African Journal of Hospitality, Tourism and Leisure, Volume 8 (1) - (2019) ISSN: 2223-814X

Modelling and forecasting Zimbabwe's immigrants using SARIMA models.

[1]Tendai Makoni*, [2]Talent. D. Murwendo, [3]Romeo Mawonike and [4]Musara Chipumuro
[1,3,4]Department of Mathematics and Computer Science
[2]Department of Pyhsics, Geography and Environmental Science
Great Zimbabwe University, Zimbabwe

Corresponding author*

Abstract

The remarkable increase in the Zimbabwe's migration is of major concern in terms of planning purposes and policy formulation by the government, Non-governmental organizations (NGOs), tourism authorities, investors and industrialists. It is important to have statistical models that can be used to project future migrants as the projections are vital in decision making purposes. The total monthly migrants (immigrants) of Zimbabwe from 2001 to 2017 are modelled using the Box Jenkins technique. The data for the study was availed by the Zimbabwe National Statistics Agency (ZIMSTAT). A time series plot determined the migration trend to Zimbabwe and December is noted as the seasonal month with a significant number of migrants. The parsimonious SARIMA $(1, 1, 1)(2,0,1)_{12}$ model fitted well to the migration data is validated by the Akaike information criterion (AIC), mean square error (MSE), root mean square error (RMSE) and other graphical measures. The model is used to project immigrants for the next 24 months. The projected migrants indicate a slow increase with the month of December being the peak month. The government of Zimbabwe, NGOs, investors and the industrialists should anticipate for an increased population due to an increase in immigrants and plan accordingly.

Keywords: Migration, immigrants, ARIMA, SARIMA, forecasting.

Introduction

Migration forecasts by any country are important as governments use the forecasts for planning their policy and social service actions, such as spending on health care, retirement programs, and antipoverty programs. The business people make use of these forecasts for strategic planning, developing new products on demand by the migrants as well as new required services. The government can treat incoming migrants as tourists since they can bring foreign currency into the country considering that some will be non-residents while others will be returning residents. According to Azam and Gubert (2006), migration is a decision that impacts in various ways on the welfare of households and the home community as well as the whole economy of a nation.

Smith, (1960), referred migration as a change in physical space while Shaw (1975) defines migration as the relatively permanent movement of persons over a significant distance. In simple terms, migration can be defined as the geographical movement of people across a specified boundary for the purpose of establishing a new permanent or semi-permanent residence. In this paper, we consider immigration as the movement of people into a new country. Furthermore an immigrant is referred to as an international migrant who enters the country from an area outside the country. Migration flows increase the destination population and decrease the source population by the same amount (Bijak 2010, Cohen 2012). Immigration positively improves trade, entrepreneurship and innovation.

Zakria and Muhammad (2009) noted overpopulation as the major issue on developing countries. The population of Zimbabwe is increasing through fertility and migration. This implies that the increase in population is associated with high crime rates, medical issues, educational aspects, food shortages, traffic issues and and also accommodation problems among others. Accurate or better migration forecasts can guide the responsible authorities in coming up with appropriate measures to mitigate problems. The Zimbabwean government, tourism managers and industrialists acknowledge the significance of migration statistics when it comes to future planning. All industry and tourism products are designed for local and international use and are used by the population (migrants) one way or the other.

The brain drain and substantial irregular emigration flows are some of the challenges being faced by Zimbabwe (ZIMSTAT Report, 2009). Managing these migration issues is difficult for the country because of the unavailability of a complete and coherent legal, institutional and policy frameworks for migration practices. Furthermore, the unavailability of adequate data and statistical analysis of the nature and extent of factors driving migration is of major concern in Zimbabwe as they hinder the government from devising appropriate migration programmes and policies for the benefit of the nation.

It is vital for Zimbabwe to be aware about the past, current and future migration trends as this helps in identifying the desired needs of the state. The migrants are sometimes treated like, or reffered to as tourists. Zakria and Muhammad (2009) stated that most developing countries have little knowledge about their population needs and size with the countries having good economic and social status being more aware of such issues. Without enough of the latest and finest migration knowledge, no single country will be scientifically and technologically able to be be on the right track concerning Migration trends are volatile over time (NRC, 2000). This means special attention is needed when determining migration trends. Migration trends are as important as tourism trends because they impact on several sectors. Migration data is an important asset to Zimbabwe and any other country since scientists make use of these data in determining significant migration trends. According to Shryock *et al.*, (1973), Jan *et al.*, (2007) and Agrawal, (2000), various methods (linear, nonlinear, regression models, simple and double exponential, simple decay and growth models, etc.) are used to forecast population and number of migrants.

