International Tourism and Hospitality Journal 1(2): 1-10 (2018)

Print ISSN: 2616-518X Online ISSN: 2616-4701



ITHJ

International Tourism and Hospitality Journal Journal Homepage: https://rpajournals.com/ithj

Forecasting Tourists Arrivals at Chinhoyi Caves Recreational Park in Zimbabwe Using SARIMA Model

Tendai Makoni*¹, Delson Chikobvu² Bongani Maphosa³

Great Zimbabwe University, Zimbabwe^{1,3} University of the Free State, South Africa²

Abstract

Forecasting tourist arrivals and volatility for Chinhoyi Caves Recreational Parks using SARIMA-GARCH models are vital for decision and planning purposes. The monthly data from the Zimbabwe Parks and Wildlife Management Authority (ZIMPARKS) for the period January 2010 to December 2016 was used. The time series plot indicated mixed fluctuations in tourist arrivals and exhibited seasonality. The series became stationary after a logarithmic transformation and a seasonal difference. The Akaike Information Criteria (AIC) and the Bayesian Information Criterion was used in selecting the best model and a SARIMA(1,0,1)(1,1,0)12 model fitted well to the data. The residuals of the model confirmed to all tests. Accuracy measures like the mean absolute error (MAE) and the mean absolute percentage error (MAPE) supported the model and an out of sample for the next 24 months (2017 to 2018) were done. From the forecasts, an increase in tourist arrivals was noted. The results imply the need for enough accommodation facilities in and around the recreational centre.

Keywords: Tourist arrivals, ARIMA models, SARIMA models, time series plot

*Corresponding author: Tendai Makoni¹; Email: tpmakoni@gmail.com

Submission Date: 9 June 2018 Acceptance Date: 17 June 2018

Introduction

Forecasting is a key tool for planning and decision-making in modern business organisations, according to Thoplan (2014). In the tourism industry, forecasting has become even more important with the increasing competition among nations and businesses for international tourists. Tourism plays a crucial role in the emerging economies all around the world. While tourism can bring into the country some foreign currency, it can also create employment opportunities for the populace and business growth opportunities for the Tour and Safari Operators, Charter aircraft operators, car hire companies, hotel and restaurant operators and the general dealers.

The tourism sector is a significant contributor to employment, tax revenues and earnings of foreign exchange. Tourism contributes 10% of World GDP, 7% of Global trade and employs one in ten jobs globally (UNWTO, 2016). Tourist arrivals to the African continent in 2016 constituted 5% of the global total at 58 million (UNWTO, 2016). The report indicates that tourist arrivals increased by 46million globally, a 4% increase on 2015 arrivals. The increase means an increase in foreign currency and infrastructure since tourism have the strong impetus for economic growth at all levels, that is, globally, continentally and nationally. The development and growth of the hotel and leisure infrastructure are dependent on the stability and growth of the tourism sector.

Zimbabwe enjoys a fair share of tourist arrivals from all over the world with numbers on the increase over the past five years. The country recorded a total of 1 956 000 tourist arrivals in 2008 while 1 794 000 and 2057000 tourist arrivals were recorded in 2012 and 2015, respectively (UNWTO, 2016). Tourist arrivals to Zimbabwe in 2016 reached 2 167 686, which is a 5% increase on 2015 arrivals. (https://tradingeconomics.com). This also translates to revenue generation of US\$819 million dollars, a 10.9% contribution to the country's GDP (ZIMSTAT, 2017). The majority of the tourists' main interest is normally to visit the country's wildlife areas on photographic and hunting safaris. We also have tourists whose interest lies in visiting cultural sites and mystical areas like the Victoria Falls' Rainforest, Chinhoyi Caves, Inyangani Mountains, Chimanimani's bridal veil, Vumba gardens and the culturally rich Great Zimbabwe Ruins.

Chinhoyi Caves Recreational Park is located about 8km north of Chinhoyi town along the Harare-Chirundu highway and is one of the 12 recreational parks out of 40 plus estates under the management and control of the Zimbabwe Parks and Wildlife Management Authority (ZIMPARKS). ZIMPARKS is a Government of Zimbabwe parastatal under the Ministry of Tourism, Hospitality Industry and Environment, responsible for the conservation and protection of the country's wildlife resource and it's habitat, management and preservation of some natural and cultural heritage sites as per the PARKS AND WILDLIFE ACT [CHAPTER 20: 14]. It occupies a very small piece of land measuring 120 hectares (Parks and Wildlife Act 20: 14) and owing to this limiting size, the high volume of tourists visiting the site every month calls for careful planning on the part of the park management.

