

Chapter- IV



DATA, RESEARCH DESIGN AND METHODOLOGY

4. DATA AND RESEARCH METHODOLOGY

This chapter provides a comprehensive framework that testifies the relationship between foreign investment and economic development empirically. First of all, the study briefly describes the selected variables. Next it highlights the scheme of investigation and elaborates methodologies.

4.1 Data

4.1.1 Sample Design and Selection of Variables

The selection of the variables is based on their economic relevance and the degree of impact on the Indian economy. The study considers both, existing theory and empirical evidence in this regard. Moreover, for selecting the variables we carefully consider the previous literature and incorporate those variables which are frequently used in the previous studies to capture the dynamic relationship between the flow of Foreign Investment and the Indian Economic Development. The study considers a broader set of variables, namely GDP, Foreign Investment, FDI, FII, etc., to capture the effect of foreign investment on Indian economic development. In this study, we consider GDP in proxy of Indian economic Development.

4.1.2 Description of the Selected Variables

Gross Domestic Product (GDP): The GDP is widely considered as one important measure of the economic welfare of a country. It is defined as the total rupees value of all finished goods and services produced within the country over a specific period of time, usually estimated on annual and quarterly basis. Economic activities, concerned

with production, consumption and distribution of goods and services, are carried out by three basic economic units, namely households, firms and government. Consequently they are the final user of goods and services. One of the three alternative methods of measuring national income is by conjoining the expenditures, private consumption expenditure by households, and non-profit organizations(C), business expenditure and home purchases by households expressing as investment expenditure (I), government spending(G), and net trade i.e. export minus imports(XN), made by these economic agents, known as expenditure approach. Generally adjusting factors income received from abroad to national income is GDP. GDP is probably the single best indicator of economic development and it also refers to as the size of the economy. Samuelson and Nordhaus (1985) neatly sum-up importance of GDP in their seminal textbook 'Principles of Economics' as it helps policy makers, central banks, investors to judge the economy is contracting or expanding, whether it requires boosts or restrains and identify whether recession or inflation is looming on the horizon. Finally, this variable has an impact on monetary and fiscal policy, economic shocks, tax and spending plan as well as reduction in the severity of business cycle. In India there are plenty of organizations providing the GDP data by summing the contributions of the three basic economic sectors to the GDP. These sectors are primary or agriculture, secondary or industry/manufacturing and tertiary or service sector.

Foreign Direct Investment (FDI): Foreign Direct Investment (FDI) is an important contributor to the economic growth of a country. It is also a major source of non-debt financial resource for the economic development of a country. Multinational Corporations (MNCs) invest in India to take advantage of enormous demand in the domestic market (both chartered and unchartered), cheaper labour cost, and also

special investment privileges like SEZ, tax holidays, etc. For a country where cross border direct investments are being made, it also means achieving tangible assets, technology & know-how, expertise management skills, and generating employment. The continuous flow of inward FDI in India, which is now allowed and absorbed across several economic units, clearly shows that the faith which cross border investors have in the country's economy. The inflow of foreign direct investment has come to India into the basic three sectors of the economy and exerts different impact on the sector. Therefore, cross border direct investment inflow particularly into primary sector, secondary sector and tertiary sector are the principal sectoral variables.

Foreign Institutional Investments (FII): The term “foreign institutional investment” or FII is used most commonly in India to refer to investments in the Indian financial markets by cross-border financial institutions. In India, International institutional investors must register with the SEBI to take part in the capital market. Positive fundamentals, amalgamation with fast-growing markets, have made India an attractive destination for foreign institutional investors. FII acts as a stimulator for the development of the country’s economy because it helps in obtaining capital at a lower cost and provides access to cheap global credit. Moreover, it complements domestic savings and investments. The net foreign institutional investment in equity and debt will be used to represent the flow of international fund in Indian capital market.

