Modeling and Forecasting Foreign Direct Investments in India

** Bulbul Singh*

Abstract

For a developing country like India, FDI can be an important source of finance, and it can contribute richly to the economy by transferring superior managerial skills and state- of- the- art technology into the country. Ever since India undertook economic reforms in 1991, it has been one of the largest recipients of FDI in the world, making it a country of huge opportunities. This paper attempted to forecast FDI in India from 2014 to 2020 using univariate ARIMA modelling. Since FDI can have an impact on other macroeconomic variables such as GDP and exports, an accurate forecasting can be valuable for policy making. Applying the Box- Jenkins methodology, the process of model identification, estimation, diagnosis, and forecasting were undertaken. As many as eight different ARIMA models were estimated from which one was short-listed after an iterative process. Several accuracy tests were used and after confirmation of white noise in residuals, that model was eventually selected, which had the least forecasting error and biasness. ARIMA Model (1,1,1) was found to be most suitable and provided the tightest fit to the data. As per the forecasted model, India can potentially receive FDI up to US $ 74,935.27 in 2020, and the average receipts over the forecasted period could be US $ 51982.39 million. The compounded annual growth rate (CAGR) of FDI inflows between the forecasted period of 2014 and 2020 is likely to be 14.31%.

Keywords: **foreign direct investment, ARIMA, forecasting, Box Jenkins methodology**

JEL Classification: C220,F170, F210, F230

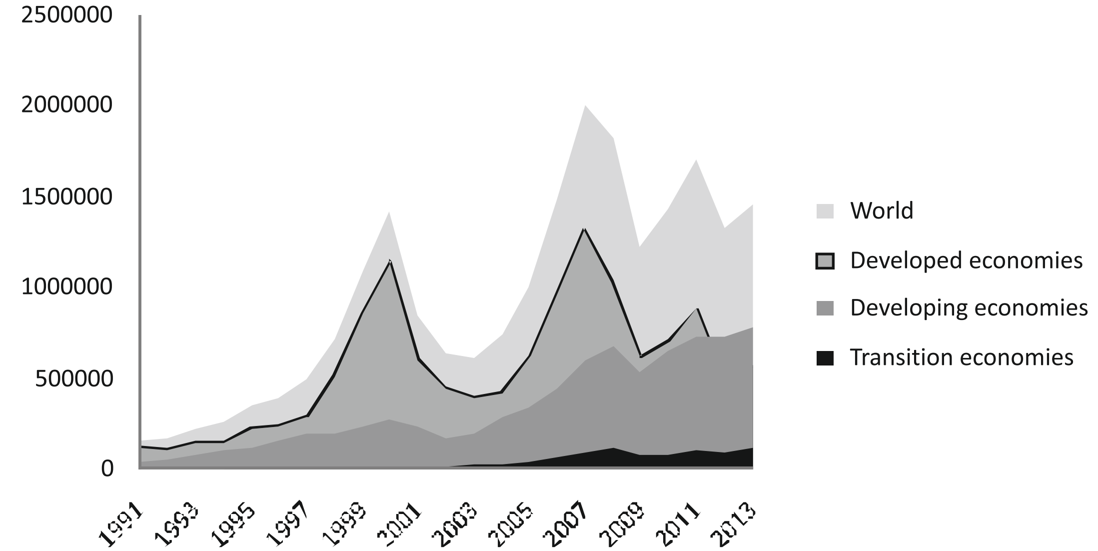
Paper Submission Date : January 27, 2015 ; Paper sent back for Revision : April 10, 2015 ; Paper Acceptance Date : April 25, 2015

Foreign direct investment (FDI) is defined as an investment involving a long-term relationship made by a company or entity based in one country, into a company or entity based in another country. Entities making direct investments typically have a significant degree of influence and control over the company into which the investment is made. The world economy in the last two decades has witnessed a fast growth in FDI flows. Between 1991 and 2013, the inward FDI flows in the world grew almost ten times - from US $155,366 million to US $ 1,451,965 million (UNCTAD, 2013 ; see Figure 1).

The subject of FDI has engaged particular interest and discussion in the contemporary economic literature. There are two reasons for this intensified attention. First, more and more nations of the world have begun liberalizing their domestic policies and have started to adapt to open market practices due to which there has been a significant increase in the volume of FDI flows between nations. Secondly, from the standpoint of developing nations, FDI has become an important source of finance. FDI can add to the investible resources and capital formation of a country. At the same time, it is also a means of transferring production technology, skills, innovative capacity, organizational and managerial practices from technologically advanced transnational corporations to companies in developing countries. The investing enterprise, in turn, gains by getting an access to a wider international market and cheaper resources. It can be observed from the Figure 1 that the FDI received by developing nations has now become larger than inflows in developed nations since 2011. Developing nations

* *Assistant Professor,* Sri Aurobindo College of Commerce and Management, Village Jhande, P.O. Threeke, Ferozepur Road, Ludhiana- 141 021, Punjab. E-mail : bulbul246@gmail.com

Figure 1. Inward Foreign Direct Investment Flows Between 1991 and 2013 (in US $ in Millions)



Source: Compiled from UNCTAD (2014)

with liberalized economic systems, skilled workforces, and good growth prospects are attracting larger amounts of FDI in recent years than the saturated developed nations. India is one of the largest recipients of FDI amongst developing nations. In 2013, India received US $28.2 billion in FDI, which accounted for more than four fifth of FDI received by South Asian nations (UNCTAD, 2014). According to UNCTAD's World Investment Prospects Survey (UNCTAD, 2010), India is ranked second in the top priority host economies for FDI for the period from 2010-2012. China was at the top position, and India was rated ahead of some of the developed nations such as United States of America, United Kingdom, Germany, Japan, and Australia.

Foreign Direct Investment Inflows in India

India initiated its economic reform process in the year 1991 (Ahluwalia, 2002). The liberalization, privatization, and globalization program saw the dismantling of the licensing system. India opened up the economy to foreign investment and put major thrust on building ports, highways, oil and gas industries, power generation and telecommunication sectors. It also undertook reforms to gradually open the consumer goods and services sector. FDI can be made in India in two ways - the automatic route and the approval route. Areas of priority interest are allowed 100% equity participation through the automatic route ; whereas, a more cautious approach is undertaken for areas which are more critical for protecting domestic interests. In these sensitive sectors, India allows foreign investments to come in a phased manner and has created slabs such as 26%, 49%, 51%, and 74% foreign equity investment. The reform process has led to a gradual increase in the FDI inflows in the country, particularly in the last decade (see Table 1). India is one of the fastest-growing economies of the world, and several transnational corporations have invested in India so that they can be a part of its growth story.

The growth progression in India has been through balanced and prudent government policies and regulations. To India's advantage, it has a highly educated, young population and offers cheap and abundant labour. This has led to India becoming an outsourcing hub, and several corporations have set up their research and development centres in the country. Instead of following a knee- jerk approach, India has undertaken a gradual liberalization path. More and more sectors are put on the automatic route with the passage of time, and the equity stipulations are relaxed in a phased manner. Apart from its educated workforce, India is also an attractive host nation for FDI

Table 1. Foreign Direct Inflows in India from 1991-2013 (in millions of US $)

Year	FDI INFLOWS	YEAR	FDI INFLOWS
1991	75	2002	5629.671
1992	252	2003	4321.076
1993	532	2004	5777.807
1994	974	2005	7621.769
1995	2151	2006	20327.76
1996	2525	2007	25349.89
1997	3619	2008	47138.73
1998	2633	2009	35657.25
1999	2168	2010	27431.23
2000	3587.989747	2011	36190.4
2001	5477.637624	2012	24195.77
		2013	28199.45

Source: UNCTAD Stat, 2014

due to its better judicial system, English language proficiency, fewer cultural barriers, and better institutional set-up as compared to several other developing countries.