Appropriate statistical time series forecasting methods like the autoregressive integrated moving average (ARIMA) and seasonal ARIMA (SARIMA) are important in migration modelling and forecasting. These methods are capable of capturing the time series annual and intra-annual patterns as they make use of available time series data. According to Keyfitz (1996), population forecasting is highly uncertain. This means careful attention and proper methods that capture all migration dynamics need to be adopted when forecasting population movement. The ARIMA and SARIMA models appear to be better and more accurate models. The construction of predictive intervals using ARIMA models is possible since they consider the time series as realisations of a stochastic process with uncertainty (Keilman et al., 2001).

The main aim of this paper is to model and forecast immigrants in Zimbabwe using a stochastic ARIMA/SARIMA model. Eveiw-7 and R software packages are used for model fitting and forecasting.

Material and methods

We obtained monthly migration data from ZIMSTAT and the data span from January 2003 to June 2017. The Box-Jenkins technique is adopted in this paper through developing ARIMA/SARIMA models. These models are not commonly used in modelling and forecasting migration, but they have the good forecast power and produce accurate forecasts.

The general ARIMA model is denoted ARIMA(p,d,q) where p, d and q refer to the orders of the non-seasonal AR, I and MA parts of the model respectively. The SARIMA model can be denoted by SARIMA(p,d,q)(P,D,Q)$_s$ where p, d and *q are ARIMA* orders and P, D and Q are the seasonal components of the seasonal AR, I and MA respectively with s being the seasonal period.

The time series plots are used to determine migration trends. The Augmented Dicky Fuller (ADF) test is used to test presence of unit root on the migration series. The autocorrelation function (ACF) and the partial autocorrelation function (PACF) are constructed for the purpose of identifying the order of the ARIMA/SARIMA model. The identified model is fitted along with other various stochastic ARIMA/SARIMA models. The AIC is used in the identification of the best model. The Mean squared error (MSE) and Root mean square error (RMSE) are used to evaluate the forecasting accuracy of the model. The graphical techniques (PACF, ACF, histogram, and normal probability plots of residuals) are also applied in validating the model. The parsimoniously and certified model is then used to forecast immigrants in Zimbabwe for the next 24 months.

Results and discussion

Descriptive statistics

As a way of describing the major features of Zimbabwean imgrants, the descripite statistics was done and is summarised in Table 1.

Table 1: Descriptive statistics of migration dataset

Minimum	Maximum	Median	Mean	Std. Dev	Skewness	Kurtosis
100965	978877	354682	366608	146730	1.274840	6.037957

(Source: Author's work)

Table 1 shows that the minimum value and maximum value of Zimbabwean immigrants is 100965 and 978877 respectively. The average number of immigrants is 366608. The data is not normal and is leptokurtic according to the positive skewness value and kurtosis value which is above 3.

Additionally, as a way of determining the monthly immigration trend, a time series plot is constructed using the original monthly migration series of Zimbabwe from January 2003 to June 2017 (Figure 1).

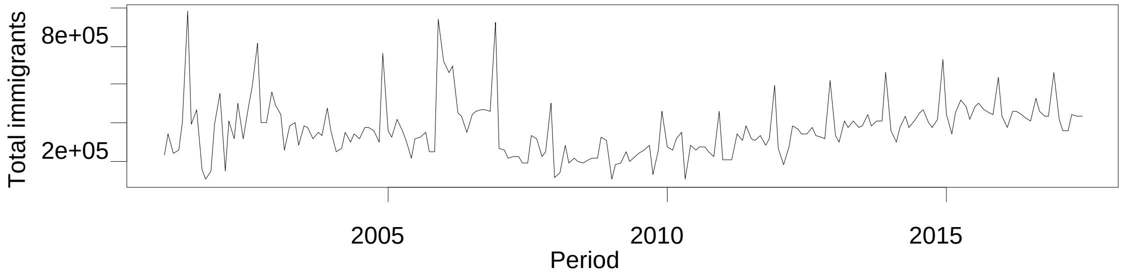

Figure 1: Time series plot for total immigrants (Source: Author's work)

Figure 1 exhibits a mixed pattern between 2001 and 2007, and an upward increasing trend starting around 2008 up to the end of June 2017, indicating a non-stationary time series. The nature of the pattern (Figure 1) could be as a result of political instability and economic meltdown, especially those years in which presidential elections were held and acquisition of land by comrades. The ADF test is conducted under the null hypothesis of the presence of a unit root on the original migration series to check stationarity of the series.