4.1.3 Data Source

To establish the long-run relationship and the short-run dynamics, the study uses quarterly data on Gross Domestic Products (GDP) at constant prices with base 2004-05, Foreign Investment (FI), Foreign Direct Investment (FDI) and Foreign

Institutional Investment (FII) in their net inflow values for the period of first quarter of 1996 to third quarter of 2016. The data are collected and composed from Data base on Indian economy and various issues of Handbook of Statistics on the Indian economy published by the Reserve Bank of India (RBI). But, the annual data for the period of 1995 to 2016 have been used in this study to examine the effect of sectoral inflow of FDI on the output growth of the respective sectors in context of Indian economy. The India's national account statistics in the form of sector-wise GDPs are collected from Ministry of Statistics and Program Implementation (MOSPI), Government of India and subsumed these data in the three distinct heads namely agricultural, manufacturing and service sector GDP. Data on sector-specific FDI inflows are collected from statistics released by Ministry of Commerce and Industry, Government of India and database of INDIASTAT. Keeping consistency with the GDP data format the data of FDI inflows are also segregated into three economic sectors. The FDI inflows in agricultural, manufacturing and service sector are denoted as FDI_AGR, FDI_MFG and FDI_SRV respectively. Likewise, the sectoral contributions to GDP for these three sectors are presented by GDP_AGR, FDI_MFG and GDP_SRV respectively.

4.1.4 Study Period

After the opened up policy was adopted by the Government of India in 1991, the Indian economy has liberalized considerably so as to allow even cross border investors to invest in India in most of the sectors of the economy as well as Indian investors are allowed to invest abroad. The most profound changes in the Indian economic scene has been witnessing since from this period. The forces of Liberalization, Privatization and Globalization (LPG) have radically changed the bases of the Indian economy. Foreign investments in the form of direct and portfolio

have been welcoming with open hands and extending it through rapidly policy transformation namely, Special Economic Zones (SEZ) policy, Export Processing Zones (EPZ) Merchandise Export India Scheme, Make in India campaign, etc. In this circumstances, the macro level empirical investigations is being carried out using quarterly data since from first quarter of 1996 to third quarter of 2016. Although, the unavailability of monthly data of sector level inflows of FDI and also sectoral output we have bounded to conduct the intended empirical investigation on annual data from 1995 to 2016.

4.2 Research Methodology

4.2.1 Statistical and Econometric Tests Used in the Study

The study will employ several statistical tools and techniques like, Mean, Median, Standard Deviation, Skewness, Kurtosis, Jarque-Bera test statistics, t-test, F-test, Chi-square test etc. to get the descriptive information of the variables and for data analysis.

With a view to accomplishing the stipulated set of objectives we have use several econometric methods related to modern time series analysis. The Unit Root test, Johansens Cointegration test, Auto Regressive Distributed Lag (ARDL) bound test, Regression analysis, Vector Autoregression (VAR) Estimation, Vector Error Correction Model (VECM) and several other method have been used to explore the long-run and short-run relationship among the variables.

To determine the long-run relationship among FDI, FII and the other economic variable the study will consider Johansen cointegration test or bound test for cointegration under ARDL framework. On the basis of the unit root property the study

employs either any of these techniques to address the long-run association between the modeled variables. If the variables specified in a model are stationary at their first difference then apply VAR based cointegration test proposed by Johansen (1988).

4.2.1.1 Unit Root test

The time series analysis is a challenging and an important field of research on macro economic analysis. Empirical relationship among the macroeconomic variables, like GDP, FDI, FII and so on, is based on time series analysis. Before we proceed to the time series analysis, it is sensible to test stationarity of a data series which allow us to draw meaningful inferences from the analysis. It is also suppose to increase the accuracy and reliability of the underlying model used. Simply, in inferential statistics, the term stationarity implies first two moments of a time series remains constant over considering time span. It explains that the future will behave very similar to the past and on the basis of the past values reliable forecast can be easily made. Sometimes, empirical time series exhibits very common type of non-stationary behavior which can be modeled by Autoregressive Integrated Moving Average process (ARIMA). Notably, which is commonly known as “unit root process” as the autoregressive polynomial related with the process contains roots which are on the unit circle. Therefore, in empirical time series it is crucial to test whether there exist one or more unit roots.

The unit root test results provide an idea about whether the underlying data series contains unit root property or not. From the test results it can be also identified the order of integration. The order of integration is much essential for conducting both the regression and co-integration analysis. If the stipulated data series have not the correct order of integration there may arise problem of spurious regression. That can make

the analysis useless and it may draw inferences in favour of significant relationship where actually the series are totally unrelated. Therefore, it is imperative to conduct unit root test for time series data to draw a valid and meaningful relational inferences.