In this paper, an attempt has been made to forecast foreign direct investment inflows in India upto the year 2020 by using the auto regressive integrated moving average (ARIMA) modeling. To the best of my knowledge, such a study has not been attempted in India so far. An accurate forecast can help economists in making better plans. Based on these estimates, planners can identify further need for opening up restricted sectors. A precise and valid forecast can give policy makers a framework for setting targets for the future, for example; they can strategize on what would be an appropriate time for relaxing equity stipulations in sectors where they have been imposed. Since FDI can affect other macroeconomic variables such as gross domestic product, exports, and employment, an assessment of prospective future trends is vital.

Review of Literature

A. T. Kearney (2011) constructed the FDI Confidence Index by using primary data from a proprietary survey administered to senior executives in world's leading corporations in 27 countries spanning 17 industries. The survey reported that investors are turning to emerging economies like China, India, and Brazil for their large, fast growing consumer markets than their cheap labour markets. In the survey, 60% of the respondents anticipated higher taxation rates in developed markets and expected increased labour regulations in developing nations as they integrated further into global markets. UNCTAD (2010) anticipated that global FDI inflows will rise and be in the range of US $ 1.6 to 2 trillion by 2012. It further reported that four major emerging markets - China, India, Brazil, and the Russian Federation - all ranked among the top five investment destinations, and developing nations in Asia were accorded higher priority for investments by respondents as compared to developed countries.

The Financial Times (2012) used the database fDiMarkets which tracks the global greenfield investment projects and reported that renewable energy was the fastest growing sector for FDI in 2011, and almost one in five FDI projects were expansion projects with ventures in manufacturing, extraction, and front and back office being principally important. Bashier and Talal (2007) used the Box- Jenkins methodology of building an ARIMA model to forecast FDI inflows in Jordon for the period from 2004- 2025. The forecasting results revealed an increasing trend over the forecasted period, with inflows likely to reach 3214.87 million Jordanian Dinar (JD) by 2025.

Turolla and Margarido (2011) also used the ARIMA model and detected a strong moving average pattern in FDI inflows of Brazil. They used monthly data between January 1995 and November 2009 and forecasted FDI inflows for 13 months up to December 2010. Their findings revealed a dynamic series for Brazil with relatively rapid adjustments towards equilibrium values.

In another forecast for Jordan's FDI inflows, Al-Rawashdeh, Nsour, and Salameh (2011) used the ARIMA model and forecasted that up to the year 2030, FDI inflows totaling 29207.06 million JD could be received by Jordan, with an average annual growth rate of 3.22%. Like Bashier and Talal (2007), Al-Rawashdeh et al. (2011) suggested a strong moving average component (0, 1, 1) in Jordanian FDI inflows.

Cheng and Ma (2007) studied the past and future outward FDI from China. They used the Gravity model and forecasted a strong upward rise in China's outward FDI based on its own past trend as well as based on experiences of its neighbouring countries, Japan and South Korea. Kumar and Dhingra (2012) studied the growth performance and forecasted FDI inflows in Sri Lanka using double exponential smoothing and Ljung Box Q statistics. They forecasted that Sri Lankan FDI inflows would rise up to US $698.02 million by the year 2020 with a compounded annual growth rate of 3.83%. This is slower than other South Asian economies which they forecasted would have a CAGR of 4.8% in the same period.

Data and Methodology

Data

In this research, an endeavor has been made to forecast FDI inflows in India from 2014- 2020. Data for FDI inflows has been sourced from UNCTAD Stat and has been shown in the Table 1. The data collected covers the time period from 1991 to 2013. This time period was considered suitable because India experienced structural breaks in 1990-91 due to opening up of the economy, and using data prior to 1991 would have led to an inaccurate modeling. The data was transformed into natural logarithm for time series processing. Data processing was carried out in the econometric package EViews 7.

Methodology

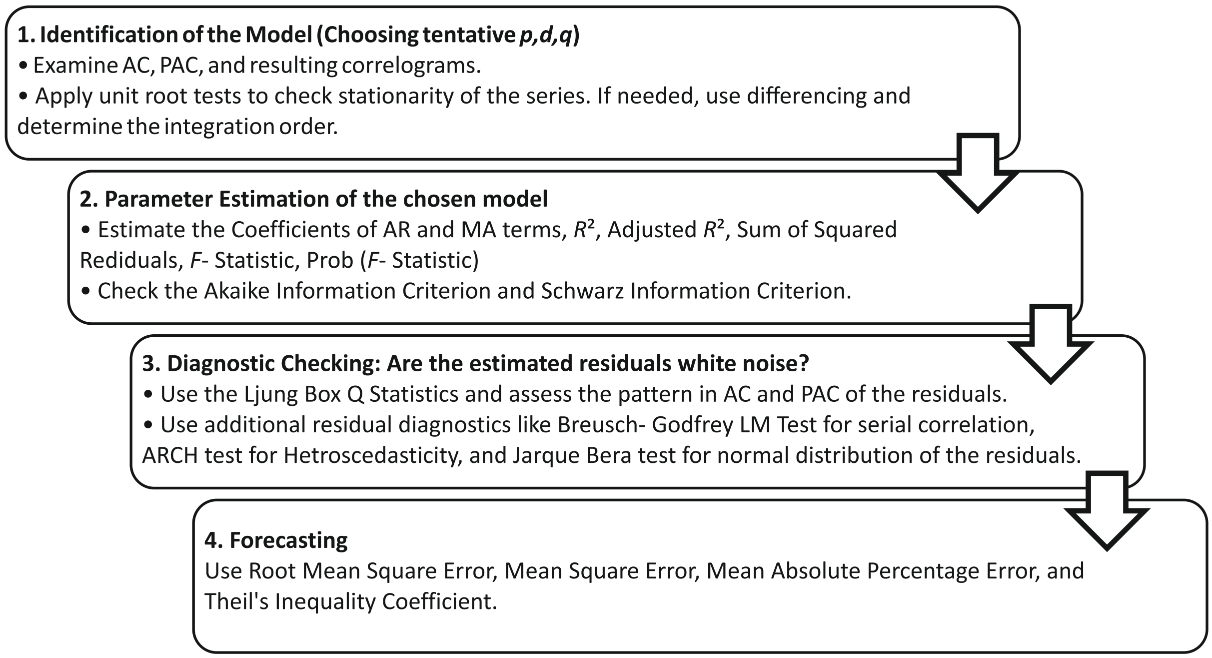
This paper uses the univariate ARIMA modeling by applying the Box- Jenkins approach for forecasting FDI inflows in India. The ARIMA model does not involve the construction of an equation, but uses the probabilistic or stochastic properties of economic time series on their own under the philosophy 'let the data speak for itself'. Unlike the regression models in which Y_t is explained by k regressors $X_1, X_2, X_3, ... X_k$, the Box- Jenkins type time series models allow Y_t to be explained by past or lagged values, values of Y itself, and stochastic error terms.

