"Science Stays True Here"
Journal of Mathematics and Statistical Science, Volume 2016, 47-58 | Science Signpost Publishing

A Bootstrapped Vector Autoregressive Model for Tourist Arrivals in Zimbabwe: A Case Study of Great Zimbabwe Monuments (2009-2012)

Makoni Tendai

Department of Mathematics and Computer Science, Great Zimbabwe University, Zimbabwe.

Abstract

The time series monthly data for local and international tourist arrivals at Great Zimbabwe Monuments from January 2009 to December 2012 were analyzed and modeled using bootstrapped vector autoregressive method. The first difference of the monthly bootstrapped data suggested stationarity. A vector autoregressive (VAR) model was estimated since there was no cointegration among the variables as suggested by the Johansen's cointegration test. Consumer Price Index (CPI), exchange rate and tourist arrivals were all accommodated in the model. Zimbabwean tourist visits are seasonal as evidenced by high numbers in August and December every year.

1. Introduction

The solid structures of Great Zimbabwe ruins were built over quite a long period from approximately 1200 years AD to 1450 years AD and they are a major center of tourist attraction in Zimbabwe, particularly Masvingo city. These ruins are one of the major tourist attraction centers in Zimbabwe. Lickorish et al., (1991) noted that monuments form an integral sector of tourism destinations and, if properly managed, they can generate substantial revenue. Great Zimbabwe Monuments destination attracts a large number of both local and international tourists. The Great Zimbabwe Ruins are sub-Saharan Africa's most important and largest stone ruins. There are large towers and structures built out of millions of stones balanced perfectly on top of one another without the aid of mortar. The skill with which the stones were laid is impressive given the lack of mortar hence it attracts more tourists across the world.

Insufficient air transport, poverty, diseases, insufficiency in facilities and accommodation were noted as major by Kester (2003). Hall and O'Sullivan (1996) noticed that political stability and political relations influences the image of destinations in tourist-generating regions, this applies to the Zimbabwean case as noticed by high volumes of tourist after theformation of the Government of National Unity (GNU) in Ferbruary 2009 and the introduction of multiple currencies the same period.According to the Zimbabwe Tourism Authority reports (2009), in 2008 tourism was the third largest foreign exchange earner in Zimbabwe after tobacco and gold. Zimbabwe started receiving more tourists after the formation of the Government of National Unity (GNU) in February 2009. The introduction of multiple currencies, political and economic stability contributed to the increase of tourist arrivals.

Christie and Crompton (2001) indicated that the tourism sector is already a growing contributor to GDP and exports in more than half of all African countries, making tourism an important sector in most developing economies. Tourism is one of the major contributor of employment and a major source of foreign currency.

Short-term forecasting methods in modelling tourists arrivals such as the univariant ARIMA approaches do not give a strong explanation on the major contributors on tourist arrivals hence these approaches need to be modified by considering other approaches such as vector autoregressive approaches. Coshall (2009) used univariate analysis, combined the ARIMA volatility and smoothing model in his study and his findings indicated that the ARIMA volatility models tend to overestimate demand while the smoothing models underestimate the number of future tourist arrivals. Song and Witt (2004) used the vector autoregressive (VAR) model to forecast international tourist flows to Macau for the period 2003 to 2008 and concluded that Macau's tourism demand will increase. These findings advocated the researcher to use the vector error correction method.

2. Data

Secondary monthly tourist's arrival data from January 2009 to December 2012 were used in this research. Arrivals data were acquired from the Zimbabwe Tourism Authority. Monthly secondary data from ZIMSTAT's Quarterly Digest of Statistics (2009-2013) for consumer price index (CPI) and exchange ratewere also used. Since vector autoregressive (VAR) models normally suffer from overfitting with too many free insignificant parameters, the researcher will consider only three variables.