There are plenty of unit roots tests available. Going through the pros and cons of different tests, we decide three of the most popular and recognized test, namely, Augmented Dickey-Fuller (ADF) test, Phillips-Perron (PP) test, and Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test.

Augmented Dickey-Fuller Test

The stationarity of a time series variables are tested by data generating process or, in other words, to know the order of integration of time series variables the analysis is gone through the stationarity process. The Augmented Dickey-Fuller (ADF) Test under Autoregressive moving Average (ARMA) (p,q) is an extension of Dickey-Fuller Test of single order auto regression. If the series is integrated at higher order lags than one, the assumption of white noise disturbances is violated in simple Dickey-Fuller unit root test. Therefore, the ADF test is the most popular test for checking stationarity of time series data in empirical research. In performing ADF test to check the unit root we have to face two practical issues. First, we have to choose the model/s involving a constant, a constant and a linear time trend, or none of them. One point should be mentioned here that use of intercept, and intercept and a linear time trend, is very popular as the other two cases are rarely applicable in macro-variable analyses. So, we apply ADF test under both the generally accepted models, constant and with constant and linear trend. The associated ADF equations are

$$\Delta Y_t = \alpha + \pi Y_{t-1} + \sum_{j=1}^p \delta \Delta Y_{t-j} + u_t$$

$$\Delta Y_t = \alpha + \beta_t + \pi Y_{t-1} + \sum_{j=1}^p \delta \Delta Y_{t-j} + u_t$$

where y_t is the variable in period t , α is a constant, β is the coefficient on a time trend, p is the lag order of the autoregressive process and u_t is white noise.

Secondly, we have to specify lag lengths on the basis of some selected criteria. Instead of arbitrarily assign lag lengths to the model Schwarz Information Criterion (SIC) is applied as lag selection criterion of ADF test. The decision depends upon rejection or acceptance of the null hypothesis, H_0 , by comparing the test statistics with critical values. Here, H_0 hypothesis advocates the series is not stationary and contains unit root, while the alternative (H_1) hypothesis assumes that the series is stationary. This hypothesis and other hypothesizes made in this study have been tested by applying the standard statistical test procedures.

Phillip-Perron (PP) Test

Phillip and Perron (1998) have proposed an alternative nonparametric method of controlling for serial correlation in the error terms without adding lagged difference terms when testing for a unit root. The main point of difference between ADF and PP test is the process of take care of serial correlation and heteroskedasticity in the error terms. More specifically, ADF test uses a parametric auto-regression to approximate the ARMA structure of the error terms in the test regression. The PP test estimates a non-augmented Dickey-Fuller test equation and modifies the t-ratio of α coefficient so that serial correlation does not affect the asymptotic distribution of the test statistic.

Here, also we have two choices to run PP test. First, we have to choose model/s and second, to specify lag length. Here, like ADF test, we have considered two most popular models which are more general in application, and the Newey-West method

is applied as lag selection criteria of PP test. The test regressions in terms of two separate models, only intercept and with intercept and trend for the PP tests are

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

$$\Delta Y_t = \alpha + \beta_t + \gamma Y_{t-1} + u_t$$

where, u_t is stationary in level, that is, $I(0)$ and may be heteroskedastic. The PP test remove the problem of auto-correlation and heteroskedasticity in the error terms u_t by directly modifying the test statistics. This feature of the PP test makes it distinctly better than ADF test by improving its robustness.

In the PP test, the decision depends upon rejection or acceptance of the null (H_0) hypothesis by comparing the statistics obtained from the test with critical values. Null (H_0) hypothesis advocates that series is not stationary and has unit root, while the alternative hypothesis (H_1) advocates that series is stationary. If the calculated value is higher than the absolute critical value, then (H_0) hypothesis is rejected and series is confirmed to be stationary.

Kwiatkowski-Phillips-Schmidt-Shin (KPSS) Test

Denis Kwiatkowski, Peter C. B. Phillips, Peter Schmidt, and Yongcheol Shin introduce an alternative test for estimating the stationary property of time series variables in 1992, and is called the KPSS test. The KPSS test statistic is applied for checking whether the data series is stationary, or non-stationary due to the presence of a unit root. The KPSS test differs from the other unit root tests described above in that the time series Y_t is assumed to be stationary under the null hypothesis. This test is typically intended to complement stationary tests, such as the ADF and PP tests.