The main attraction at the Chinhoyi Caves is "the sleeping pool" (Chirorodziva), a huge craven in a limestone rock formation with constant level sky-blue water and constant temperatures throughout the year. Many locals believe and consider the pool to have magical and cultural relevance to their lives. The huge craven is surrounded with numerous tunnels laden with stalactites and stalagmite rock formations that afford one numerous angled view of "the sleeping pool". These features need delicate handling to ensure their longevity.

The management needs to come up with appropriate strategies to handle the influx of visitors; hence it is paramount to appropriately forecast tourist arrivals at the site. The site attracts curious deep-water divers, who continually crave to dive to the bottom of "the sleeping pool", a feat no living being has been able to achieve. Others visit the site for spiritual redemption, others for cultural purposes and others to satisfy their ever hungry minds for knowledge of the history or geographical beauty of the area. Yet, others just get in to get a snapshot or two through their cameras for long-term memories of its beauty.

With the varied tastes derived from the site by different types of tourists, the numbers of tourists arriving at the site are bound to grow beyond limits. Due to its location along the highway, hence ease of access, Chinhoyi Caves Recreational Park draws most of

the international tourists using the Harare-Chirundu highway, who, out of curiosity make a stopover to view and relax at the mystical site. Unlike the Victoria Falls (Rainforest) that draws very high international tourist arrivals to its mystical beauty, Chinhoyi Caves draws a large volume of local visitors the bulk of which are scholars in search for more knowledge on the history and culture of the Makonde people that used the caves in the pre-19th century.

The location aspect of Chinhoyi Caves Recreational Park presents a challenge in planning. It has a limited space that cannot accommodate large volumes of tourists at any time and there has been a steady rise in the number of tourist arrivals over the years, a situation that requires appropriate interventions before the site loses its aesthetic value. Presently, ZIMPARKS does not have a tourist arrival forecasting model for Chinhoyi Caves Recreational Park, a situation that leaves huge gaps in the management plans. The monthly variations in tourist arrivals at Chinhoyi Caves make it very difficult for the ZIMPARKS management to effectively plan for the materials and other resource allocations for the site. This creates the need to devise an appropriate model that would enable the management to precisely predict the tourist numbers so as to manage the limited available space, plan for the tourism activities and arrange the essential materials and resources to ensure its long-term sustainability.

Historical trends and seasonal patterns are explored in time series modelling in order to predict the future of the variable based on the identified trends and seasonal components (Rangika, 2016). This means tourism patterns can be identified by making use of time series models; hence a time series approach is adopted in this paper. Therefore, the aim of the paper is to determine the behaviour of monthly tourist arrivals at Chinhoyi Caves Recreational Park for the period January 2010 to December 2016 and develop a time series forecasting model for tourist arrivals.

Literature Review

The methods employed in tourist arrivals forecasting are numerous, including the Naive Models, the exponential smoothing techniques, autoregressive integrated moving average (ARIMA), seasonal autoregressive integrated moving average (SARIMA) and the autoregressive moving average (ARMA) and some causal methods like the regression analysis. These approaches have varying degrees of accuracy, with the accuracy of each method generally going a long way in ensuring successful business planning by those employing them.

Various researchers have different views about the best forecasting method, but Witt and Witt (1995) concluded that no single forecasting method that performs consistently across different situations. Saayman and Botha (2015) noted that the demand for accurate predictions of tourist arrivals and tourism revenue has led to the application of various modelling techniques such as ARIMA and SARIMA as these methods are capable of producing accurate tourism forecasts.

Turner and Witt, (2001) noted the need for short-term tourist arrival forecasts and acknowledged that they can be achieved through the use of ARIMA and SARIMA models. They further indicated that tourism industry should become more short-term focused, and designed to change rapidly with changing market demand. This means ARIMA and SARIMA models can be applied in coming up with short-term tourism forecasts. Claveria and Torra, (2014) used the data for 2001 to 2009 tourist arrivals to Catalonia and observed that ARIMA models outperformed Neural Network models especially for short horizons. In Turkey, Yilmaz (2015) used monthly tourism data and the results revealed that the

SARIMA models provide more accurate short-term forecasts when compared to the structural time series models. The findings are further supported by Kim and Moosa (2005) who, in their comparative study on the forecasting accuracy of SARIMA models against Regression based models and structural models confirm the superiority of SARIMA models. Goh and Law (2002) used the SARIMA and Multivariate ARIMA (MARIMA) models, and compared their forecasting accuracy against a number of exponential smoothing models, moving average models, as well as a naive model. The SARIMA and MARIMA models outperformed all the models across.

