ISSN 2249-5908
Issue 3, Vol.6 (November-2013)

Identification of SARIMA as a Model for Forecasting Indian Leather Export

*Jagathnath Krishna K.M., NithiyananthaVasagam S., Giriyappa, K. and Chandramouli, D.

Economics Research Division
CSIR- Central Leather Research Institute,
Adyar, Chennai-600 020, India

Abstract

Leather industry occupies a prominent role in the Indian economy in view of its massive potential for export. To assist in decision making process, this article proposes a time series model based on leather export values from India during January, 1999 to March, 2013. Based on the information criterion, we identified ARIMA (0, 1, 1) (1, 2, 1) model as a reasonable model to forecast the export values of leather. This proposed model is used to forecast the export values of leather for the year 2013 – 2014.

Key words: AIC, BIC, Leather Export, Forecast, SARIMA model, Time series.

1. Introduction

The leather industry in India holds an important position in contributing towards the economy of the nation as it is one of the oldest manufacturing industries. The industry not only contributes towards development of the country's economy by way of production, but also by the way of offering a wide range of employment opportunities. The industry is showing a massive potential of offering more export growth and employment opportunities. Recently, the export of leather and leather products gained great momentum and the export of Indian leather products have shown marginal growth. This is because of the best planning made by the industry for optimum utilization of available raw materials and resources.

Over several years of its existence, the leather industry in the nation has undergone drastic changes from being a mere exporter of raw materials in the early 60's and 70's. Now, the country has become one of the largest exporters of value-added leather products and finished goods due to the contribution made by the big and small scale industries. One of the reasons behind the development of leather industry in the nation is because of the various policy initiatives taken by the Indian government. To frame such policies, detailed research on the trends in exports and its statistics is required.

In India, the leather sector is spread across and produces a comprehensive range of products from raw hides to fashionable shoes. The industry consists of firms in all capacities, including small artisans to major global players. These have been an increasing emphasis on the planned development of industry, which is aimed at optimum utilization of available raw materials for maximum returns, especially from exports.

The forecasting of trends in leather export is necessary for decision making and framing policies to support the export. There are some studies related to marine products (Venugopal and Prajneshu, 1996) and rubber export (Tatiporn and Kanchana, 2012) which make use of statistical modelling techniques for forecasting the export trends. The statistical modelling of the export trends in leather are not studied by any of the researchers. This motivated us to look into the export trends through scientific methods like, time series modelling. The time series modelling is widely applied in finance, business, marketing, economics, astronomy, meteorology, medicine (Imhoff et. al. 1997), etc (see Brockwell and Davis(1996)). For the present study, we considered monthly data from January, 1999 to March, 2013. The next section deal with the methodology adopted for the study, in section 3, the model identification is given and in last section some concluding remarks are given.

2. Methodology

The method applied for analysing the 13 years of leather export data is discussed in this section. According to Schumway and Stoffer (2006), time series is the characteristics of data that seemingly fluctuate in a random fashion over time. Usually the time series data are assumed to be stationary. If the series is stationary, the modelling of data can be done using auto regressive moving average (ARMA) process. The autoregressive models are formed based on the idea that the current value of the series x_t, can be explained as a function of *m* past values $x_{t-1}, ..., x_{t-m}$, where *m* determines the number of steps into the past needed to forecast the current value. The extent to which it is capable of forecasting a data series from its own past values can be assessed by looking into the autocorrelation function. Whereas, a moving-average model is conceptually a linear regression of the current value of the series against current and previous (unobserved) white noise error terms or random shocks. The ARMA model is defined as follows.

Definition 2.1: A time series $\{x_t; t = 0, \pm 1, ...\}$ is *ARMA (p, q)* if it is stationary and

$$x_t = \phi_1 x_{t-1} + ... + \phi_p x_{t-p} + w_t + \theta_1 w_{t-1} + ... + \theta_q w_{t-q}$$

with $\phi_p \neq 0, \theta_q \neq 0$ and $\sigma_w{}^2 > 0.$ The parameters p and q are called the orders of autoregressive and the moving average respectively.

