Indian Journal of Extension Education
Vol. 57, No. 2 (April-June), 2021, (131-134)

ISSN 0537-1996 (Print)
ISSN 2454-552X (Online)

Research Note

Box-Jenkins ARIMA Approach to Predict FDI Inflow in India

Megha Goyal[1], Suman Ghalawat[1], Amita Girdhar[1], Nitin Agarwal[2] and Joginder Singh Malik[3]*

ABSTRACT

The study attempts to model and predict the FDI inflows in India through use of time series data for FDI inflow in India from 1980 to 2019. The Autoregressive integrated moving average (ARIMA) model developed by Box and Jenkins (1976) was used to develop the model. UBJ identification included the determination of appropriate AR (autoregressive) and MA (moving-average) polynomials orders i.e. values of p and q. Orders were determined from the autocorrelation functions and partial autocorrelation functions of the stationary series. FDI data were found to be non stationary and a single order differencing was sufficient to obtain the appropriate stationary series. The study determined a low AIC value and subsequently introduced the ARIMA model (1,1,0) as a suitable FDI predictor model in India. The promised FDI inflows for the years 2016-17 to 2018-19 were within the scope of confidence and the percentage deviation of predictable and observable values ensures that our predicted prices close to real prices.

Keywords: ARIMA, Forecast, Box and Jenkins, Stationary, Foreign direct investment (FDI)

INTRODUCTION

Foreign Direct Investment (FDI) is an investment that involves long-term relationships, interests and the influence of a single investor (foreign investors / parent business) in a foreign-based business other than a foreign investor and plays an important role in developed and developing countries. India after liberating and growing the global economy in 1991, witnessed a significant increase in FDI flows. It has played a key role in the development process over the past two decades. To a large extent, FDI is a non-performing source of additional foreign investment. At a minimum, FDI is expected to improve outreach, technology, skills levels, employment and linkages with other sectors and regions of the economy. Maggon (2012) reviewed FDI's economic policy policies and suggested possible improvements to the current policy framework. Thabani Nyoni (2018) used ARIMA to predict FDI entry in number in Zimbabwe. Anupama and Rupashree (2019) applied ARIMA model in R to software to forecast FDI inflow in INDIA. Mohammed Ershad Hussain, Mahfuzul Haque (2016) analyzed about Foreign Direct Investment, Trade, and Economic Growth of Bangladesh. Verma (2015) described Regression, ARIMA and ARIMAX analyses on GDP, inflation, exchange rate, export, import, energy generation and trade balance to estimate the foreign direct investment (FDI) in India. Time series data (TS) refers to the dynamics that occur over time. The TS movement of the chronological data can be solved by the tendency, periodic (say, seasonal), rotation and irregularities. One or two of these items may cover the rest of the series. The basic assumption in any TS analysis is that some aspects of the previous pattern will continue to remain in the future. The most widely used method of modeling and predicting TS data is the

[1]Assistant Professor, Department of Business Management, CCS Haryana Agricultural University, Hisar, Haryana
[2]Research scholar, Guru Jambheshwer University, Hisar, Haryana
[3]Professor, Department of Extension Education, College of Agriculture, CCSHAU, Hisar, Haryana
*Corresponding author email id: jsmalik67@gmail.com

Box-Jenkins's Autoregressive integrated moving average (ARIMA) methodology.

METHODOLOGY

Data for the 39-year Foreign Direct Investment (FDI) period series from 1980-81 to 2018-19 were collected from various issues of RBI bulletins. The statistical analysis was performed to improve the relevant relationship by the Autoregressive integrated moving average (ARIMA) of the FDI forecast in India. Data from 1980-2016 were used for model development and data for the next three years were used to assess model validity.

The univariate ARIMA approach was first popularized by Box and Jenkins and the models developed through this approach are referred to as univariate Box-Jenkins (UBJ) models. The ARIMA model building process is executed in four steps; Identification of the autoregressive integrated moving average model where the order is (p,d,q); Estimation of the coefficients; A test is constructed on the residuals projected and the model is subjected to a set of Diagnostic Testing and Forecasting the future from given set of data. The general functional form of ARIMA (p,d,q) model is :

$$\phi_p(B)\Delta^d y_t = c + \theta_q(B)a_t \qquad \text{... (1)}$$

where, y = Variable under forecasting, B = Lag operator, a = Error term (Y- $\hat{Y}$, where $\hat{Y}$ is the estimated value of Y), t = time subscript, ϕ_p (B) = non-seasonal AR i.e. the autoregressive operator, represented as a polynomial in the back shift operator, Δ^d = non-seasonal difference

$\theta_q(B)$ = non-seasonal MA i.e. the moving-average operator, represented as a polynomial in the back shift operator ϕ's and θ'_s are the parameters to be estimated

Here the order of the AR model and MA model can be expressed through p and q respectively. The number of time series differences is expressed through d where d, p and q are all sets of integers, show the projected residual for each period of time. For ideal circumstances, the model should be impartially distributed as random normal set of observations.