Wan et al., (2013) adopted a SARIMA model and compared it against with a seasonal moving average model and a Holt-Winter model and noted that the SARIMA model was the best performing model under all three different h-step-ahead forecasting horizons (where his one month, three-months and twelve-months ahead). Kulendran and Wong, (2005) suggested that ARIMA models provide more accurate forecasts for a time series that has fewer seasonal variations, while SARIMA provides more accurate forecasts for a time series that has a strong seasonal variation. Tourism series usually exhibit seasonality hence can be modelled using SARIMA models. Lim and McAleer, (2002) estimated various ARMA models over the period 1975(1) to 1996(12) for monthly tourist arrivals to Australia from Hong Kong, Malaysia and Singapore and found out that the SARIMA models outperform the non-seasonal ARMA models.

In forecasting tourist arrivals in Singapore, Chu, (2004) concluded that ARIMA-type models outperform other models. According to Song et al., (2003b) the ARIMA models are often considered as delivering more accurate forecasts than econometric forecasting techniques. Lin, et al., (2011), uses ARIMA, ANNs and Multivariate Adaptive Regression Splines (MARs) to model tourist arrivals forecasting to Taiwan and concludes that ARIMA models outperform ANNs and MARs approaches and provide an effective model for forecasting tourist arrivals.

According to Song and Li (2008), naive models are benchmark models normally used for comparing forecasting performance of other models. Smeral and Wuger (2005) applied the ARIMA, SARIMA and Naïve methods for forecasting tourist arrivals and revealed that advanced models like ARIMA or SARIMA could not even outperform the simple Naive model. Fethi (2014) in a study on the application of the Naive method on monthly tourism data for Tunisia finds that its forecasting accuracy is relatively poor compared to SARIMA model. Apart from this, Turner, et al., (1995a) compared the forecasting performance of the ARIMA model, the Holt Winters exponential smoothing method and the naive model. The models were fitted to quarterly international tourist flow data to New Zealand, from June 1978 to September 1992. Forecasting performance was compared between eleven different countries and world regions. They concluded that the Winters and ARIMA methods outperform the Naive model.

Kulendran and Witt, (2003) use a number of ARIMA specifications, a causal structural time-series model (STSM), a basic structural model (BSM), as well as, the naive model of no change. Their findings suggest that the ARIMA models exhibit superior predictive ability in the short-run forecasts; however, none of the models could outperform the naive model in the medium-run forecasts. This claim is further supported by Goh and Law, (2002) who indicated that the overall performance of ARIMA models is superior to that of the Naive models and Smoothing techniques when they applied ARIMA models with interventions in forecasting tourism demand using ten arrival series for Hong Kong. Vu and Turner, (2006) compared ARIMA models against a BSM for the case of Thailand and the ARIMA models showed in their study a better forecasting accuracy.

Methodology

A quantitative approach has been adopted in forecasting the monthly tourist arrivals at Chinhoyi Caves. The data analysis will be done using R statistical package. Explanatory data analysis is done as a way of getting a summary of the descriptive statistics. Time series plots are used to visualise the behaviour of the tourist arrivals over the study period. The Box-Jenkins (1971) approach is used in coming up with the best ARIMA or SARIMA model. The monthly data on tourist arrivals to Chinhoyi Caves Recreational Park for the period January 2010 to December 2016 accessed from the ZIMPARKS Headquarters in Harare, Zimbabwe is used.

ARIMA and SARIMA models

The ARIMA model can be expressed as:

$$\phi(B)Y_t = \Theta(B)a_t \tag{1}$$

where $\phi(B)$ and $\Theta(B)$ are polynomials of B used for model coefficients, Y_t is the monthly tourist arrival series and a_t is an error term. A SARIMA model can be expressed as:

$$\phi_p \phi_{PS}(B^s) \nabla^d \nabla^{DS} Z_t = \theta_q \Theta_{QS}(B^s) a_t \tag{2}$$

Here, B and B^s are the ordinary backward shift operator and seasonal backshift operator respectively. ∇^d and ∇^{Ds} are the differencing operator and the seasonal differencing operator respectively. The differencing operator is for making the data stationary as the models can only be fitted using stationary data.

