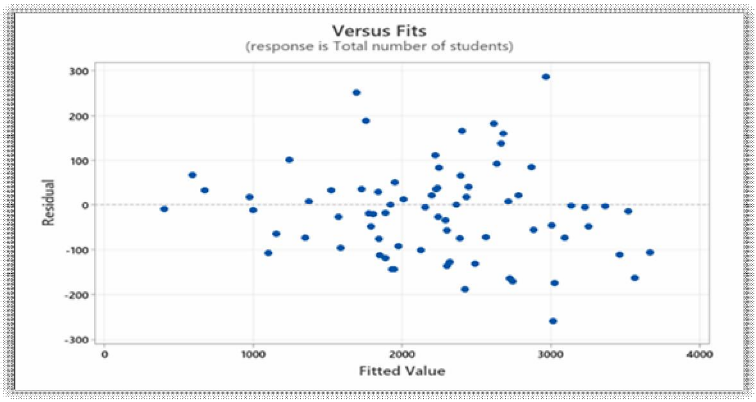




e. Residual versus fit:

Figure (10) appears to be devoid of all the regular patterns and moves that could be used to improve the model, the data randomly oscillating around the zero line

Figure (10): residuals versus fits



f. Modified Box and Pierce (Ljung-Box) Chi-Square Statistic

The following table 6 represent the computer output regarding this statistic:

Table (6): Box and Pierce (Ljung-Box) Chi-Square Statistic

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	**12**	**24**	**36**	**48**
Chi-Square	**5.2**	**14.2**	**18.7**	**23.2**
DF	**7**	**19**	**31**	**43**
P-Value	**0.633**	**0.769**	**0.960**	**0.994**

The P-values corresponding to the lags k = 12, 24, 36, 48 are all considerably large, so at k = 12, we find that the P-value corresponding to this statistic is equal to 0.633, and this indicates the existence of a collective random pattern in the first 12 autocorrelation coefficients for errors. Likewise, the corresponding P value of this scale at k = 24 indicates the presence of a collective random pattern in the first 24 autocorrelation coefficients of errors, and so on. These indicators greatly support the goodness of the identified model.

g. The first difference model of residuals (Δe_t)

ACF for Δe_t cuts off after lag 1, while PACF of Δe_t dies down. Thus Δe_t has pure moving average model of order one. The model is fitted for Δe_t and the results are shown in table (7).

Table (7): ARIMA model for Difference of residuals

Final Estimates of parameters				
Type	**Coef**	**Se Coef**	**T**	**P**
MA (1)	**0.9667**	**0.0351**	**27.50**	**0.000**
Number of observations: Original series 74, after differencing 73 **Residuals: SS = 802959 (backforecasts excluded)** **MS = 11152 DF = 72**				
Modified Box-Pierce (Ljung-Box) Chi-Square Statistic				
Lag	**12**	**24**	**36**	**48**
Chi-Square	**5.8**	**14.9**	**19.7**	**23.8**
DF	**11**	**23**	**35**	**47**
P-Value	**0.886**	**0.900**	**0.985**	**0.998**

Thus Δe_t follows has pure moving average model of order one with estimate 0.9667, and we must conduct the following test:

$H_0 : \theta = 1 \quad ; \; H_1 : \theta \neq 1$

To conduct this test, we use the standard normal statistic as follow:

$$|Z| = \left|\frac{\hat{\theta} - 1}{SE(\hat{\theta})}\right| = \left|\frac{0.9667\text{-}1}{0.0351}\right| = 0.95 < 2$$

This means that the real model parameter is not significantly different from 1. In addition, we find that the corresponding P-values for the adjusted Box and Peirce statistic support the fit of this model. In short, it can be inferred that the appropriate model for the series of first difference of the residuals resulting from the fitting the SARIMA (0, 1, 2) (0, 1, 3) model for the quarterly series is the MA (1) model with a parameter that does not differ significantly from the one. This is another indication that errors represent pure random errors.

5.3 Underfitting

i. Fitting the model SARIMA (0, 1, 1) (0, 1, 3).

When fitting this model, we found that the autocorrelation and partial autocorrelations functions have spikes at lag 2, this means that the SARIMA (0, 1, 1) (0, 1, 3) model fails to be a suitable replacement for the ARIMA SARIMA (0, 1, 2(0, 1, 3) model.

ii. Fitting the model SARIMA (0, 1, 2) (0, 1, 2).

When fitting this model, we found that the autocorrelation and partial autocorrelations functions have spikes at lag 12, this means that the SARIMA (0, 1, 2) (0, 1, 2) model fails to be a suitable replacement for the SARIMA (0, 1, 2(0, 1, 3) model.