If the data is non-stationary, either transformation technique or differencing is applied at appropriate lag to achieve stationarity. In many situations, especially in economics and finance, the data are not stationary in general. So a refined model that could accommodate the differencing between the lag becomes important. Box and Jenkins (1970) has developed autoregressive integrated moving average (ARIMA) model that could accommodate the non-stationarity for modelling and prediction. The integrated ARMA or ARIMA model is a broadening of the class of ARMA models to incorporate differencing.

Definition 2.2: A process, $\{x_t\}$ is said to be *ARIMA (p, d, q)* if

$$\nabla^d x_t = (1-B)^d x_t$$

is*ARMA(p, q)*. In general, we will write the model as

$$\phi(B)(1-B)^d x_t = \theta(B) w_t .$$

If $E\left[\nabla^d x_t\right] = \mu,$ we write the model as

$$\phi(B)(1-B)^d x_t = \alpha + \theta(B) w_t \text{, where } \alpha = \mu\left(1-\phi_1 - ... - \phi_p\right).$$

Like any other data analysis, there are a few steps to fit ARIMA models to time series data. First plot a time series graph and search for some trends or outliers. If the variability increases over time, the variance in the data has to be stabilized using transformation. The second step is to identify the values for autoregressive of order *p*, order of differencing *d* and moving average order *q*. After fixing the value of *d*, by looking into the sample auto correlation function (ACF) and partial autocorrelation function (PACF) of the corresponding differenced data, the values for p and q can be identified. After identifying the models, the parameter involved in the model has to be estimated.

The monthly leather export data for a period of 13 years considered for the present study. The autocorrelation plot shows regular fluctuations indicating the presence of seasonal variation. Hence seasonal ARIMA model, a modified ARIMA model to account seasonal and non-stationary behaviour is considered for our study. The seasonal ARIMA or SARIMA model is defined as follows.

Definition 2.3: The seasonal autoregressive integrated moving average model or SARIMA of Box and Jenkins (1970) is given by

$$\Phi_P\left(B^s\right)\phi\left(B\right)\nabla_s{}^D\,\nabla^d x_t = \alpha + \Theta_Q\left(B^s\right)\theta\left(B\right)w_t,$$

where w_t is the usual Gaussian white noise processes. The general model is denoted as *ARIMA (p, d, q) (P, D, Q)*$_s$. The ordinary autoregressive and moving average components are represented by polynomials $\phi\left(B\right)$ and $\theta\left(B\right)$ of orders p and q respectively and the components $\Phi_P\left(B^s\right)$ and $\Theta_Q\left(B^s\right)$ of orders P and Q, and ordinary and seasonal differencing component by $\nabla^d = \left(1-B\right)^d$ & $\nabla^D = \left(1-B^s\right)^D$.

In the final step, among the class of identified models, one model has to be considered which satisfies all the statistical criteria for forecasting. This model should have minimum Akaikie information criteria (Burnham & Anderson, 1998) and Bayesian information criteria (Schewarz, 1978).

Definition 2.4: Akaikie's information criteria (AIC) is defined by

$AIC = \ln L + 2k,$ where L is the Gaussian likelihood function and k is the number of free parameters.

Definition 2.5: Bayesian information criteria (BIC) is given by

$$BIC = -2\log L + k\log n.$$

Apparently, the difference between the AIC and BIC is the penalty term, instead of *2k*, it *is k log n.* However, BIC gives an asymptotically consistent estimate of the order of the true model.

3. Results and Discussions

The data used for the present study was obtained from Council of Leather Export (CLE), Government of India. The data consists of 171 observations on monthly export values of leather from January, 1999 to March, 2013. The time series plot for the data is given in Figure 1.