The Augmented Dickey-Fuller (ADF) test is used to test the presence of unit root on the dataset. The autocorrelation functions (ACF) and partial correlation functions (PACF) are used to identify model lags. The Akaike Information Criterion (AIC) is used in the selection of the best ARIMA or SARIMA model.

Predictive performance

A number of measures are used in judging forecasting accuracy of fitted models. Majority of the tourist arrival forecasting studies used Mean Absolute Percentage Error (MAPE), Root Mean Square Error (RMSE), Root Mean Square Percentage Error (RMSPE) and Theil's U statistic to examine the performance of forecasting in their tourism arrivals forecasting models. However, the Mean Absolute Percentage Error (MAPE) is adopted in this study and can be expressed as:

$$MAPE = \frac{1}{n} \sum_{t=1}^{n} \left| \frac{\varepsilon_t}{Y_t} \right| 100 \tag{3}$$

Here, $\mathcal{E}_t = Y_t - f_t$, Y_t is the actual value and f_t is the forecasted value and n is the size of the observations in the dataset. Minimum MAPE value indicates the best model.

Result and Discussion

Data analysis of the monthly tourist arrivals at Chinhoyi Caves Recreational Park is done using R, E-views 7 and Excel software packages.

Time series plot

Time series plots are employed to investigate the existence of a trend, seasonality, regular and irregular fluctuations. Figure 4.1 exhibits the observed patterns.

December 2016

9000
8000
7000
6000
4000
3000
2000
1000
2010
2011
2012
2013
2014
2015
2016

Figure 4.1: Tourist arrivals pattern to Chinyoyi Caves from January 2010 to December 2016

Figure 4.1 shows monthly time series plot of the observations on tourist arrivals to Chinhoyi Caves Recreational Park for the period January 2010 to December 2016. High peaks in the plot normally occur during school terms whereby schools send their students to the Chinhoyi Caves for leisure and educational purposes. A positive trend line in the series is observed and some seasonal fluctuations. The presence of a trend line and seasonal fluctuations indicate the non-stationarity of the series.

Augmented Dickey-Fuller test

To verify the presence of unit root, an ADF test is conducted and the results suggested the acceptance of the null hypothesis of presence of a unit root in the original data series. This implies that the series is not stationary. The data is transformed using a logarithm transformation then followed by a seasonal difference of period 12. ADF test is repeated and results are displayed in Table 4.1.

Table 4.1: Unit root result after log-transform and seasonal difference

	T-Statistic	P-value
Augmented Dickey-Fuller test statistic	-8.121575	0.0000
Test critical values: 1% level	-3.524233	
5% level	-2.902358	
10% level	-2.588587	

Table 4.1 above suggest the rejection of the null hypothesis of presence of a unit root in the original data series. This is supported by the critical values, which are less than the ADF

test statistic at all levels, and also the p-value, which is less than 5%. The data can now be used to fit a model.

Model selection

The ACF and PACF plot are used to identify the order of the proposed model. The ACF and PACF plots (see Appendix 1) suggested a SARIMA $(1,0,1)(1,1,0)_{12}$ model. This suggested model is fitted together with other models.

Table 4.2: SARIMA models and their AIC and BIC

Model	AIC	BIC
SARIMA(2,0,2)(1,1,0)12	17.77	17.86
SARIMA(2,0,1)(1,1,0)12	17.834	17.92
SARIMA(1,0,1)(1,1,0)12	17.148	17.288
SARIMA(1,0,2)(2,1,0)12	17.636	17.725
SARIMA(2,0,1)(2,1,0)12	17.92	18.01

The results in Table 4.2 suggested that the SARIMA $(1,0,1)(1,1,0)_{12}$ model is the better model according to the AIC and BIC because of the smallest AIC and BIC among the fitted models. The optimal parameters of this model are obtained by the Maximum Likelihood Estimation (MLE) method. The model parameters summarised in Table 4.3

Table 4.3: SARIMA $(1,0,1)(1,1,0)_{12}$ model parameters

Coefficient	Value	T-value	P-value
ϕ_1	0.965289	49.12564	0.0000
θ_1	-0.981252	-48.81358	0.0000
ϕ_1	-0.501204	-4.334637	0.0001

From Table 4.3, optimal parameters for the model are displayed and all are significant at 1% level of significance. The model is checked for adequacy, before it is used for forecasting. ADF unit root test is applied and the results are displayed in Figure 4.2.