5.4 Overfitting

i. Fitting the model SARIMA (0, 1, 3) (0, 1, 3).

The estimate of the added parameter θ_3 is 0.0484 with p-value 0.762. This means the added parameter does not differ significantly from zero and should be deleted from the model, this means that the SARIMA (0, 1, 3) (0, 1, 2) model fails to be a suitable replacement for the SARIMA (0, 1, 2(0, 1, 3) model.

ii. Fitting the model SARIMA (0, 1, 2) (0, 1, 4).

The estimate of the added parameter Θ_4 is 0.0613with p-value 0.656. This means the added parameter does not differ significantly from zero and should be deleted from the model, this means that the SARIMA (0, 1, 2) (0, 1, 4) model fails to be a suitable replacement for the SARIMA (0, 1, 2(0, 1, 3) model.

The above analysis may be summarized by saying that all the results of diagnostic tests and examinations support the appropriateness of using the identified model to analyze the data and the absence of clear reasons to doubt the suitability of the statistical hypotheses on which this model relies for the data. Then this model can be used in forecasting, as we will see.

6. Forecasting

Forecasting is the last stage of the modern analysis of time series. the SARIMA model (0,1,2) model, which successfully passed all the diagnostic tests and checks, has been used in forecasting the next six observations (years 2021-2022 and 2022-2023) and the results were as follows:

Table (8): forecasting (years 2021-2022 and 2022-2023)

Period	Forecast	Lower 95%	Upper 95%	Actual
79	**2379.09**	**2167.87**	**2590.31**	**2335**
80	**2479.16**	**2216.64**	**2741.69**	**2270**
81	**1797.7**	**1519.23**	**2076.18**	**1739**
82	**2499.16**	**2056.36**	**2941.97**	**?**
83	**2523.79**	**2000.72**	**3046.87**	**?**
84	**1670.51**	**1109.27**	**2231.76**	**?**

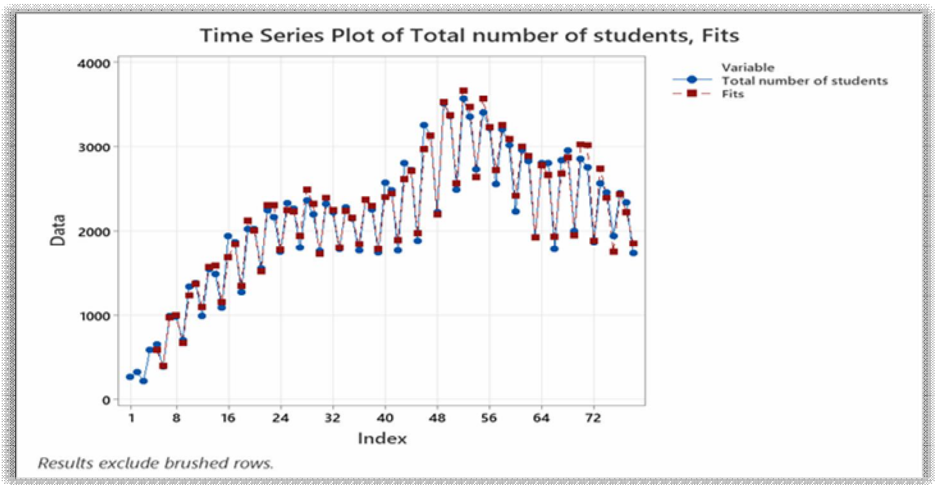
Table (9): the percentage of error

year	Semester	Time	Actual	Forecast	Percentage of error
2021-2022	1	79	2335	2590.31	-10.9%
	2	80	2270	2741.69	-20.8%
	3	81	1739	1797.70	-3.4%
Average Percentage of errors					-11.7%

On 15\9\2022 I had the actual number of students enrolled in the College of Administrative Sciences at Kuwait University for summer semester of the academic year 2021/2022. From table (9) it clear that all actual values fall within **the** 95% confidence intervals for all semesters and the average of the error percentage is 11.7.

From figure (11) The actual data and the fitted data are visualized in the following graph. This visualization allows us to see and track the number of students and their fitted numbers by date and compare between them. It is clear from the visualization that the fitted data are very close from the actual data and the movements and fluctuations of two series are very similar. The fitted graph can absorb the fluctuation or the pattern that recurs over a one-year period for the number of students, which occurs in the actual data.