3. Review of Methods Used

Bootstrapping is a method done by repeatedly drawing random samples, from the original data sample (Efron and Tibshirani, 1993). These resamples contain the same number of data points, $N = 48$, as the original sample. The fundamental nature of the bootstrap method is that if there is an observed sample of tourist arrivals, CPI and exchange rate, $x = \{x_2, x_3, \dots, x_n\}$ from unknown population distribution F(X) with mean μ and standard deviation δ. In order to obtain the unbiased maximum likelihood estimation of a population statistic μ, in this case μ represents the mean, the researcher resampled the observed sample X of tourist arrivals, CPI and exchange rate 1000 times (B = 1000). Resampling mimics the random process of the underlying system that generated F(X) and is done with replacement so that the bootstrap samples $x^*{}_b = \{x^*{}_1, x^*{}_2, x^*{}_3, \dots, x^*{}_n\}$, $\mathrm{b} = 1, 2, 3, \dots, 1000$ will be obtained. Since values are drawn with replacement the same sample value can occur more than once within a resample and the desired statistic $\hat{\mu}$ is calculated from each resample and is denoted by $\hat{\mu}^*$. The bootstrap estimate of μ, that is $\hat{\mu}^*$ was obtained by the formulae:

$$\hat{\mu}^* = \sum_b^B \frac{\hat{\mu}^*{}_b}{B}.$$

Johansen's (1988, 1991) maximum eigenvalue and trace tests on cointegration was used since it is commonly used if all variables in the system are I(1). Vector autoregressive (VAR) model can also be used in tourism forecasting. Vector autoregressive (VAR) modeltests and impose weak exogeneity restriction. In this research, 20 bootstrapped samples were considered. It was found that there were two cointegrating equations hence 20 vector error correction models were estimated. A VAR(p) model of the $(m \times 1)$ vector of time series $Z_t = (Z_{1t}, Z_{2t}, Z_{3t}, \dots, Z_{mt})$ with autoregressive order p can be expressed by the equation:

$$Z_t = a + A_1 Z_{t-1} + A_2 Z_{t-2} + \dots + A_p Z_{t-p} + \varepsilon_t$$

where A_i are $(m \times m)$ coefficients matrices and a is a $(m \times 1)$ intercepts, ε_t is a $(m \times 1)$ vector of disturbances that possesses the following properties.

The stochastic quadratic response model was as follows,

$$E(\varepsilon_t) = 0, (mean\ zero),$$
$$E(\varepsilon_t \varepsilon_t{}') = \textstyle\sum_\varepsilon (variance - covariance\ matrix),$$
$$E(\varepsilon_t \varepsilon_t{}') = 0\ \ for\ \ r \neq t\ \ (no\ serial\ correlation).$$

Final coefficients, A_i's of the vector error correction model will be obtained by the formulae:

$$\hat{A}^* = \sum_{b=1}^{1000} \frac{\hat{A}^*{}_{jb}}{1000}.$$

Estimation and evaluation of the forecast quality were based on data of monthly tourist arrivals in Masvingo at Great Zimbabwe Monuments. Forecasting performances of the estimated vector autoregressive (VAR) model was later evaluated using the mean square error (MSE) which can be expressed as:

$MSE = \frac{1}{n}\sum_{t=1}^{n}(A_t - F_t)^2$, where A_t is the actual value and F_t is the forecast value. The forecasted values for the year 2014 were compared with the actual values of that year. Multivariate normality test of residuals was done using the Jarque-Bera's multivariate normality test.

4. Results

4.1 Tourist Arrivals Time Series Plot

The pattern of the tourist arrivals at Great Zimbabwe Monuments was determined by the use of time series plots.