Figure 4.2: Results of Unit Roots Test for the Model

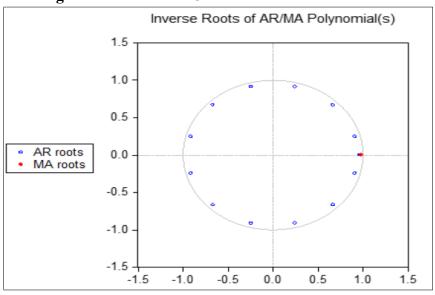


Figure 4.2 above shows that no unit lies outside the circle, hence the model is both stationary and invertible. The Jarque-Bera test for normality also reveals that the model residuals are normally distributed because of the test statistic of 1.244492 and the p-value of 0.236737. Furthermore, forecasting accuracy is examined. The model has smaller RMSE of 0.3478821, MAE of 0.2433325 and MAPE of 3.10758. Since the model seems to be good, it is then used to come up with out of sample forecasts for the year 2017 and 2018. Figure 4.3 exhibit the out-of-sample forecasts for the next 24 months.

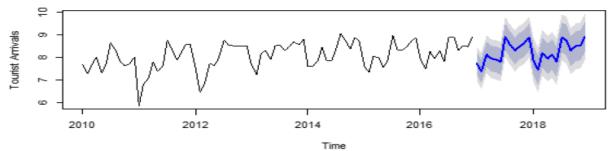


Figure 4.3: 24-Out-of-sample tourism forecasts from the SARIMA(1,1,1)(1,1,0)₁₂ model

Figure 4.3 shows seasonal variations and a positive trend pattern signalling continued growth in tourist arrivals numbers at Chinhoyi Caves. Forecasts from the model indicated an increase in future tourist arrivals hence the authorities managing Chinhoyi Caves must put in place enough accommodation and ablution facilities for the estimated tourist arrivals.

Conclusion

Forecasting of tourist arrivals to Chinhoyi Caves Recreational Park is very important for planning in Management of the fragile ecosystem that forms the core of the 120ha area. Mass tourism might negatively impact the area's aesthetic value. In this research, original monthly tourist arrival series were non-stationary and become stationary after a log-transformation followed by a seasonal difference of period 12. A SARIMA (1,0,1)(1,1,0)₁₂ model fitted well to the dataset. The developed model was tested for adequacy through residual analysis with recorded values of RMSE, MAE and MAPE and proved to be a good model. Forecasts from the model indicated an increase in future tourist arrivals hence the authorities managing Chinhoyi Caves must put in place enough accommodation and ablution facilities to accommodate the predicted increase number of tourist arrivals.

References

Athanasopoulos, G. & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. Journal of Tourism Management, 29, 19-31.

Box, G.E.P., and Jenkins, G.M., (1976). Time series analysis, forecasting and control, San Francisco, CA: Holden Day.

Brida, J. G. & Risso, W. A., (2011). Research note: Tourism demand forecasting with SARIMA models the case of South Tyrol. Tourism Economics 17 (1), 209-221.

Burger, C.J.S.C, Dohnal, M, Kathrada, M & Law, R (2001). A practitioners guide to time series methods for tourism demand forecasting, A case study of Durban, South Africa, Tourism management, 22: 403-409...