➪ **The ARIMA Theory :** ARIMA (autoregressive integrated moving average) models are generalizations of the simple AR model that uses three tools for modeling the serial correlation in the disturbance (Schwert 2009). The first tool is the autoregressive, or AR, term. The AR(1) model uses only the first order term, but in general, higher-order AR terms may be used. Each AR term corresponds to the use of a lagged value of the residual in the forecasting equation for the unconditional residual. An autoregressive model of order p, AR*(p)* has the form :

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots\ldots + \rho_p u_{t-p} + \varepsilon_t$$

The second tool is the integration order term or *I(d)*. Each integration order corresponds to differencing the series being forecast. A first- order integrated component means that the forecasting model is designed for the first difference of the original series. A second- order component corresponds to using second differences, and so on.

Figure 2. The ARIMA Process Adopted



The third tool is the MA, or moving average term. A moving average forecasting model uses lagged values of the forecast error to improve the current forecast. A first- order moving average term uses the most recent forecast error, a second- order term uses the forecast error from the two most recent periods, and so on. An MA*(q)* has the form:

$$u_t = \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots\ldots + \theta_p\varepsilon_{t-p}$$

The autoregressive and moving average specifications can be combined to form an ARMA *(p,q)* specification:

$$u_t = \rho_1 u_{t-1} + \rho_2 u_{t-2} + \ldots\ldots + \rho_p u_{t-p} + \varepsilon_t + \theta_1\varepsilon_{t-1} + \theta_2\varepsilon_{t-2} + \ldots\ldots + \theta_p\varepsilon_{t-p}$$

Although ARIMA is typically applied to residuals from a regression model, the specification can also be applied directly to a series. This latter approach provides a univariate model, specifying the conditional mean of the series as a constant and measuring the residuals as differences of the series from its mean.

➯ **The Procedure:** The ARIMA modeling of forecasting involves using a combination of the three building blocks *(p, d, q)* described above. The procedure followed consists of four steps which are shown in the Figure 2.

(1) Identification of the ARIMA Model : For identification of the ARIMA terms *(p,d,q)*, we shall make use of correlograms. The correlograms give information about autocorrelations, partial autocorrelations, and Q statistics along with its probability values.

(a) Autocorrelations (AC) : The autocorrelation of a series Y at lag k is estimated by :

$$\tau_k = \frac{\Sigma^T_{t=k+1} (Y_t - \overline{Y})(Y_{t-k} - \overline{Y})}{\Sigma^T_{t=1} (Y_t - \overline{Y})^2}$$

where,
$\overline{Y}$ is the sample mean of Y (Schwert, 2009). This is the correlation coefficient for values of the series k periods apart. If τ_k is non- zero, it means that the series is first order serially correlated. If τ_k dies off more or less geometrically with increasing lag k, it is a sign that the series obeys a low- order autoregressive (AR) process. If τ_k has a significant spike and drops to zero after a small number of lags, it is a sign that the series obeys a low- order moving average (MA) process. The dotted lines in the plots of the autocorrelations are the approximate two standard error bounds computed as ± $2(\sqrt{T})$. If the autocorrelations are within these bounds, these are not significantly different from zero at approximately 5% significance level.

(b) Partial Autocorrelations (PAC) : The partial autocorrelation at lag k is the regression coefficient on Y_{t-k} when Y_t is regressed on a constant, Y_{t-1}, ..., Y_{t-k} . This is a partial correlation since it measures the correlation of Y values that are k periods apart after removing the correlation from the intervening lags. If the pattern of autocorrelation is one that can be captured by an autoregression of order less than k, then the partial autocorrelation at lag k will be chosen to zero.

The PAC of a pure autoregressive process of order p, AR*(p)*, cuts off at lag p, while the PAC of a pure moving average (MA) process asymptotes gradually to zero. The PAC is estimated at lag k recursively by :

$$\varphi_k = \frac{\tau_k - \Sigma^{k-1}_{j=1} \varphi_{k-1,j}\tau_{k-j}}{1 - \Sigma^{k-1}_{j=1} \varphi_{k-1,j}\tau_{k-j}} \quad \text{for } k > 1$$

where,
τ_k is the estimated autocorrelation at lag k and where,

$$\varphi_{k,\mathrm{j}} = \varphi_{k-1,\mathrm{j}} - \varphi_k \varphi_{k-1,k-j}$$

The dotted lines in the plots of the autocorrelations are the approximate two standard error bounds computed as ± 2 ($\sqrt{T}$). If the partial autocorrelations are within these bounds, it is not significantly different from zero at approximately 5% significance level. If the autocorrelations are zero after one lag, then a first- order autoregressive model is appropriate. Alternatively, if the autocorrelations are zero after one lag and the partial autocorrelations declined geometrically, a first order moving average process would seem appropriate.

(c) Q- Statistics : The last two columns reported in the correlogram are the Ljung- Box Q- statistics and their p-values. The Q- statistics at lag k is a test for the null hypothesis that there is no autocorrelation up to order k and is computed as:

$$Q_{LB} = T(T+2)\Sigma^{k}_{j=1} \frac{\tau^2_j}{T - J}$$

Where τ_j is the j-th autocorrelation and T is the number of observations. The Q statistic is used as a test of whether the series is white noise.

(d) Stationarity Testing : A series is said to be stationary if the mean and autocovariances of the series do not depend on time (Gujarati, Porter, & Gunasekar, 2009). Stationarity of data is essential in time series analysis as otherwise, it can lead to spurious results. Though several procedures are available to test for unit roots (non-stationary), this research has used the Augmented Dickey Fuller test (ADF) and the Phillips- Perron unit root tests (PP). The ADF test relies on rejecting a null hypothesis of unit root in favour of the alternative hypothesis of stationarity. The general form ADF is estimated by the equation (1):

$$\Delta Y_t = a_0 + Z_t + a_1 Y_{t-1} + \Sigma^{p}_{i=1} a_i \Delta Y_{t-1} + \varepsilon_t$$

where,

a_0 is a constant, t is a deterministic trend, and enough lagged differences are included to ensure that the error term becomes white noise. If the autoregressive representation of Y_t contains a unit root, the t- ratio for a_1 should be consistent with the hypothesis $a_1 = 0$.

The PP unit root tests differ from ADF tests mainly in how they deal with serial correlation and heteroscedasticity in the error terms. The test regression in Phillips- Perron test is :

$$\Delta Y_t = \varphi Y_{t-1} + \alpha + \beta t + u_t$$

where,
u_t is a stationary process (which may also be heteroscedastic). An advantage of the PP test over the ADF tests is that the PP tests are robust to general forms of heteroscedasticity in the error terms. Another advantage is that the user does not have to specify a lag length for the test regression.

If a series is stationary without any differencing, it is designated as I (0) or integrated of order 0. On the other hand, a series that has stationary first difference, it is designated as *I (I)* or integrated of order (1).

(2) Estimating ARIMA Models : To specify the ARIMA model, we shall difference the FDI series, if necessary, to account for the order of integration. Next, we shall describe the structural regression model by adding AR and MA terms. An iterative process will be used to test different models. The goal of ARIMA analysis is to get a parsimonious representation, and various parameters will be inspected to get the best fitting model. We shall estimate the coefficients of AR and MA terms, R^2, Adjusted R^2, F- Statistic, Prob (F- Statistic) and check the models with low values of the Akaike information criterion and Schwarz information criterion.

(3) Diagnostic Checking : To check if the model is a reasonable fit to the data, we shall apply a number of diagnostic checks, which are listed as follows:

(a) Correlograms : The AC and PAC of the equation residuals are checked as per the Ljung- Box Q Statistic and its p - values. If the residuals are white noise, the null hypothesis that there is no serial correlation should be accepted.