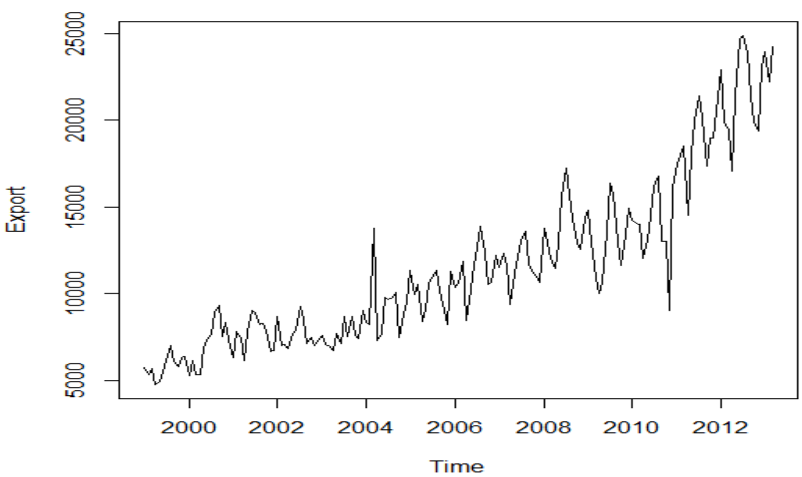

Figure 1: Time series plot for leather export

It can be observed from the Figure 1, that the time series is not stationary and has seasonal variation. Therefore we differenced the original data once at lag 1 and also taken the seasonal difference once and the plot is given in Figure 2.

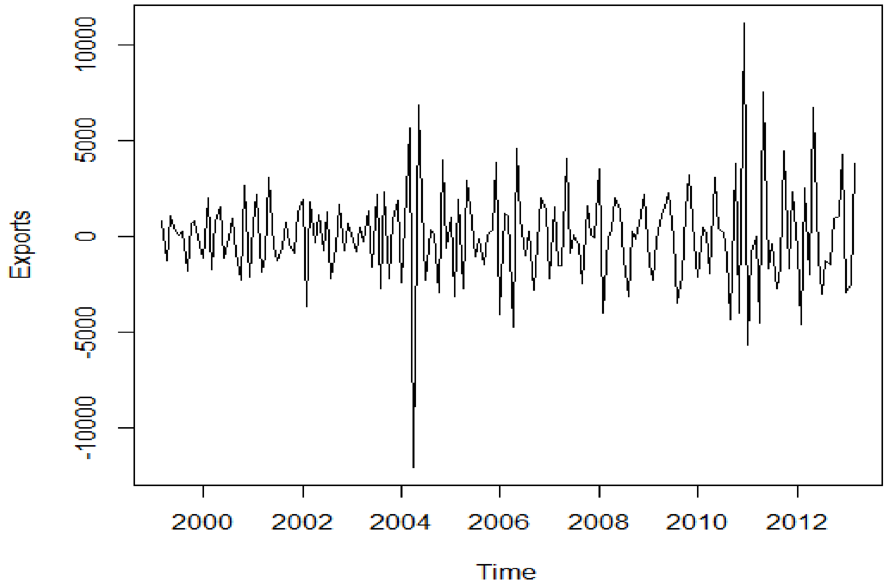

Figure 2: Time series plot with ordinary difference 1 and seasonal difference 1.

The differenced series is appeared to be stationary and we confirmed the stationarity by conducting the unit root test, Augmented Dickey Fuller (1979) test. As the p-value for the test is less than 0.05, we reject the unit root null hypothesis at 5% level of significance, which shows the data are stationary. Even though now our data is stationary, to search for possible models for forecasting, we tried with one more seasonal differencing and conducted the unit root test, which also showed significance and is given in Figure 3.