Table 2: ADF test results on original series

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-1.908649	0.3278
Test critical values:	1% level	-3.465780	
	5% level	-2.877012	
	10% level	-2.575097	

(Source: Author's work)

Table 2 results confirm the presence of unit root at 1%, 5% and 10% on the original series as supported by the ADF test statistic which is less than all the critical values. Hence the data is not stationary. We transformed the original data by taking the first difference and further conducted the ADF test.

Table 3: ADF test results on first difference of migration series

		t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic		-5.776597	0.0000
Test critical values:	1% level	-3.465780	
	5% level	-2.877012	
	10% level	-2.575097	

(Source: Author's work)

Table 3 results indicate that the first differenced data is stationary, therefore can be used to develop an ARIMA/SARIMA model. ACF and PACF of the differenced series are used to identify the tentative model.

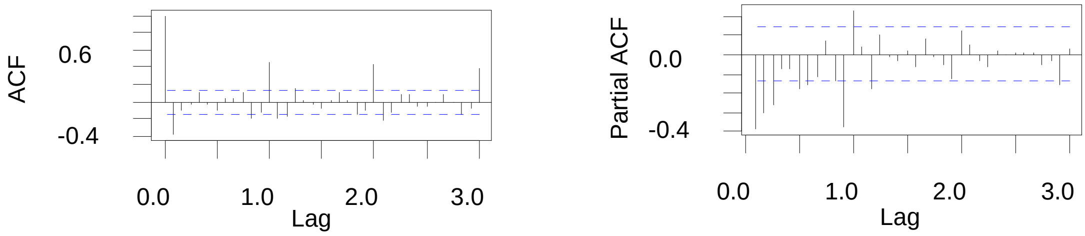

Figure 2: ACF and PACF of first differenced series (Source: Author's work)

African Journal of Hospitality, Tourism and Leisure, Volume 8 (1) - (2019) ISSN: 2223-814X

A SARIMA(1,1,1)(2,0,1)$_{12}$ model is being suggested by the ACF and PACF in Figure 2. This model is fitted together with other models and we apply the AIC criterion to identify the best model.

Table 4: Estimated models and AIC values

Model	AIC
SARIMA(1,1,2)(2,0,0)$_{12}$	5172.96
SARIMA(1,1,1)(2,0,0)$_{12}$	5164.15
SARIMA(0,1,1)(0,0,1)$_{12}$	5178.19
SARIMA(1,1,1)(2,0,1)$_{12}$	5154.09
SARIMA(1,1,2)(2,0,1)$_{12}$	5163.84

(Source: Author's work)

According to the AIC criterion, the SARIMA(1,1,1)(2,0,1)$_{12}$ model proved to be the best model because of its lowest AIC value as shown in Table 3. The coefficients of the selected model are summarised in Table 5.

Table 5: SARIMA(1,1,1)(2,0,1)$_{12}$ model parameters

SARIMA(1,1,1)(2,0,1)$_{12}$ model					
	ar1	ma1	sar1	sar2	sma1
Coefficients	0.3167	-0.8822	1.0957	-0.0967	-0.9683
s.e.	0.0964	0.0553	0.0955	0.0951	0.0787

(Source: Author's work)

All the coefficients displayed in Table 5 are statistically significant. Analysis of the model residuals is done starting with the ACF residual plot.

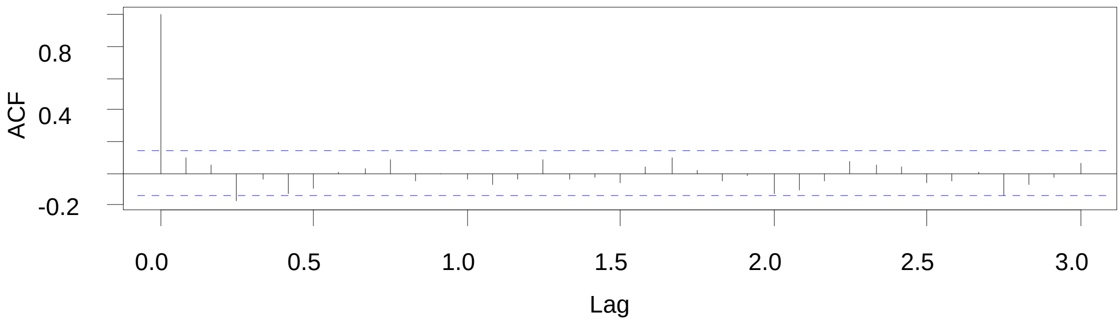

Figure 3: ACF plot of the SARIMA(1,1,1)(2,0,1)$_{12}$ model residuals (Source: Author's work)

Figure 3 suggests the rejection of the null hypothesis of autocorrelation on the model residuals besides one significant spike at lag one. Furthermore the Box-Ljung test statistic (= 23.554) associated with the p-value (0.2624) supports the conclusion of no autocorrelation at 5%. We performed a normality test of the model residuals through the plotting of the normal probability plot and the Jarque Bera Test.