(b) Normal Distribution of Residuals : To check the normal distribution of the residuals, we apply the Jarque-Bera statistic. This test statistic measures the difference of the skewness and kurtosis of the series with those from normal distribution. The statistic is computed as:

Jarque- Bera = $\frac{N}{6}\left(S^2 + \frac{(K-3)^2}{4}\right)$, where S is the skewness and K is the kurtosis. Under the null hypothesis of a normal distribution, the Jarque- Bera statistic is distributed as χ^2 with 2 degrees of freedom. The reported probability is the probability that a Jarque- Bera statistic exceeds the observed value under the null hypothesis - a small probability value leads to the rejection of the null hypothesis of a normal distribution.

(c) Serial Correlation LM Test : In a time series, the residuals are found to be correlated with their own lagged values. This serial correlation violates the standard assumption of regression theory that the disturbances are not correlated with other disturbances. Unlike the Durbin-Watson test for AR(1) errors, the Lagrange Multiplier (LM) test may be used to test higher order ARMA errors. The null hypothesis is that there is no serial correlation. We shall estimate the observed R-squared statistic of the Breusch- Godfrey LM test to rule out serial correlation in the residuals.

(d) The ARCH LM Test of Heteroscedasticity : The ARCH test is a Lagrange multiplier (LM) test for autoregressive conditional heteroscedasticity (ARCH) in the residuals. This particular heteroscedasticity specification was motivated by the observation that in many financial time series, the magnitude of residuals appeared to be related to the magnitude of recent residuals. Ignoring ARCH effects may result in loss of efficiency. The ARCH LM test statistic is computed from an auxiliary test regression. To test the null hypothesis that there is no ARCH in the residuals, we run the regression:

$$e^2_t = \beta_0 + (\Sigma^q_{s=1} \beta_s e^2_{t-s}) + v_t ,$$

where, e is the residual. The F- Statistic (that measures the joint significance of all lagged squared residuals) and the observed-R-squared statistic (which is the LM test statistic) computed as the number of observation times the R^2 from the test regression should not be significant.

(4) Forecasting : The last step would involve forecasting the values of FDI inflows from 2012- 2020. After an iterative process of checking the best fitting model, we shall apply dynamic forecasting to get the output. Suppose the forecast sample is $j = T+1, T+2\ldots\ldots T+h$, and denote the actual and forecasted value in period t as y_t and $\hat{y}_t$, respectively (Schwert, 2009). The forecast error statistics computed are as follows :

Root Mean Square Error = $\sqrt{\Sigma^{T+h}_{t=T+1} (\hat{y}_t - y_t)^2/h}$

Mean Square Error = $\Sigma^{T+h}_{t=T+1} |\hat{y}_t - y_t|/h$

Mean Absolute Percentage Error = $100\,\Sigma^{T+h}_{t=T+1} \left| \dfrac{\hat{y}_t - y_t}{y_t} \right| /h$

Theil Inequality Coefficient = $\dfrac{\sqrt{\Sigma^{T+h}_{t=T+1} (\hat{y}_t - y_t)^2/h}}{\sqrt{\dfrac{\Sigma^{T+h}_{t=T+1} \hat{y}_t^2}{h}} + \sqrt{\dfrac{\Sigma^{T+h}_{t=T+1} y_t^2}{h}}}$

The first two forecast error statistics depend on the scale of the dependent variable. These should be used as relative measures to compare forecasts for the same series across different models; the smaller the error, the better the forecasting ability of that model according to that criterion. The remaining two statistics are scale invariant. The Theil inequality coefficient always lies between zero and one, where zero indicates a perfect fit.

The mean squared forecast error can be decomposed as:

$$\Sigma(\hat{y}_t - y_t)^2 / h = ((\Sigma \hat{y}_t / h) - \bar{y})^2 + (s_{\hat{y}} - s_y)^2 + 2\,(1 - r)\, s_{\hat{y}} s_y$$

Where $\Sigma\hat{y}_t / h, \bar{y}, s_{\hat{y}}, s_y$ are the means and (biased) standard deviations of $\hat{y}$ and y, and r is the correlation between $\hat{y}$ and y. The proportions are defined as:

Bias Proportion = $\dfrac{((\Sigma\hat{y}_t / h) - \bar{y})^2}{\Sigma(\hat{y}_t - y_t)^2 / h}$

Variance Proportion = $\dfrac{(s_{\hat{y}} - s_y)^2}{\Sigma(\hat{y}_t - y_t)^2 / h}$

Figure 3. Trend Line Showing Log of FDI Inflows in India (1991-2013)

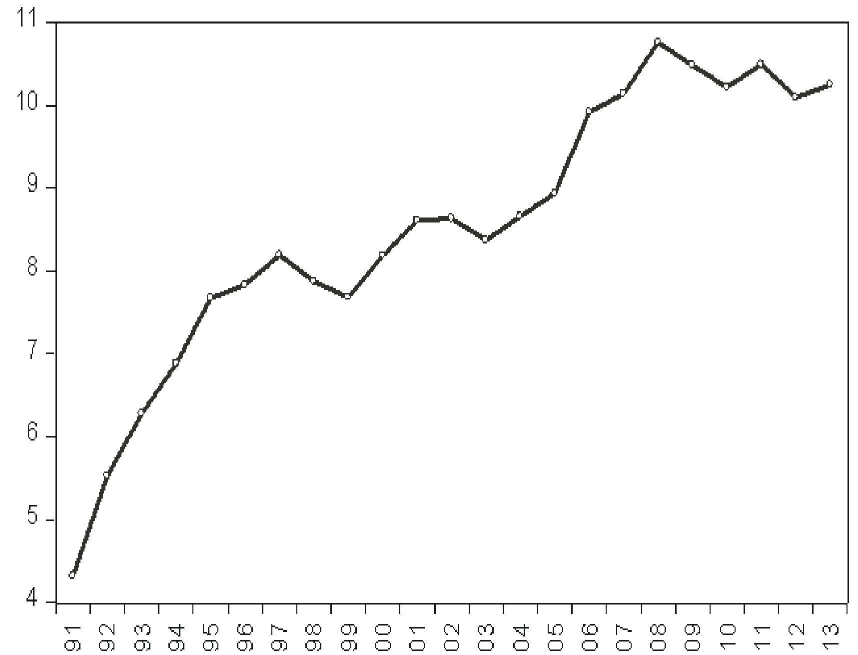
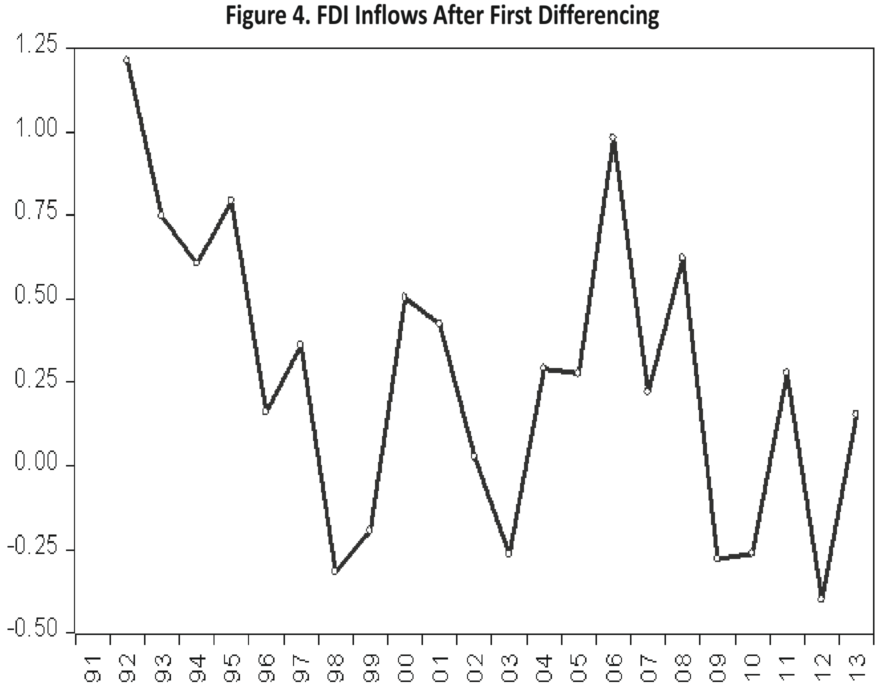