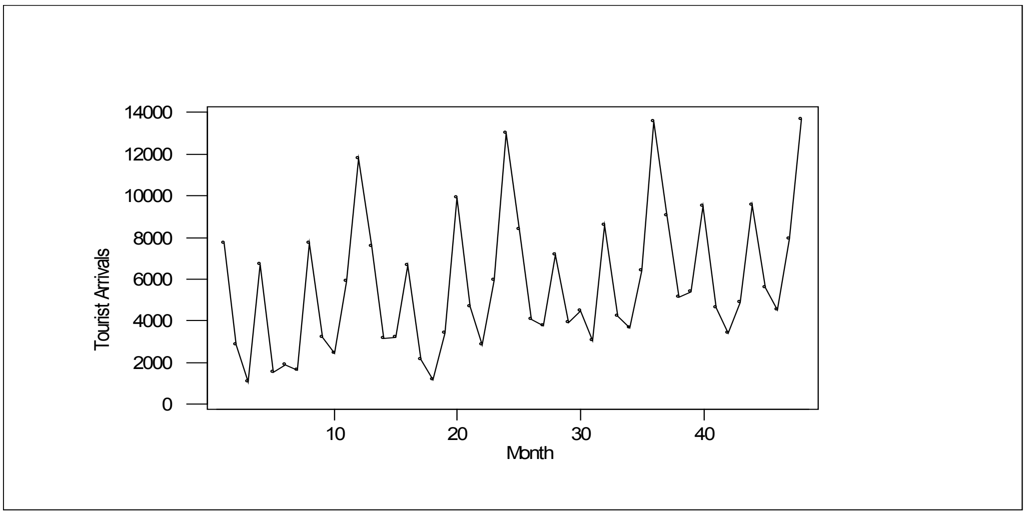

Figure 4.1 Great Zimbabwe Monuments time series plot of the monthly tourist arrivals

The Great Zimbabwe Monuments time series plot for the monthly tourist arrivals suggested existence of a pattern. Anoted high values in the tourist arrivals especially in every August December suggests a seasonal variation. Seasonal variations may result from public holidays in August and December. The time series plots also suggest that the data is non stationarity. After noticing that the, first, second, third and fourth difference of the data still suggest non stationarity, the researcher then bootstrapped the data. Bootstrapped data became stationary after the first difference as suggested by Figure 4.2 below.

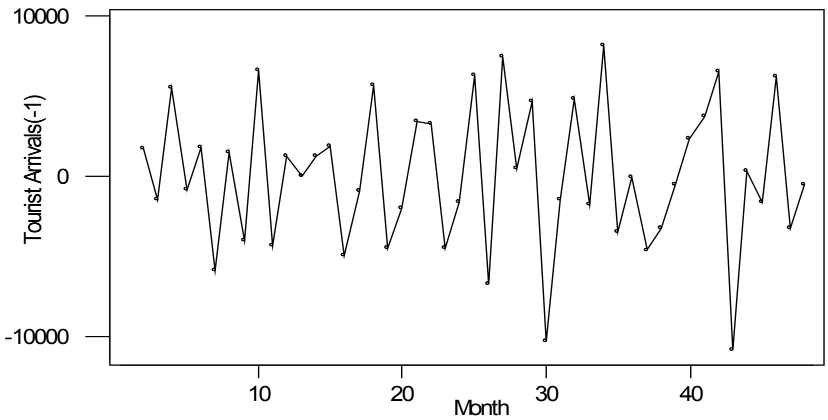

Figure 4.2 Stationary bootstrapped monthly tourist arrivals

CPI and exchange rate data were not stationary; hence, it was also bootstrapped. First difference of the bootstrapped data became stationary.

4.2 Cointegration Test

Since the bootstrapped variables are I(1), the Johansen Maximum Likelihood (ML) procedure was then used to determine whether a stable long-run relationship exists between the variables. Johansen's cointegration test was conducted on the bootstrapped data since this is a multivariate analysis whereby more than one cointegrating vector being present. Cointegration test was carried under the null hypothesis which states that there is no cointegration. Results obtained from the test are summarised in Table 4.1 below.

Unrestricted Cointegration Rank Test (Trace)				
Hypothesized . No. of CE(s)	Eigenvalue	Trace Statistics	0.05 Critical Value	Prob.**
None	0.338354	24.39313	29.79707	0.1843
At most 1	0.126415	6.633061	15.49471	0.6207
At most 2	0.018926	0.821633	3.841466	0.3647

Trace test indicates nocointegratingeqn(s) at the 0.05 level				
* denotes rejection of the hypothesis at the 0.05 level				
**MacKinnon-Haug-Michelis (1999) p-values				
Unrestricted Cointegration Rank Test (Maximum Eigenvalue)				
Hypothesized No. of CE(s)	Eigenvalue	Trace Statistics	0.05 Critical Value	Prob.**
None	0.338354	17.76007	21.13162	0.1391
At most 1	0.126415	5.811428	14.26460	0.6375
At most 2	0.018926	0.821633	3.841466	0.3647
Max-eigenvalue test indicates no cointegration at the 0.05 level				
* denotes rejection of the hypothesis at the 0.05 level				
**MacKinnon-Haug-Michelis (1999) p-values				

Table 4.1 Cointegration results

From the above Table 4.1, all the nullhypotheses are being accepted since all the *p*-values are greater than 0.05, suggesting that there are no cointegrating equations. Non-existence of cointegrating equations implies that a vector autoregressive (VAR) modelwas going to be estimated.