Initial stage in making an autoregressive integrated moving average model in equation (1) is to find the stationarity of the time series data of observations. To obtain the stationarity, at the primary step the ACF or the Auto Correlation Function and the PACF or the Partial Auto Correlation Function of the given time series data sets are to be drawn, if the series are not stationary then proper differenced order is required to make the series under consideration stationary. Using Correlogram analysis, the p and q of the model is to be fitted, this is based on iterative process. To test for the Goodness of Fit, the BIC alternatively speaking the Bayesian Information criteria is examined here. For autoregressive model, AR (p) the Auto Correlation Function is tailing off at the level p but the Partial Auto Correlation Function cuts off. For moving average model, MA (q), the Auto Correlation Function cuts off whereas the Partial Auto Correlation Function is tailing off in the order of q. Moreover for Autoregressive Moving Average, ARMA (p, q) neither of the Auto Correlation Function or the Partial Auto Correlation Function is tailing off. The standard model developed in the equation (1) needs to be assessed through iterative method until the sum of the squares of the residual in its least is obtained. The appropriateness of the constructed model can be tested through analytical diagnostic examination. This contains the process of scrutinizing the residuals from the model thus fitted to inspect if there exists the indication of non-randomness. Here from the residuals correlogram is calculated, it is found out to what extent there is significant difference from zero among the coefficients. To test for the randomness of the model's residuals the Ljung Box Statistic (Ljung and Box, 1978) is utilized. The formulation for the Ljung Box-statisticis

$$Q = N(N+2)\sum_{k=1}^{K}\frac{\hat{\rho}_k^2}{N-k} \qquad \text{... (2)}$$

Here N indicates the number of the observations, the autocorrelation order of lag is denoted by $\hat{\rho}_k$. The statistics Q approximately follows a chi-square distribution with (K-m) degree of freedom, where m is number of parameters estimated in ARIMA model. If Q is large it says that the residual autocorrelations as a set are significantly different from zero and random shocks of estimated model are probably auto-correlated, then reformulate the model. Finally the ARIMA model is used to forecast results, with upper and lower limits, these limits give us the confidence interval. RMSE or the root mean square is used to identify the measurement errors, this helps us to identify the robustness of the model.

$$\text{RMSE}=\sqrt{\frac{1}{N}\sum_{t=1}^{N}\left(y_t - \hat{y_t}\right)^2} \quad \text{... (3)}$$

Equation (3) shows the expression of the RMSE or the root mean square. Where y_t is actual observation and $\hat{y_t}$ is estimated observation.

Table 1: Autocorrelations : FDI

Lag	Auto-correlation	Std. error (a)	Box-Ljung Statistic Value	df	Sig.(b)
1	.899	.160	33.998	1	.000
2	.796	.259	61.412	2	.000
3	.689	.316	82.501	3	.000
4	.594	.352	98.613	4	.000
5	.518	.377	111.212	5	.000
6	.462	.395	121.567	6	.000
7	.417	.408	130.264	7	.000
8	.327	.419	135.775	8	.000
9	.257	.426	139.295	9	.000
10	.167	.430	140.838	10	.000
11	.073	.431	141.143	11	.000
12	-.026	.432	141.182	12	.000
13	-.093	.432	141.710	13	.000
14	-.122	.432	142.660	14	.000
15	-.146	.433	144.081	15	.000
16	-.166	.434	145.994	16	.000

RESULTS AND DISCUSSION

The analysis was performed on the data of the time series from 1980-81 to 2018-19 of FDI. The ARIMA method was used to estimate FDI in India. UBJ identification includes the determination of appropriate AR and MA polynomials orders i.e. values of p and q. Orders are determined from the autocorrelation functions and partial autocorrelation functions of the stationary series. FDI data were found to be non stationary and a single order differencing was sufficient to obtain the appropriate stationary series. The estimated acf is shown in Table 1. After experimenting with different lags of the moving average and autoregressive processes, ARIMA (1,1,0) was introduced to estimate FDI in India. The Marquardt algorithm (1963) was used to minimize the sum of squared residuals. Log Likelihood, Akaike's Information Criterion, AIC (1969), Schwarz's Bayesian Criterion, SBC (1978) and residual variance decided the criteria to estimate AR and MA coefficients in the model. Parameter estimates of the fitted models are given in Table 2.

The fitted model ARIMA (1,1,0) may be elaborated as:

$$(1-\phi_1 B)\,(1-B)\,Y_t = a_t$$

$$Y_t - (1 + \phi_1)\,BY_t + \phi_1 B^2\,Y_t = a_t$$

$$Y_t=(1 + \phi_1)Y_{t-1}-\phi_1 Y_{t-2} + a_t$$

The ARIMA (1,1,0) fitted model indicates the presence of lagged values of dependent variable i.e. autoregressive component. The residual acf along with the associated 't' tests and Chi-squared test suggested

Table 2: Parameter estimates of ARIMA model

		Estimates	Std. error	t	Approx sig.
Non-Seasonal Lags	AR1	.083	.183	.453	.654
Constant		1603.759	883.74	1.815	.079

Melard's algorithm was used for estimation.