Figure 4. FDI Inflows After First Differencing



$$\text{Covariance Proportion} = \frac{2\,(1 - r)\, s_{\hat{y}} s_y}{\Sigma(\hat{y}_t - y_t)^2 / h}$$

The bias proportion indicates how far the mean of the forecast is from the mean of the actual series. The

Figure 5. Correlogram of Log of FDI Inflows at Level

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.789	0.789	16.289	0.000
		2	0.627	0.009	27.037	0.000
		3	0.471	-0.069	33.415	0.000
		4	0.349	-0.013	37.092	0.000
		5	0.268	0.034	39.391	0.000
		6	0.173	-0.087	40.408	0.000
		7	0.117	0.021	40.896	0.000
		8	0.033	-0.107	40.937	0.000
		9	-0.047	-0.074	41.026	0.000
		10	-0.096	-0.001	41.432	0.000
		11	-0.113	0.036	42.042	0.000
		12	-0.155	-0.119	43.291	0.000

Figure 6. Correlogram of Log of FDI Inflows after First Differencing

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.255	0.255	1.6376	0.201
		2	0.167	0.109	2.3729	0.305
		3	-0.042	-0.117	2.4218	0.490
		4	-0.193	-0.193	3.5157	0.475
		5	-0.011	0.113	3.5195	0.620
		6	-0.184	-0.167	4.6375	0.591
		7	-0.072	-0.040	4.8218	0.682
		8	-0.085	-0.044	5.0940	0.747
		9	-0.016	0.040	5.1038	0.825
		10	-0.097	-0.186	5.5148	0.854
		11	0.099	0.191	5.9843	0.874
		12	0.067	-0.007	6.2240	0.904

variance proportion specifies how far the variation of the forecast is from the variation of the actual series. The covariance proportion measures the remaining unsystematic forecasting errors. The bias, variance, and covariance proportions add up to one. If the forecast is "good", the bias and variance proportions should be small and most of the bias should be concentrated on the covariance proportions.

➪ **Criteria for a Good Forecasting Model :** The following criteria shall be used to get the best fitted model:

(1) The *F*- Statistics should be high enough to reject the null hypothesis that no model is possible.

(2) The values of AIC and SIC should be low.

(3) The R^2 and adjusted R^2 values should be high.

(4) The estimated coefficients should be statistically significant.

(5) The model should be parsimonious, simple, and effective and should not have too many coefficients.

(6) The model should be stationary.

Table 2. ADF Test and PP Test Results

Variable	Test for Unit Root in	Included in Test equation	ADF TEST Statistics	PP TEST Statistics
Log FDI Inflows	Level	Intercept	-3.524129**	-3.247833**
		Trend and Intercept	-3.117769	-3.080563
		None	-2.090937	1.637248
	First Difference	Intercept	**-3.830653*** **	**-3.830653** **
		Trend and Intercept	**-3.952903** **	**-3.944438** **
		None	**-3.456068*** **	**-3.476421*** **

* Significant at 10%, **Significant at 5%, *** Significant at 1%

(7) There should not be any patterns left in the AC and PACs.

(8) The residuals must be white noise.

(9) The root mean square error is low relative to other models and Theil's inequality coefficient is close to zero.

(10) The plots of actual against fitted values must indicate a good fit and withhold data well.

Results

(1) Model Identification : I applied the Box- Jenkins methodology. The process starts with model identification. I first conducted a graphical analysis to check the trend of FDI inflows over the last 21 years. As can be inferred from the Figure 3, there is an increasing trend in FDI inflows, and the series exhibits a random walk with non-stationary mean and variance.

Since the plotted graph after logarithmic transformation is non- stationary in its level form, the first difference of the series is plotted next. It can be observed from the Figure 4, that after taking the first differences, the series is found to be stationary with constancy in its mean and variance.

Next, the correlogram analysis is undertaken to examine the ACs, PACs and Q-Statistics. The correlogram presented in the Figure 5 shows that FDI is not stationary in its level form. The Q-Statistics are significant, suggesting that the series in its level form is not white noise. The ACs of the first two lags and PAC of the first lag lies outside the dotted lines of the two standard error terms. The ACs decline gradually as the number of lags increase and the PAC has a significant spike at lag 1 after which it dies off. This is suggestive of a strong autoregressive component.

After an initial examination of AC, PAC, and Q- Statistics of the series in its level form, the correlogram analysis of the series with first differencing is undertaken. When the series is taken in its first differenced form, no outliers are noticed in the ACs and PACs (see Figure 6). The ACs and PACs lie within the two dotted lines of the two standard error terms. The Q- Statistics is not significant and the p - values are > 0.05 at all the lags. This means that after taking first differences, the series has become white noise. Since the p -value > 0.05, the null hypothesis that there is no autocorrelation can be accepted.

The correlogram analysis has shown that after taking first differences, the series becomes stationary. To further corroborate the integration order of the series, the unit root tests are undertaken. The ADF test and PP test for unit root testing are performed in the series. The tests are conducted under the hypotheses: the series are stationary at levels and subsequently, they are stationary after first differencing (first differencing was undertaken only if the results at level show presence of unit root). The results have been compiled by considering three cases. The first case is that the equation includes the intercept, the second case is that the equation includes the trend and intercept, and the third case is that neither trend nor intercept are included into the equation. The results of the unit root tests are shown in the Table 2.