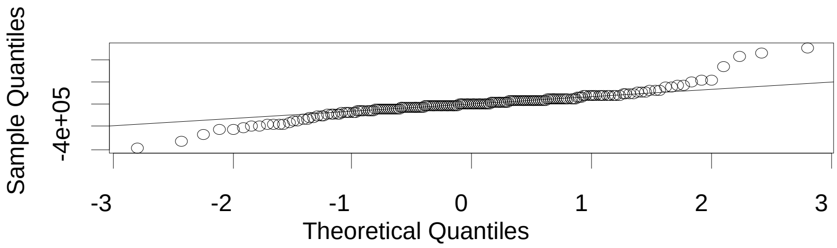

Figure 4: Q-Q plot of the SARIMA(1,1,1)(2,0,1)$_{12}$ model residuals (Source: Author's work)

The Q-Q plot displayed in Figure 4 shows normality of the model residuals indicated by an approximate straight line of points. Furthermore, the Jarque Bera test statistic (268.04) and the p-value (0.051) suggest normality of residuals at 5%. Since the model seems to be good according to diagnostic tests, the forecasting accuracy is tested before the model is used to come up with future immigrant estimates.

Table 6: Accuracy measures

Model	**RMSE**	**MAE**	**MAPE**	**MASE**
SARIMA(1,1,1)(2,0,1)$_{12}$ model	106355.3	68927.25	22.21034	0.698699
SARIMA(1,1,2)(2,0,1)$_{12}$ model	116301.4	73804.39	23.63879	0.7481375

(Source: Author's work)

The SARIMA(1,1,1)(2,0,1)$_{12}$ model has the lowest RMSE, MAE, MAPE and MASE as shown in Table 6 above. This means the model's forecasting performance is better than its counterpart, consequently, can be used in the prediction of future immigrants. The SARIMA(1,1,1)(2,0,1)$_{12}$ model is used to predict future immigrants for the next 24 months.

Table 7: Expected future immigrants

Month	**Forecast**	**Lo 80**	**Hi 80**	**Lo 95**	**Hi 95**
Jul-17	428825	288496	569155	214209	643441
Aug-17	476544	323532	629555	242533	710555
Sep-17	451305	294148	608462	210955	691655
Oct-17	404061	244448	563674	159954	648169
Nov-17	415611	253991	577231	168434	662787
Dec-17	654611	491128	818094	404585	904637
Jan-18	429216	263979	594454	176507	681925
Feb-18	383847	216847	550848	128442	639252
Mar-18	419552	250811	588293	161485	677619
Apr-18	440672	270209	611135	179972	701372
May-18	431697	259530	603864	168390	695004
Jun-18	446080	272226	619935	180193	711967
Jul-18	437697	257920	617473	162752	712641
Aug-18	476960	294079	659840	197268	756651
Sep-18	455727	270454	641001	172376	739079
Oct-18	406486	219038	593934	119809	693164
Nov-18	419078	229533	608624	129193	708963
Dec-18	658755	467142	850369	365707	951803
Jan-19	434886	241256	628517	138754	731019
Feb-19	391048	195432	586664	91879	690217
Mar-19	430571	232990	628153	128397	732746
Apr-19	445515	245988	645042	140365	750665
May-19	436616	235162	638070	128519	744713
Jun-19	452110	248748	655472	141094	763126

Table 7 shows the estimated immigrants together with the 80% and 95% confidence interval of immigrants. There is an overall increase of immigrants over the forecasted period with the month of December receiving highest number of immigrants. The graphical display of the forecasted immigrants is shown in Figure 5.