Table 3: Diagnostic checking of residuals autocorrelations: FDI

Model	Number of Predictors	Model Fit statistics R-squared	RMSE	MAPE	Normalized BIC	Ljung-Box Q Statistics	Sig.
FDI (1,1,0)	0	.926	4797.53	1000.89	17.15	7.09	.982

Table 4: FDI estimates along with upper and lower confidence limits based on ARIMA model

Models	FDI (1,1,0) (million US $) 2016-17	2017-18	2018-19
Forecast	57893.8	59558.3	61167.1
UCL	67654.4	73946.2	79060.6
LCL	48133.3	45170.5	43273.6

UCL & LCL - Upper and lower confidence limits (95%)

Table 5: Percent deviation between Actual & fitted values India's FDI from 2016-17 to 2018-19. (Value in Million USD)

Year	FDI Actual (Y)	FDI Forecast (F)	Relative Deviations
2016-17	60220	57893.8	-3.86
2017-18	60974	59558.3	-2.32
2018-19	64375	61167.1	-4.98

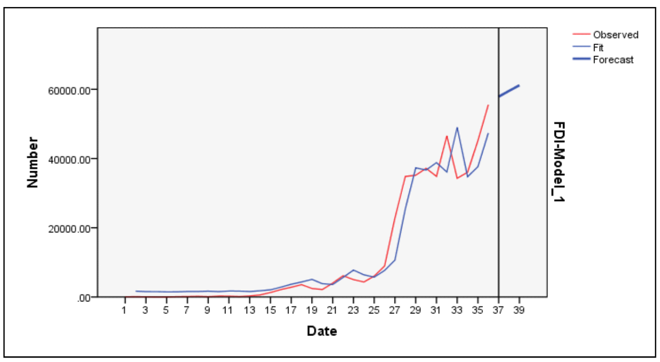

Figure 1:

by Ljung and Box (1978) were used for the checking of random shocks to be white noise (Table 3). The observed and forecasted values of FDI along with lower and upper confidence limits and are shown in Table 4.

$$\text{Percent deviation (RD \%)} = \frac{\text{Forecasted yield} - \text{Actual yield}}{\text{Actual yield}} \text{ X } 100$$

ARIMA model could be used successfully for modelling as well as forecasting of yearly FDI of India. It has been found that there is a significant increasing trend in FDI of India. The forecast values of FDI during 2016-17 to 2018-19 are close to the actual valuesas percent deviation of the forecasted and observed figures is in acceptable limits shown in Table 5. The level of accuracy achieved by ARIMA (1,1,0) was found adequate for estimating FDI and residuals were white noise. Three-step ahead (out-of-model development period i.e. 2016-17, 2017-18 and 2018-19) forecasted values of FDI along with their estimated and observed values over the years are shown in Figure 1.

Paper received on : February 27, 2021
Accepted on : March 30, 2021

REFERENCES

Akaike, H. (1969). Fitting autoregressive models for prediction, *Annals of the Institute of Statistical Mathematics,* **21**, 243-47.

Anupama, G. and Rupashree, R. (2019). Forecast of Foreign Direct Investment Inflow (2019-2023) with reference to Indian Economy, *International Journal of Research in Engineering, IT and Social Sciences*, **9**(4), 22-26.

Box, G.E.P. and Jenkins, G.M. (1976). Time series analysis: Forecasting and control. Holden Day, San Franscisco: 575p.

Hussain, M.E. and Haque, M. (2016). Foreign Direct Investment, Trade, and Economic Growth: An Empirical Analysis of Bangladesh, *Economies*, **4**(2), 7. doi:10.3390/ economies 4020007.

Ljung, G.M. and Box, G.E.P. (1978). On a measure of lack of fit in time series models, *Biometrika,* **65**, 297-303.

Maggon, M. (2012). Economic and policy determinants of FDI, an empirical analysis in context of India, *Vivekananda Journal of Research,* **1**(1), 13-23.

Marquardt, D.W. (1963). An algorithm for least-squares estimation of non-linear parameters. *Journal of the Society for Industrial and Applied Mathematics,* **2**, 431-41.

Schwarz, G. (1978). Estimating the dimension of a model, *The annals of Statistics,* **62**, 461-464.

Thabani, N. (2018). Box-Jenkins ARIMA approach to predicting net FDI inflows in Zimbabwe, *Munich Personal RePEc Archive*, 87737p.

Verma, P. (2015). Regression, ARIMA and ARIMAX Models to Study the Factors affecting Foreign Direct Investment in India, *Management Asian Journal of Research in Business Economics and Management,* **5**(12), 1-10.