4.3 Bootstrapped Vector Autoregressive Model

Let X = Tourist Arrivals, Y = CPI and Z = Exchange rate. Table 4.2 below shows a summary of the VAR model.

	X	**Y**	**Z**
X(-1)	0.117557 (0.15614) [0.75287]	-1.20E-06 (1.9E-05) [-0.06156]	1.61E-05 (1.5E-05) [1.06270]
X(-2)	-0.084148 (0.15701) [-0.53594]	4.09E-06 (2.0E-05) [0.20847]	-4.15E-05 (1.5E-05) [-2.71448]
Y(-1)	0.000241 (0.00011) [2.21101]	1.198797 (0.07454) [16.0827]	0.035016 (0.05810) [0.60273]
Y(-2)	-26.74136 (643.065)	-0.193077 (0.08027)	0.010826 (0.06256)

	[-0.04158]	[-2.40534]	[0.17304]
Z(-1)	3.63E-05 (1.39E-05) [2.970743]	0.480220 (0.18171) [-0.31793]	-0.057770 (0.14162) [3.39083]
Z(-2)	-2203.487 (1431.24) [-1.53956]	-0.033441 (0.17865) [-0.18719]	0.251348 (0.13924) [1.80512]
C	7120.853 (13555.3) [0.52532]	0.322083 (1.69202) [0.19035]	-2.142757 (1.31876) [-1.62482]
R-squared	0.578745	0.888955	0.746773
Adj. R-squared	0.62986	0.877256	0.707815
Sum sq. resids	4.75E+08	7.404584	4.498006
S.E. equation	3490.760	0.435731	0.339608
F-statistic	25.55594	18.0211	19.16865

Table 4.2 Estimated Bootstrapped Vector autoregressive (VAR) model results.

From the Table 4.2, there are three dependent variables, namely X, Y and Z. Furthermore, there are two lags at each variable. Summarized below are the three equations obtained from the table above.

$$X = C(1)*X(-1) + C(2)*X(-2) + C(3)*Y(-1) + C(4)*Y(-2) + C(5)*Z(-1) + C(6)*Z(-2) + C(7) \quad (4.1)$$

$$Y = C(8)*X(-1) + C(9)*X(-2) + C(10)*Y(-1) + C(11)*Y(-2) + C(12)*Z(-1) + C(13)*Z(-2) + C(14) \quad (4.2)$$

$$Z = C(15)*X(-1) + C(16)*X(-2) + C(17)*Y(-1) + C(18)*Y(-2) + C(19)*Z(-1) + C(20)*Z(-2) + C(21) \quad (4.3)$$

4.4 Model Coefficient Test

A coefficient test helps to streamline unnecessary coefficients in the models. Significance of coefficients in equation 4.1, 4.2 and 4.3 were tested and the results for the test are summarised in Table 4.3.