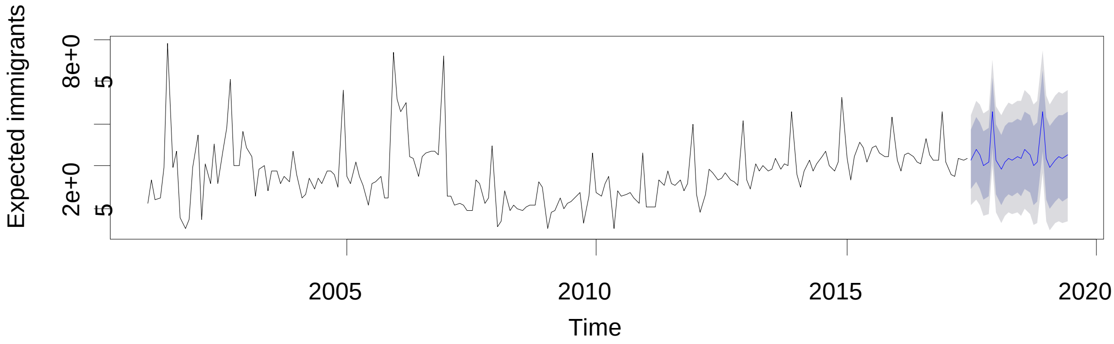

Figure 5: Immigrants forecasts from SARIMA $(1,1,1)(2,0,1)_{12}$ model (Source: Author's work)

Figure 5 displays the trend of the actual immigrants and forecasted immigrants along with the 80% and 95% confidence limits for the next 24 months. Accommodation and transport facilities may be increased during the peak month and this can be benefitial to the tourism sector and government in terms of money (foreign currency).

Conclusions and recommendations

It is noted that the SARIMA $(1, 1, 1)(2,0,1)_{12}$ is an appropriate model for forecasting total immigrants to Zimbabwe for the next 24 months. The model is supported by the AIC criteria, MSE, RMSE and other graphical measures. The highest number of immigrants is expected in the month of December and it is a seasonal month (Figure 5). These findings are very crucial to the government of Zimbabwe, the investors, the NGOs as well as the industrialists for future planning, investment planning, foreign direct investments and projects. The government of Zimbabwe should take note of the estimates in planning about the health, educational and recreational facilities to cater for these numbers and to put in place crime control measures as there might be an increase in crime rates The industrialists should design their products, anticipating these expected immigrants and should be in a position of providing the necessary products to the increased number of immigrants.

Acknowledgements

The authors would like to thank the ZIMSTAT for providing some of the research data and the Great Zimbabwe University for funding the research.

References

Agrawal, U.D. (2000). Population projections and their accuracy. Available at: https://www.amazon.com/Population-projections-their-accuracy-Agrawal/dp/8176461733 [Date Accessed: 18 October 2018].

Azam, J.P. & Gubert, F. (2006). Migrants' Remittances and the Household in Africa: A Review of the Evidence. *Journal of African Economies,*15(2), 426-462.

Bijak, J. (2010). Forecasting International Migration in Europe: A Bayesian View. Springer Series on Demographic Methods and Population Analysis, 24. Dordrecht: Springer.

Cohen, J. (2012). Projection of net migration using a gravity model. Tenth Coordination Meeting on International Migration, Population Division, United Nations, New York, 9-10 February 2012.

Das Gupta, B. & Lashley, R. (1975). Migration from Villages: An Indian Case Study. *Economic and Political Weekly, Annual Number*.

Jan, B., Ali, I. & Shah, S. (2007). Selecting of mathematical model for projections of NWFP population. *The Journal of Humanities and Social Sciences*, XV(2): 69-78.

Keilman, N., Pham, D. Q. & Hetland, A. (2001). Norway's uncertain demographic future. Social and Economic Studies 105. Oslo: Statistics Norway. Available at: https://www.researchgate.net/publication/242019483_Norway's_uncertain_demographic_future [Date Accessed: 2 January 2019].

NRC [National Research Council] (2000). *Beyond Six Billion. Forecasting the World's Population.* National Academies Press, Washington DC. Available at: https://www.hse.ru/data/2011/10/23/1268838271/Beyond%20Six%20Billion.pdf [Date Accessed: 20 August 2018].

Shaw, R.P. (1975). Migration theory and fact. Bibliography Series No.5. Philadelphia: Regional Science Research Institute.

Shryock, H. S. & Seigel, J. S. (1973). The methods and materials of demography. *Social Forces*, 52(1), 129–131,https://doi.org/10.1093/sf/52.1.129-a.

Zachariah, K. C., Prakash, B. A. & Rajan. S. (2002). Gulf Migration Study: Employment Wages and Working Conditions of Kerala Emigrants in United Arab Emirates". Centre for Development Studies (Thiruvananthapuram) Working Paper No.326. Available at: www.cds.edu [Date Accessed: 25 November 2018].

Zakria, M. & Muhammad, F. (2009). Forecasting the population of Pakistan using arima models. *Pak. J. Agri. Sci.,* 46(3).

Zimbabwe National Statistics Agency (ZIMSTAT) Report. (2009). Government of Zimbabwe, Harare, Zimbabwe.