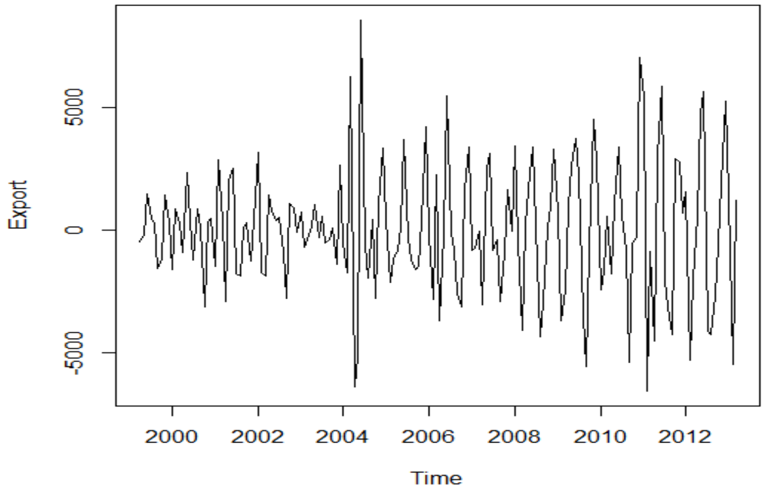

Figure 3: Time series plot with ordinary difference 1 and seasonal difference 2.

From Figure 2 and 3, it is observed that the series is stationary with few outliers. The ACF and PACF plots of ordinary differencing at lag 1 and seasonal differencing at lag 2 are shown in Figure 4.

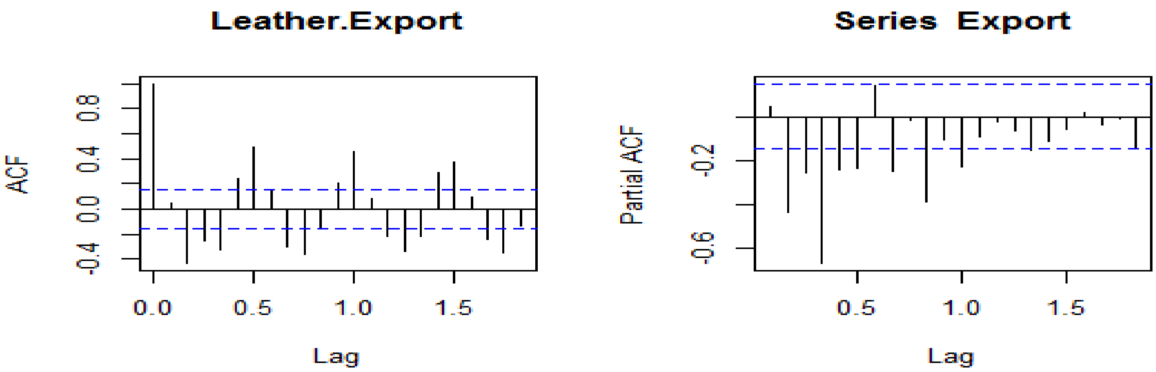

Figure 4: ACF and PACF plot for ordinary difference 1 and seasonal difference 2.

Based on the ACF and PACF models have been identified. Model selection was made using AIC and BIC, which are a goodness of fit for particular model by balancing the errors of the fit against the number of parameters in the model. For each of the identified model, AIC and BIC values are estimated and are given in Table 1.

ISSN 2249-5908
Issue 3, Vol.6 (November-2013)

Table 1: Comparison of SARIMA models

Model	AIC	BIC
ARIMA(0,1,1)(1,1,1)	2725.4	2737.65
ARIMA(0,1,1)(1,2,1)	**2590.59**	**2602.52**
ARIMA(0,1,1)(0,2,1)	2616.57	2625.52
ARIMA(1,1,1)(0,2,1)	2617.34	2629.28

From Table 1, the model 2, ARIMA (0,1,1) (1,2,1) which has the least information criteria (AIC, BIC) is considered as the best fit and the model is given by

$$\left(1-\Phi_1 B^{12}\right)\left(1-B^{12}\right)^2(1-B)x_t=\left(1+\Theta B^{12}\right)(1+\theta B)w_t \qquad (1)$$

For ARIMA (0,1,1)(1,2,1) model, the estimated parameter values are $\theta=-0.5849, \Theta=-0.9973$ and $\Phi=-0.4311$, with standard errors 0.0748, 0.0737 and 0.0737 respectively. Hence the model (1) can be rewritten as

$$\left(1+0.4311\times B^{12}\right)\left(1-B^{12}\right)^2(1-B)x_t=\left(1-0.9973\times B^{12}\right)(1-0.5849\times B)w_t \qquad (2)$$

Where $\{x_t\}$ is the original time series on export of leather and $\{w_t\}$ is the white noise.