Figure (11): time series plot for y(t) vs fits



7. Summery and conclusions:

The principal objective of this article is to model and forecast the series of number of students enrolled in the College of Administrative Sciences at Kuwait University from the academic year 1995/1996 to the academic year 2020/2021using the Box and Jenkins methodology which has been developed in 1970.It has been found the identified SARIMA (0, 1, 2) (0,1,3) model has successfully passed all the diagnostic tests and checks of invertibility, residuals, underfitting , and overfitting. In addition, this model has been used to forecast the number of quarterly students expected to be enroll in the college in the next two academic (2021-2022 and 2022-2023) . These forecasts may be useful for education strategic management and planning such as preparing enough teachers, sections or classrooms and course schedule managements for students who are expected to enroll in the near future. On the sixteenth of August 2022, I was able to obtain the actual number of students who were enrolled in the fall, spring, and summer semesters for the academic year 2021-2022. These numbers were compared to the corresponding forecasts; the average error percentage for these forecasts was -11.7% and all actual numbers were within the 95% confidence intervals.

REFFERNCES

[1] Abraham, B. and Ledolter (2005). Statistical Methods for Forecasting (2nd Ed.). New York: Wiley.

[2] Box, G.E.P., Jenkins, G.M. (1970). Time Series Analysis: Forecasting and Control, John Wiley & Sons.

[3] Box, G.E.P., Jenkins, G.M, Reinsel, G.C. and Ljung, G.M. (2016). Time Series Analysis: Forecasting and Control, fifth Edition, John Wiley & Sons.

[4] Bowerman, B. L. and O'Connell, R. T. (1993). Forecasting and Time Series: An Applied Approach (third Ed.). California: Duxbury Press.

[5] Chatfield, C. (2019). The Analysis of Time Series: Theory and Practice. An Introduction with R, Seventh Edition. Chapman & Hall.

[6] Harvey, A.C. (1993). Time Series Models. Second edition. The MIT Press

[7] Liu, L. M. (2009). Time Series Analysis and Forecasting, Second Edition, Villa Park, IL: Scientific Computing Associates.

[8] Shaarawy, S. M. Introduction of Modern Analysis of Time Series (in Arabic) (2005). Center of Scientific Publication, King Abdul-Aziz University, KSA.

[9] Shaarawy, S. M., Soliaman E., Ahmed T., Ali SH. (2014). Introduction to time series analysis and its applications in economics and management with the use of the Minitab program (in Arabic)

الملخص

نمذجة وتوقع عدد الطلاب المسجلين في كلية العلوم الإدارية بجامعة الكويت باستخدام نموذج SARIMA الضربي

هدف الدراسة: الهدف الرئيسي لهذه الدراسة هو إيجاد أفضل نموذج ممكن للسلسلة الزمنية الموسمية لعدد الطلاب المسجلين في كليه العلوم الإدارية بجامعة الكويت والتنبؤ بأعداد الطلاب المسجلين في الفصول الستة التالية.

منهجيه الدراسة: اعتمدت الدراسة على استخدام منهجية بوكس و جينكنز في الحصول على النموذج الأكثر ملاءمة من بين نماذج SARIMA (p, d, q) (P, D, Q) الممكنة، ودراسة واقع التغير الكمي في سلسلة الطلاب الملتحقين بكلية العلوم الإدارية بجامعة الكويت من العام الاكاديمي ١٩٩٦/١٩٩٥ الى العام الاكاديمي ٢٠٢١/٢٠٢٠.

بيانات الدراسة: بالاستعانة بعمادة القبول والتسجيل بجامعة الكويت تم الحصول على أعداد الطلاب المسجلين من العام الدراسي ١٩٩٥/١٩٩٦ الى العام الدراسي ٢٠٢٠/٢٠٢١ لجميع الفصول الدراسية (الخريف-الربيع-الصيف).

نتائج الدراسة: أوضحت الدراسة أن النموذج SARIMA(٢ ,٠،١) (٠،١،٣)) هو الافضل من بين جميع النماذج الموسمية الممكنة لتحليل السلسلة الزمنية لعدد الطلاب المسجلين في كلية العلوم الإدارية بجامعة الكويت. وقد اجتاز هذا النموذج جميع الاختبارات والفحوصات التشخيصية بنجاح، وتم استخدامه في التنبؤ بالعامين الدراسيين التاليين(٢٠٢١-٢٠٢٢) و (٢٠٢٢-٢٠٢٣). ومما لا شك فيه هذه التنبؤات قد تكون مفيدة للإدارة والتخطيط الاستراتيجي للتعليم مثل إعداد ما يكفي من المعلمين أو الأقسام أو الفصول الدراسية وإدارة جدول المواد للطلاب المتوقع تسجيلهم في المستقبل القريب.