	Coefficient	Std. Error	t-Statistic	Probability
C(1)	0.117557	0.156144	0.752875	0.4530
C(2)	-0.084148	0.157011	-0.535940	0.5930
C(3)	0.000241	0.000109	2.211009	0.0250
C(4)	-26.74136	643.0649	-0.041584	0.9669
C(5)	3.63E-05	1.39E-05	2.970743	0.0034
C(6)	-2203.487	1431.242	-1.539563	0.1264
C(7)	7120.853	13555.29	0.525319	0.6004
C(8)	-1.20E-06	1.95E-05	-0.061557	0.9510
C(9)	4.09E-06	1.96E-05	0.208469	0.8352
C(10)	1.198797	0.074540	16.08270	0.0000
C(11)	-0.193077	0.080270	-2.405341	0.0177
C(12)	-0.057770	0.181708	-0.317929	0.7511
C(13)	-0.033441	0.178653	-0.187186	0.8518
C(14)	0.322083	1.692025	0.190353	0.8494
C(15)	1.61E-05	1.52E-05	1.062698	0.2901
C(16)	-4.15E-05	1.53E-05	-2.714477	0.0076
C(17)	0.035016	0.058096	0.602735	0.5479
C(18)	0.010826	0.062562	0.173038	0.8629
C(19)	0.480220	0.141623	3.390827	0.0010
C(20)	0.251348	0.139242	1.805115	0.0736
C(21)	-2.142757	1.318762	-1.624825	0.1069

Table 4.3 Coefficient test for bootstrapped data

Wald---x(-1) and x(-2) jointly cannot influence X

Six coefficients out of the twenty-one coefficients in Table 4.3 above are significant to explain the dependant variables since their *p*-values are all less than 0.05. After taking note of the significant coefficients, the following equations were derived.

$$X = 27.7777767107*Y(-1) + 1934.19128651*Z(-1) \quad (4.4)$$

$$Y = 1.19879709864*Y(-1) - 0.193076565358*Y(-2) \quad (4.5)$$

$$Z = - 4.14641783044e\text{-}05*X(-2) + 0.480219750173*Z(-1) \quad (4.6)$$

4.5 Model Checking

Diagnostic checks are done before models are used for specific purposes in order to make sure that they represent the data satisfactorily.

4.5.1 Autocorrelation Test of Residuals

Autocorrelation of residuals was tested using the Portmanteau Tests for Autocorrelations and the results are summarized in Table 4.4.

Lags	Q-Stat	Prob.	Adj Q-Stat	Prob.	df
1	2.146166	NA*	2.193858	NA*	NA*
2	0.313289	NA*	5.504942	NA*	NA*
3	1.25904	0.7932	11.86552	0.7532	16
4	6.91205	0.8850	18.05690	0.8400	25
5	5.18472	0.8634	27.33844	0.7838	34
6	32.71343	0.8727	35.99646	0.7666	43
7	2.14903	0.8335	47.12562	0.6656	52
8	6.97134	0.9068	52.96316	0.7584	61
9	3.06292	0.9343	60.53647	0.7829	70
10	9.32663	0.9519	68.54010	0.7934	79
11	9.48757	0.9273	81.89448	0.6630	88
12	5.79813	0.9453	90.43230	0.6680	97

Table 4.4 VAR Residual Portmanteau Tests for Autocorrelations

From the summarized results in Table 4.4 above, the researcher failed to reject the null hypothesis which says no autocorrelations on residuals up to lag h as evidenced by *p*-values which are more than 5%.

4.5.2 Normality Test

The multivariate generalization of the Jarque-Bera test (Jarque Bera 1987) was used to test the multivariate normality of the u_t since this tests the skewness and kurtosis properties of the u_t.

Componet	Skewness	Chi-sq	df	Prob
1	0.540713	2.241510	1	0.1343
2	0.464124	1.651483	1	0.1988
3	0.004739	0.000172	1	0.9895
Joint		3.893164	3	0.2732
Component	Kurtosis	Chi-sq	df	Prob
1	2.627840	0.265464	1	0.6064
2	3.305887	0.179336	1	0.6719
3	3.099757	0.019074	1	0.8902
Joint		0.463873	3	0.9268
Component	Jarque-Bera		df	Prob
1	2.506973		2	0.2855
2	1.830819		2	0.4004
3	0.019246		2	0.9904
Joint	4.357038		6	0.6285

Table 4.5 Normality results

Residuals are multivariate normally distributed as evidenced by all the p-values which are more than 5%.