To validate our fitted model, normality test on residuals were performed. The null hypothesis for the test is that the residuals are white noise in nature. The fitted autocorrelation function (ACF) for the residuals shows the residues are following normal distribution. Since the Ljung–Box randomness test fail to reject the null hypothesis and the Figure 5 gives a clear indication that the residuals are normally distributed. Hence our model 2, *ARIMA(0,1,1)(1,2,1)* is appropriate for prediction.

ISSN 2249-5908
Issue 3, Vol.6 (November-2013)

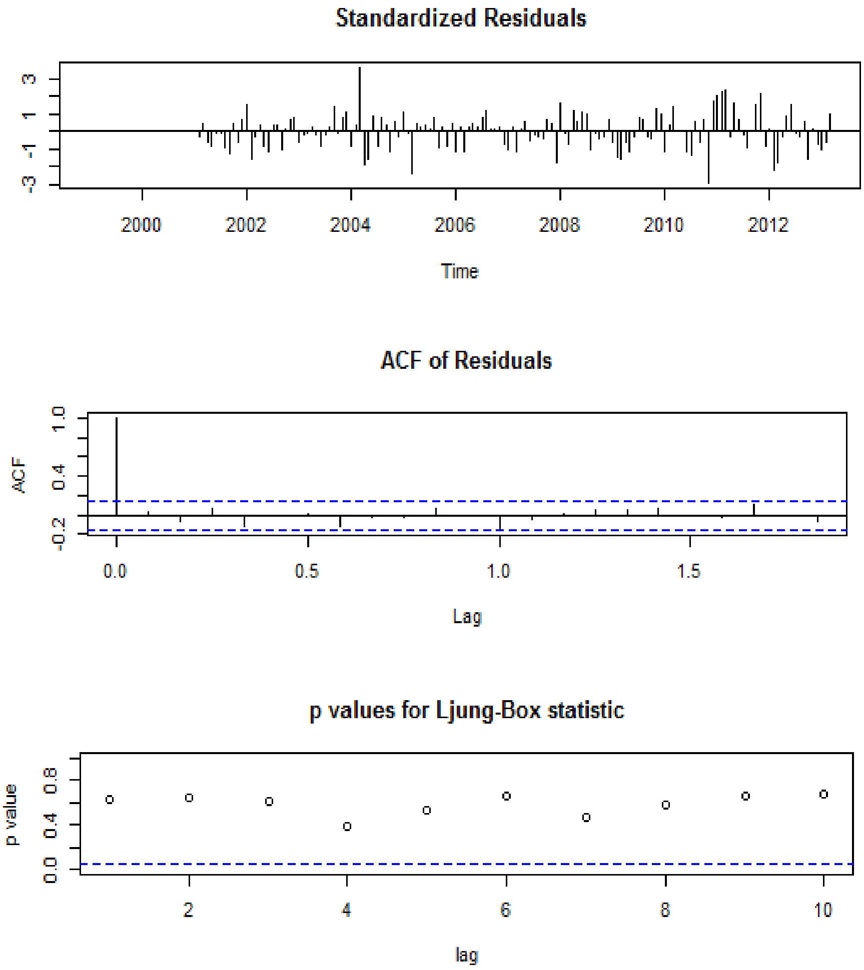

Figure 5: Plots for residues

As we had observed a few outliers in the series, we tried to fit a time series model by removing the outliers. These outliers in the series are replaced by local average. The local average is calculated by considering the data corresponding to a particular month for which the outlier is detected over a period of time. After the data cleaning, time series analysis is made to identify the best model which has minimum information criteria. This model is compared with *ARIMA (0,1,1)(1,2,1)*and observed that *ARIMA (0,1,1)(1,2,1)* is having the least information criteria. Hence the model *ARIMA (0,1,1)(1,2,1)* is used for forecasting the export of leather for the year 2013-2014 and is given in Table 2.