Table 3. ARIMA Model Estimation

Values in parenthesis ( ) show standard errors and [ ] show *t* - statistics

Model (*p,d,q*)	C	AR(1)	AR(2)	MA(1)	MA(2)	R^2	Adj. R^2	*F*-Statistic	Prob. (*F*-Stats)	AIC	SIC
(1,1,0)	0.207179* (0.117698) [1.760258]	0.257333 (0.193875) [1.327317]	---	---	---	0.084856	0.036691	1.761771	0.200133	1.075572	1.175051
(0,1,1)	0.276835** (0.117889) [2.348268]	---	---	0.262717 (0.213096) [1.232855]	---	0.069398	0.022868	1.491458	0.236192	1.288515	1.387701
(1,1,1)	0.133706** (0.047580) [2.810150]	0.677052*** (0.190220) [3.559316]	---	-0.999753*** (0.126001) [-7.934481]	---	0.317577	0.241753	4.188308	0.032098	0.877379	1.026596
(2,1,1)	0.214298 (0.325253) [.658864]	1.034678*** (0.335756) [3.081634]	-0.185946 (0.2.8375) [-0.892365]	-1.749250*** (0.590276) [-2.963447]	---	0.642480	0.575445	9.584256	0.000736	0.301805	0.500952
(1,1,2)	0.156710 (0.162577) [0.963913]	0.631073*** (0.080767) [7.813465]	---	-1.651422*** (0.477837) [-3.456033]	-0.429706 (0.507575) [-0.846588]	0.748230	0.703799	16.84061	0.000024	-0.024514	0.174442
(2,1,2)	0.169999 (0.158226) [1.074409]	-0.163255 (0.166234) [0.982079]	0.577754*** (0.120919) [4.778021]	-0.825132 (0.491815) [-1.677728]	-1.85277*** (0.487808) [-3.798154]	0.732711	0.661434	10.27978	0.000327	0.110944	0.359877
(0,1,2)	0.251675 (0.167041) [1.506665]	---	---	0.352926* (0.192565) [1.832757]	0.541507** (0.204377) [2.649555]	0.207778	0.124386	2.491589	0.109407	1.218434	1.367212
(2,1,0)	0.166868 (0.134155) [1.24384]	0.118098 (0.234489) [0.503637]	0.171774 (0.220085) [0.780492]	---	---	0.065913	-0.04397	0.599800	0.560132	1.162184	1.311544

Note : *Significant at 10%, **Significant at 5%, ***Significant at 1%

Table 4. Residual Analysis of the Selected Models

Model *(p,d,q)*	Normal Distribution Jarque- Bera Statistic (Probability)	Serial Correlation Breusch- Godfrey LM Test		Heteroscedasticity ARCH LM Test	
		F- Statistic (Prob. *F*-stat)	Observed- R^2 (Prob. χ^2)	*F*- Statistic (Prob. *F*-stat)	Observed- R^2 (Prob. χ^2)
(1,1,1)	0.67728	1.130114	2.398113	0.800259	0.851327
	(0.712739)	(0.3475)	(0.3015)	(0.3828)	(0.3562)
(2,1,1)	1.144093	25.55983	15.64425	5.91764	4.906053
	(0.564369)	(0.0000)	(0.0004)	(0.0263)	(0.0268)
(1,1,2)	1.100414	25.89255	16.27798	3.181992	3.004431
	(0.576830)	(0.0000)	(0.0003)	(0.0913)	(0.0830)
(2,1,2)	1.779906	21.83717	15.37832	3.553301	3.284763
	(0.410675)	(0.0001)	(0.0005)	(0.0766)	(0.0699)

As can be seen from the Table 2, log of FDI inflows is not stationary at the level, which confirms the findings from the correlogram analysis. Therefore, tests were conducted again by taking first differences. The results of both the ADF test and PP test show that the series becomes stationary after taking first differences. Due care has been taken to avoid over- differencing. Thus, the integration order of the series is $I(1)$.

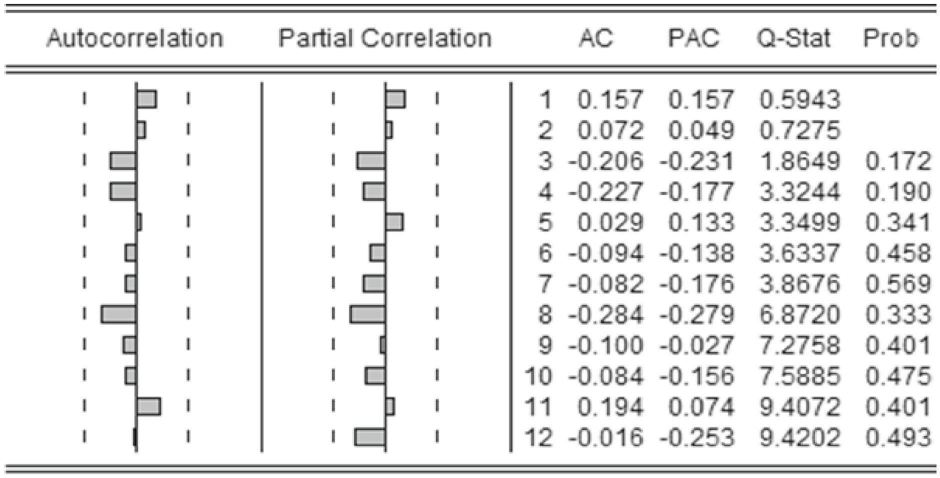
(2) Model Estimation : The graphical analysis, correlogram analysis, and unit root testing using ADF and PP tests have confirmed the integration order of the ARIMA model to be $I(1)$. Next, we need to determine the order of AR and MA terms. An examination of the patterns in ACs and PACs has indicated the presence of an autoregressive component, but a more robust testing is required to confirm these findings. My aim is to get a model that fits the actual data as closely as possible. To achieve this aim, I used an iterative process and examined eight different combinations of the p (autoregressive order) and q (moving average order) ; the integration order, d being 1. These eight combinations of *(p,d,q)* are (a) (1,1,0), (b) (0,1,1), (c) (1,1,1), (d) (2,1,1), (e) (1,1,2), (f) (2,1,2), (g) (0,1,2), and (h) (2,1,0).

The regression results of these eight models have been compiled and are presented in the Table 3. Out of these, the best fitted models have to be chosen by using the criteria mentioned in the section titled - Criteria for a Good Forecasting Model. Four models (1,1,0), (0,1,1), (0,1,2), and (2,1,0) have F- statistics which are not significant enough to reject the null hypothesis that no model is possible. These four models are, therefore, rejected. Of the remaining four models, (1,1,2) returned the lowest AIC and SIC values, while the other three models (1,1,1), (2,1,1), and (2,1,2) have returned regression outputs which are also acceptable. Model (1,1,1) is the only model in which all the coefficients are individually significant. Since the aim is to choose the best fitting model which is parsimonious and effective, I shall, for the time being, consider all of these four models and perform diagnostic checks to see if the residuals of each of these models are white noise.

(3) Model Diagnostics Checking : An accurate regression model is one in which the residuals are white noise. The four models that I have screened from the previous section are now subjected to diagnostic checks to see if the residuals are normally distributed and that there is no serial correlation and heteroscedasticity in the residuals. As described earlier, I shall use the Jarque- Bera test for checking the normal distribution of the residuals. The Breusch Godfrey LM test was applied to check serial correlation and the ARCH LM test was used to examine the presence or absence of ARCH in the residuals. The results of these three tests are reported in the Table 4.

The null hypothesis is that there is no serial correlation and ARCH cannot be accepted in Model (2,1,1).