- Claveria, O., & Torra, S. (2014). Forecasting tourism demand to Catalonia: Neural networks vs. time series models. Economic Modelling 36 (C), 220-228.
- Chan, F., Lim, C., & McAleer, M., (2005). Modelling Multivariate International Tourism Demand and Volatility. Journal of Tourism Management, 26, 301-479.
- Chu, F.L., (1998). Forecasting Tourist Arrivals. Journal of Travel Research, 36, 79-84.
- Coshall, J. T. (2009). Combining volatility and smoothing forecasts of UK demand for international tourism. Tourism Management 30 (4), 495-511.
- Du Preez, J. & Witt, S.F., (2003). Univariate versus multivariate time series forecasting: An application to international tourism demand, International Journal of forecasting, 19:435-451.
- Frances, P. H., (1998). Time Series Models for Business and Economics Forecasting, Cambridge University Press, Cambridge.
- Goh, C. & Law, R., (2002). Modelling and forecasting tourist demand for arrivals with stochastics non-stationary seasonality and intervention, Tourism management, 23:499-510.
- Kulendran, N., Witt, S. F., (2003). Forecasting the demand for international business tourism. Journal of Travel Research 41 (3), 265-271.
- Lim, C., & McAleer, M., (2005). Analysing the Behaviour Trends in Tourist Arrivals from Japan. Journal of Travel Research, 43, 414-421.
- Lim, C. & McAleer, M., (2000). Seasonal Analysis of Asian Tourist Arrivals to Australia. Journal of Applied Economics, 32, 499-509.
- Pai, P.-F., Hung, K.-C. & Lin, K.P., (2014). Tourism demand forecasting using novel hybrid system. Expert Systems with Applications 41 (8), 3691-3702.
- Rangika I. P, H., (2016). A seasonal ARIMA Model of tourism forecasting: The case of Sri Lanka, Journal of Tourism, Hospitality and Sports.
- Song, H., & Witt, S.F., (2003). Tourism forecasting: the general-to-specific approach. Journal of Travel Research, 42, 65-74.
- Thoplan R., (2014). Qualitative v/s Quantitative Forecasting of Yearly Tourist Arrivals in Mauritius, International Journal of Statistics and Applications, 198.
- Vu, J. C. & Turner, L. W., (2006). Regional data forecasting accuracy: The case of Thailand. Journal of Travel Research 45 (2), 186-193.

Appendix

ACF and PACF plots

Autocorrelation	Partial Correlation	-	AC	PAC	Q-Stat	Prob
.* .	.* .	1	-0.170	-0.170	2.4762	0.116
*** .	*** .	2	-0.368	-0.408	14.240	0.001
.1. 1	.* .	3	0.023	-0.165	14.287	0.003
. **	. [*.	4	0.252	0.085	19.969	0.001
.i. í	. *.	5	0.029	0.097	20.047	0.001
****	*** .	6	-0.487	-0.410	41.763	0.000
.1. 1	** -	7	-0.023	-0.283	41.812	0.000
. **	.* .	8	0.335	-0.073	52.373	0.000
.* .	** -	9	-0.102	-0.229	53.364	0.000
** .	** .	10	-0.227	-0.232	58.363	0.000
.].]	** .	11	0.008	-0.272	58.369	0.000
. ***	.1. 1	12	0.435	0.029	77.167	0.000
.1. 1	.1. 1	13	0.038	0.039	77.312	0.000
** .	.1. 1	14	-0.284	0.031	85.588	0.000
.1. 1	.1. 1	15	0.067	0.061	86.056	0.000
. *.	.1. 1	16	0.161	-0.035	88.792	0.000
.1. 1	. *.	17	0.072	0.197	89.343	0.000
*** .	.1. 1	18	-0.396	-0.043	106.40	0.000
.1. 1		19	-0.031	0.068	106.51	0.000
. *.	.* .	20	0.195	-0.075	110.76	0.000
. .	. *.	21	0.019	0.114	110.80	0.000
** .	.* .	22	-0.281	-0.177	119.95	0.000
.1. 1	.1. 1	23	0.054	0.032	120.29	0.000
. **		24	0.329	0.026	133.21	0.000
.1. 1	.1. 1	25	0.039	0.008	133.39	0.000
** .	.1. 1	26	-0.220	0.004	139.37	0.000
	.]. [27	0.024	-0.025	139.44	0.000
. *.	.* .	28	0.172	-0.069	143.26	0.000
.1. 1	.1. 1	29	0.039	-0.031	143.45	0.000
** .	.1. 1	30	-0.288	800.0	154.52	0.000
.* .	.* .	31	-0.090	-0.134	155.62	0.000
. *.	.* .	32	0.191	-0.079	160.67	0.000
.1. 1	.1. 1	33	0.026	0.012	160.76	0.000
.* .	.1. 1	34	-0.147	0.042	163.87	0.000
.* .	.1. 1	35	-0.079	-0.032	164.78	0.000
. **	.1. 1	36	0.267	0.001	175.46	0.000