All the above processes were done on all the 20 bootstrapped samples implying that 20 VAR model were estimated. Averaging the coefficients of the 20 bootstrapped VAR models resulted in coming up with the final models which are:

$$X = 8152.8530+0.5567*X(-1) - 0.4148*X(-2) + 1234.1913*Z(-1) - 2803.487*Z(-2) + 21.6878*Y(-1) - 14.7445*Y(-2) \quad (4.7)$$

$$Z = 2.1461*X(-1) +9.1885*X(-2) -1.2180*Z(-1) + 3.2354*Z(-2) + 0.0089*Y(-1) + 3.6501*Y(-2) - 5.2743 \quad (4.8)$$

$$Y = 1.5342- 3.1359*X(-1) + 10.0857X(-2) - 0.8533*Z(-1) + 1.3008*Z(-2) + 9.87971*Y(-1) - 0.7647*Y(-2) \quad (4.9)$$

The forecasting accuracy of the above models was evaluated using the mean square error (MSE).

5. Conclusions

Forecasting results established by the estimated vector autoregressive model seems to produce relatively accurate forecasts as evidenced by the minimal variance between the actual values of the year 2013 and the forecasted ones. Furthermore, tourist arrivals generally increased in Zimbabwe since 2009 mainly because of the formation of the Government of National Unity (GNU) that causes political stability as well as the introduction of a multi currency system which led to the stability of the economy.

5.1 Recommendations

There is need to improve infrastructure by the Zimbabwe Tourism Authority and the government, especially accommodation and road networks so to ease accommodation problems during holidays. Though, variables such as (CPI) and exchange rates were found to be very important in modeling tourist arrivals in Zimbabwe, other variables such as tourist receipts and transport cost need to be captured and included in further studies.

References

[1]. Christie, I., T., and Crompton, D., E., (2001). Tourism in Africa: Africa Region Working Paper Series, 12 Washington DC: World Bank.

[2]. Coshall, J.T. (2009). Combining volatility and smoothing forecast of UK demand for international tourism. Tourism Management, 30 (4), 495-511.

[3]. Efron, B. (1979). Bootstrap methods: Another look at the jackknife. Annals of Statistics, 7, 1-26.

[4]. Efron, B., Tibshirani, R. J. (1993). An introduction to the bootstrap. New York: Chapman Hall.

[5]. Hall, M., and O'Sullivan, V., (1996). Tourism, political stability and violence. In P. Abraham & M, Yoel (Eds.), Tourism crime and international security issues. John Wiley & Sons, New York.

[6]. Johansen, S, 1991, “Estimation and Hypothesis Testing of Cointegration Vectors in Gaussian Vector Autoregressive Models,” Econometrica, Vol. 59, No. 6, pp. 1551-1580.

[7]. Johansen, S., 1988, Statistical Analysis of Cointegration Vectors, Journal of Economic Dynamics and Control, Vol. 12, No. 2-3, pp. 231-254.

[8]. Kester, J., G., C., (2003). International tourism in Africa: Tourism Economics, Vol 9, No 2, pp 203-221.

[9]. Lickorish, L.J., Jefferson, A., Bodlender, J., & Jenkins, C.J. (1991). Developing Tourism Destinations: Policies and Perspectives. Essex, England: Longman Group, Westgate House.

[10]. Liu, T-R.,M.E. Gerlow and S.H. Irwin (1994), The performance of alternative VAR models in forecasting exchange rates, International Journal of Forecasting, 10, 419-433.

[11]. Simkins, S. (1995), 'Forecasting with vector autoregressive (VAR) models subject to business cycle restrictions', International Journal of Forecasting, 11, 569-583.

[12]. Song, H., and Witt, S. F. (2004), Tourism Demand Modelling and Forecasting: Modern Econometric Approaches, Pergamon, Oxford.

[13]. Stock, J. H. and M. W. Watson (2001). Vector Autoregressions. The Journal of Economic Perspectives 15, 101-115.

[14]. Tan, A. Y., Mccahon, C., and Miller, J. (2002). Modeling tourist flows to Indonesia and Malaysia. Journal of Travel and Tourism Marketing, 13 (1/2), 63-84.

[15]. Tourism Trends and Statistics Reports (2009-2012), Zimbabwe

[16]. Zimbabwe Tourism Authority, (2008).

[17]. Zimbabwe Tourism Authority, (2009). Zimbabwe Tourism Statistics and Economic Indicator Highlights.