Figure 7. Residual Diagnostic Using Correlogram for Model (1,1,1)



Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	0.157	0.157	0.5943	
		2	0.072	0.049	0.7275	
		3	-0.206	-0.231	1.8649	0.172
		4	-0.227	-0.177	3.3244	0.190
		5	0.029	0.133	3.3499	0.341
		6	-0.094	-0.138	3.6337	0.458
		7	-0.082	-0.176	3.8676	0.569
		8	-0.284	-0.279	6.8720	0.333
		9	-0.100	-0.027	7.2758	0.401
		10	-0.084	-0.156	7.5885	0.475
		11	0.194	0.074	9.4072	0.401
		12	-0.016	-0.253	9.4202	0.493

Table 5. Forecasting Error Statistics of the Two Models

Forecasting Error Statistic	Model (1,1,1)
Root Mean Square Error	**0.361301**
Mean Absolute Error	**0.315720**
Mean Absolute Percentage Error	**3.547082**
Theil's Inequality Coefficient	**0.020224**
• **Bias Proportion**	**0.009259**
• **Variance Proportion**	**0.000535**
• **Covariance Proportion**	**0.990205**

Figure 8. Forecasted FDI Inflows-Model (1,1,1)

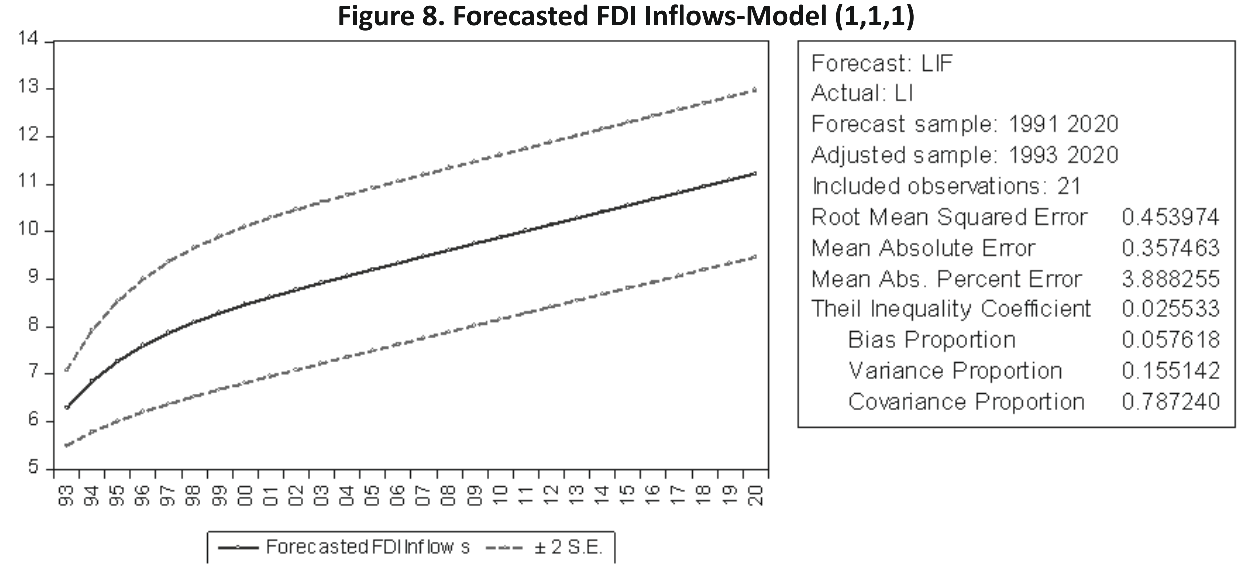
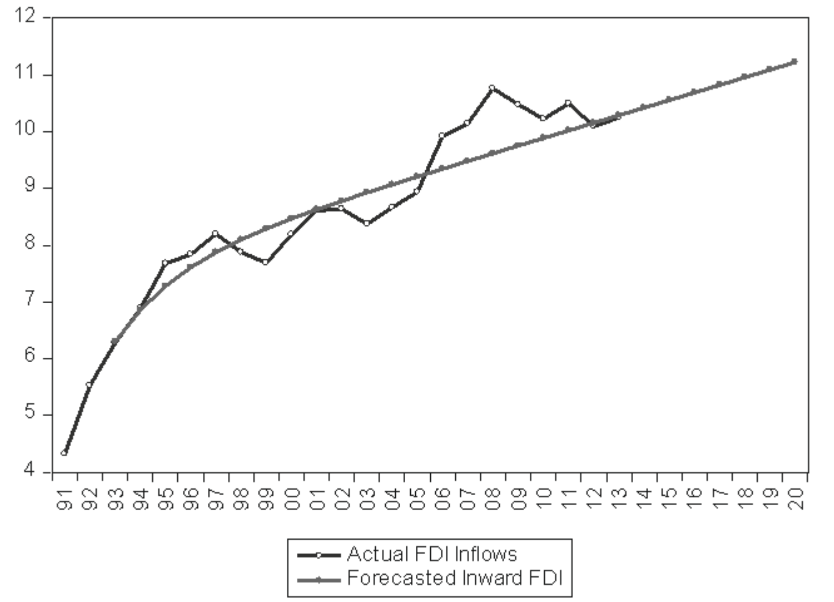


Figure 9. Actual Data Versus Fitted Model (1.1.1)



Therefore, this model is rejected as it is not white noise and selecting it would lead to an inaccurate forecast. On the other hand, the Models (1,1,2) and (2,1,2) have residuals which are normally distributed but have serial correlation and also have marginal degrees of heteroscedasticity. The only model which has residuals that are normally distributed and which does not have serial correlation and heteroscedasticity is Model (1,1,1). I further execute the correlogram analysis and check the residuals using Ljung- Box test statistics for this model. The output of this analysis is presented in the Figure 7 for Model (1,1,1). The ACs and PACs of the selected model lies within the two dotted lines of two standard errors. The Q- Statistics is not significant (*p* - values at all lags > 0.05). Thus, it can be confirmed that the model has residuals that are purely random and these error terms have zero mean, constant variances, and are serially uncorrelated.

(4) Model Forecasting : After using an iterative process of ARIMA modeling, the best-fitting model has been short-listed for forecasting FDI inflows in India. I shall now apply a number of forecasting error statistics before finally forecasting FDI. The criterion for final selection of the model has already been discussed in the section titled - Criteria for a Good Forecasting Model. The chosen model must have a lower forecasting error. This can be decided by examining the root mean square error, mean absolute error, and mean absolute percentage error. The smaller the values of these statistics, the better would be its forecasting ability. Additionally, I shall examine Theil's inequality coefficient, the bias proportion, variance proportion, and covariance proportion. In Theil's inequality coefficient, the model will be chosen only if it represents a close fit. The model must have smaller values of bias proportion and variance proportion, and larger covariance proportion ideally. The results of the forecasting error statistics are shown in the Table 5.

It is evident from the Table 5 that Model (1,1,1) is a good model. Model (1,1,1) has an acceptably low root mean square error, mean absolute error, and mean absolute percentage error. Its Theil inequality coefficient is close to zero, indicating a good fit. Model (1,1,1) can be a good forecasting model because the values of bias and variance proportion are small and most of the bias is concentrated on the covariance proportion, which is a desirable case.

The Figure 8 shows the forecasted FDI inflows as per Model (1,1,1) ; whereas, Figure 9 shows the plots of the actual FDI data versus the fitted model. From these plots, it becomes all the more evident that Model (1,1,1) is an accurate model as the forecasted line shows a good fit to the data. This is in conformity with the correlogram

Table 6. Forecasted FDI Inflows in India (2014- 2020)

Year	Forecasted Log of FDI Inflows	-2 S.E. of Forecasted Log of FDI	+ 2 S.E. of Forecasted Log of FDI	Forecasted FDI Inflows in US$ (in Millions)	-2 S.E. of Forecasted FDI in US $ (in Millions)	+2 S.E. of Forecasted FDI in US $ (in Millions)
2014	10.42181	8.681314	12.16231	**33584.16651**	5891.78331	191435.4586
2015	10.55563	8.812518	12.29874	**38392.98348**	6717.813443	219419.7849
2016	10.68942	8.943704	12.43514	**43889.04397**	7659.515423	251484.34
2017	10.82318	9.07487	12.57149	**50170.3754**	8733.050368	288223.0677
2018	10.95692	9.206024	12.70782	**57349.53417**	9956.929302	330319.6166
2019	11.09065	9.33717	12.84413	**65555.34325**	11352.23591	378560.0531
2020	11.22438	9.468322	12.98044	**74935.27351**	12943.15039	433843.0015

analysis carried out earlier which had indicated an even presence of autoregressive and moving average components. India's FDI inflows show a pattern different from Brazil's (Turolla & Margarido, 2011) as Brazil had a stronger auto- regressive component with AR(4) and MA(1) *p,q* values. It is also different from Jordon, as Bashier and Talal (2007) showed a bent towards moving average pattern with *p,d,q* (0,1,1). The Model (1,1,1), which is the final selection, can now be used in univariate modeling to generate forecasts of the future values of the series.