ISSN 2249-5908
Issue 3, Vol.6 (November-2013)

Table 2: Forecast of Leather export during the year 2013-2014.

Forecast	Export Value in Rs. Million	95% confidence interval	
		Lower Limit	Upper Limit
Apr-13	19957.06	17068.35	22845.76
May-13	24345.56	21217.86	27473.27
Jun-13	27268.70	23919.01	30618.39
Jul-13	28058.05	24500.20	31615.90
Aug-13	26701.17	22946.68	30455.66
Sep-13	23873.00	19931.67	27814.33
Oct-13	23718.56	19598.85	27838.26
Nov-13	23307.18	19016.51	27597.85
Dec-13	26873.55	22418.47	31328.63
Jan-14	28179.76	23566.41	32793.11
Feb-14	25681.65	20915.85	30447.45
Mar-14	26778.93	21865.40	31692.45

The exporters always wish to have updated information on the trends in exports. But usually the export information is available three to five months prior to the date. Hence, this forecast is useful to the leather industry to know the trends in leather industry and thereby take necessary strategy for their business.

4. Conclusion

In this article we identified the SARIMA model, *ARIMA (0,1,1)(1,2,1)* as a suitable model for forecasting the overall leather export from India. The model is arrived at based on the information criteria like, AIC and BIC. Among all the identified models, *ARIMA (0,1,1)(1,2,1)* model has the least information criteria. Also statistical test for goodness of fit and Ljung-Box test for independence is also carried out for confirming the significance of the model. As *ARIMA(0,1,1)(1,2,1)*model satisfies all the criteria, this model is made use for forecasting the total leather export value for the period 2013-2014.

Acknowledgements:

The authors are thankful to the Council of Leather Exports (CLE), government of India for the financial aid (CNP – 1098) and CSIR- Central Leather Research Institute, Chennai for the support to carry out the study. The authors are also thankful to Dr. Sudheesh Kumar Kattumanil for discussions and suggestions throughout the preparation of the article. Also we acknowledge the CSIR-CLRI Communication No. 1001.

Reference

1. Box, G.E.P. and Jenkins, G.M. (1970). *Time Series Analysis*, *Forecasting and Control*, Oakland, C A: Holden-Day.
2. Brockwell, P. J. and Davis, R. A., (2002). *Introduction to Time Series and Forecasting*, SecondEdition, Springer, New York.
3. Burnham, K. P. and David, R. A. (1998). *Model Selection and Inference: A Practical Information-Theoretical Approach*, New York: Springer-Verlag.
4. Dickey, D. A. and Fuller, W. A. (1978). Distribution of the Estimators of the Autoregressive Time Series with a Unit Root, *Journal of American Statistical Association*, 74 (366), 427-431.
5. Imhoff M., Bauer M., Gather U., Löhlein D. (1997). Time series analysis in intensive care medicine, *Applied Cardiopulmonary Pathophysiology*, 6, 263-281.
6. Schwarz, G. (1978). Estimating the Dimension of a Model, *Annals of Statistics*, 6, 461-464.
7. Shumway, R. H. and Stoffer, D. S. (2006). *Time Series Analysis and its Applications: With R Examples*, Second Edition, Springer, New York.
8. Tatiporn, P. And Kanchana, K. (2012). Forecasting Model for Para Rubber’s Export Sales, *KMITL Science and Technology Journal*, 12 (2), 198 – 202.

ISSN 2249-5908
Issue 3, Vol.6 (November-2013)

9. Venugopalan, R. and Prajneshu (1996). Trend Analysis in all- India Marine Products Export using Statistical Modelling Techniques, *Indian Journal of Fisheries*, 43 (2), 107 – 113.