As a final step, the chosen ARIMA Model (1,1,1) is used in forecasting FDI inflows in India from 2014 to 2020. The forecasted log of FDI inflows is first tabulated along with ±2 standard errors, which show the lower and upper limits of the forecasts at 95% confidence intervals. These logarithmic values are then transformed back into natural numbers to get the values of forecasted FDI inflows in US $ (in millions). The results are depicted in the Table 6.

The range of FDI inflows over the forecasted period (2014- 2020) is US $ 33,584.16 million to US $ 74,935.27 million. The average annual FDI inflow expected in India over the forecasted period is US $ 51,982.39 million. The compounded annual growth rate (CAGR) of FDI inflows between 2014 and 2020 is 14.31%. These results show an increasing trend of FDI inflows in India. The phenomenal CAGR of 14.31% expected over the forecasted period is indicative of excellent investor confidence in the Indian economy.

Research and Policy Implications

In this paper, ARIMA modeling is used to forecast FDI inflows in India. To the best of my knowledge, this may be one of the first studies that has attempted to use the Box- Jenkins methodology for forecasting FDI inflows in India. Since FDI can have an impact on other macroeconomic variables such as GDP and exports, an accurate forecasting can be useful for policy making. The findings of the paper have shown that FDI inflows are likely to grow at 14% per annum and potentially reach US $ 50 billion by 2017.

The Government of India has already launched the 'Make in India' campaign in a bid to attract foreign investors. However, there has been some skepticism with regard to ease of doing business in India. Policy makers should use this opportunity of huge investor interest in the economy to improve its infrastructure, expedite land acquisition reforms, and increase supply of skilled labour. Increased FDI inflows also have ramifications for domestic industries, which are likely to get exposed to higher competition. Thus, policies need to be framed to make Indian industry more competitive, for example, providing easier access and credit for technology upgradation. So far, huge market size and high GDP growth rates have caused an influx of FDI in India (Chadha, Singh, & Natarajan, 2014). Perhaps, policy makers in India should now plan to make this unidirectional relationship a bi-directional relationship by attracting FDI in areas which contribute to its GDP growth.

Conclusion

For a developing country like India, FDI can be an important source of finance, and it can contribute richly to the economy by transferring superior managerial skills and state of the art technology into the country. Ever since India undertook economic reforms in 1991, it has been one of the largest recipients of FDI in the world. This paper has attempted to forecast FDI in India from 2014 to 2020 using univariate ARIMA modeling. Applying the Box-Jenkins methodology, the process of model identification, estimation, diagnosis, and forecasting were undertaken. As many as eight different ARIMA models were estimated, from which one was identified as the best-fitting model after an iterative process. Several accuracy tests were used and after confirmation of white noise in residuals, the model was eventually selected, which had the least forecasting error and biasness.

The ARIMA Model (1,1,1) was found to be most suitable and provided the tightest fit to the data. The FDI inflows exhibit an even autoregressive and moving average trend in India. As per the forecasted model, the average annual FDI inflows over the forecasted period would be US $ 51982.39 million. The compounded annual growth rate (CAGR) of FDI inflows between 2014 and 2020 is likely to be 14.31%. The high level of foreign investment in India is indicative of its strong economic fundamentals. As per this forecast, the FDI inflows in the year 2020 could potentially be US $ 74,935.27 million.

Limitations of the Study and Scope for Further Research

The study uses univariate ARIMA modeling for forecasting FDI in India. Using univariate models has limitations as complex economic subjects like FDI may be dependent on several other factors. Research studies using panel data estimation have proven that gross domestic product, openness of economy, institutional efficiency, exchange rates, inflation, literacy rates, possession of strategic assets, and so forth are also important determinants of FDI in India (Singh, 2011). Perhaps, future researchers could attempt using multivariate ARIMA modeling for more precise forecasting.

References

A. T. Kearney. (2011). *Foreign direct investment (FDI) confidence index*. Retrieved from http://www.atkearney.com/gbpc/foreign-direct-investment-confidence-index

Ahluwalia, M. S. (2002). Economic reforms in India since 1991: Has gradualism worked? *The Journal of Economic Perspectives, 16* (3), 67-88. DOI: 10.1257/089533002760278721

Al- Rawashdeh, S.T. ,Nsour, J. H., & Salameh, R.S. (2011). Forecasting foreign direct investment in Jordon for years (2011- 2030). *International Journal of Business and Management, 6* (10), 138- 154. DOI: http://dx.doi.org/10.5539/ijbm.v6n10p138

Bashier, A. A., & Talal, B. (2007). Forecasting foreign direct investment inflow in Jordon : Univariate ARIMA model. *Journal of Social Sciences, 3* (1), 1-6. DOI : 10.3844/jssp.2007.1.6

Chadha, S. K., Singh, B., & Natarajan, V. S. (2014). Interaction among foreign direct investment, economic growth and foreign trade: Evidence from India and China. *International Journal of Process Management and Benchmarking, 4* (2), 145-166. DOI: 10.1504/IJPMB.2014.060405

Cheng, L. K., & Ma, Z. (2007). *China's outward FDI: Past and future* (Working Paper No. 200706001E). Beijing : School of Economics, Renmin University of China.

FDI Intelligence. (2012). *The fDi report 2012: Global greenfield investment trends*. Retrieved from http://ftbsitessvr01.ft.com/forms/fDi/report2012/files/The_fDi_Report_2012.pdf

Gujarati, D.N., Porter, D.C., & Gunasekar, S. (2009). *Basic econometrics* (5th ed.). New Delhi : Tata McGraw-Hill Education Private Ltd.

Kumar, G., & Dhingra, N. (2012). Growth performance and forecasts of FDI inflows to Sri Lanka. *E3 Journal of Business Management and Economics, 3*(8), 307-317.

Schwert, G. (2009). *Eviews 7 user's guide II*. LLC, Irvine, California : Quantitative Micro Software.

Singh, R (2011). *Foreign direct investment in India: A sectoral analysis* (Unpublished Doctoral Thesis). Faculty of Management Studies, University of Delhi, New Delhi.

Turolla, F. A., & Margarido, M. A. (2011). Modeling and forecasting foreign direct investment into Brazil with ARIMA. *Economia Global e Gastao, 16* (2), 86-100.

UNCTAD. (2007). *World investment report 2007*. United Nations: New York and Geneva.

UNCTAD. (2010). *World investments prospects survey 2010- 2012*. United Nations: New York and Geneva.

UNCTAD. (2012). *World investment report 2012*. United Nations: New York and Geneva.

UNCTAD. (2013). *World investment report 2013*. United Nations: New York and Geneva.

UNCTAD. (2014). *Inward foreign direct investment flows annual: 1970- 2011*. Retrieved from http://unctadstat.unctad.org/TableViewer/tableView.aspx?ReportId=88