The test is performed through regressions in terms of two separate models, only intercept and with intercept and trend, for the KPSS tests are

$$Y_t = \alpha + \mu_t + u_t$$

$$Y_t = \alpha + \beta_t + \mu_t + u_t$$

where, $\mu_t = \mu_{t-1} + \varepsilon_t$ and ε_t is white noise. u_t is stationary in level, that is, $I(0)$ and may be heteroskedastic. Here μ_t is used to denote a pure random walk having innovation variance σ_s^2 . In KPSS test the null hypothesis that Y_t is stationary in level, i.e., $I(0)$, is synthesized as $H_0: \sigma_s^2 = 0$, which assumes μ_t as a constant. On the other hand, the alternative hypothesis presumes the series to be nonstationary due to the presence of a unit root property. The KPSS test statistic is calculated with the help of the Lagrange Multiplier (LM) technique, and if the computed value is greater than critical value at a specific significance level, the null hypothesis of stationarity is rejected at given level of significance.

4.2.1.2 Selection of Optimum Lag Length

The autoregressive model is sensitive to lag length. Therefore, one needs to determine the appropriate lag length in the model. At the same time one should be known about the number of independent variables including the variable(s) with lags, since the larger the number of the independent variables, the smaller will be the degrees of freedom and lesser will be the accuracy of the test results. There is no technique that is commonly agreed upon regarding how to select the lags and variables structure while the outcome of the estimation heavily depends on the estimated settings. The study determines the optimum lag length based on the Akaike Information Criteria (AIC), Schwarz Information Criteria (SIC), and HannanQuinn Information Criteria (HQC). Generally the AIC and HQC criteria suggest a higher lag length. In most of the cases the study tries to impede to take the risk of over parameterization by considering lags that are too high for the VAR model.

The procedure to choose the lag order and the rank of the structure of short-run restrictions is carried out by minimizing the following modified information criteria (see; Vahid and Issler, 2002; Hecq, 2006).

$$AIC(p, s) = \sum_{i=n-s+1}^T \ln(1 - \lambda_i^2(p)) + \frac{2}{T} \times N \dots\dots\dots(1)$$

$$HQ(p, s) = \sum_{i=n-s+1}^T \ln(1 - \lambda_i^2(p)) + \frac{2 \ln(\ln T)}{T} \times N \dots\dots\dots (2)$$

$$SC(p, s) = \sum_{i=n-s+1}^T \ln(1 - \lambda_i^2(p)) + \frac{\ln T}{T} \times N \dots\dots\dots (3)$$

$$N = [n(n(p-1)) + nr] [s(n(p-1) + (n-s))]$$

where n is the number of variables in model (2) and N is a number of parameters. N is obtained by subtracting the total number of mean parameters in the VECM (i.e., $n^2(p-1) + nr$), for given r and p, from the number of restrictions the common dynamics imposes from $s(n(p-1)) - s(n-s)$.

4.2.1.3 Johansen's Cointegration Test

The cointegration test is a technique used to determine whether a set of endogenous variables have in our study, (macroeconomic and sector level variables) share a common long-run stochastic trend (having a long-run relationship). Although, there also may have the possibility to observe the short-run divergences. The presence of cointegration implies the co-movement of the endogenous variables, which may be the result of linkage between the economic development and the foreign investment at both the macro-level and sector-level. In the time series analysis it is ultimately useful(relevant) when the considered time series are non-stationary in level and all the variables used in the study should be integrated in the same higher order. In typical econometric sense, two or more variables are referred co-integrated when they move together or in other words they share a common trend. Specifically, a vector of

variables, which overcome the non-stationary problem after differencing, can have linear combinations which are stationary in levels. This linear combination is referred to as the cointegration equation. This equation depicts and explains long-run equilibrium relationship among the variables. The cointegration technique has been used to analyze the long-run relationship between economic development and foreign investment at both the macro level and sector level.

The idea of the cointegration test is simple. Suppose Y_t and X_t are integrated of order one, or $Y_t \sim I(1)$ and $X_t \sim I(1)$. Then Y_t and X_t are said to be cointegrated if and only if u_t obtained from the long run relationship regression is integrated of order zero or $u_t \sim I(0)$. Therefore, if the cointegration condition is met, then Y_t and X_t move together in the long run, such that they cannot drift arbitrarily far apart from each other as time goes on.

A great number of proposals have been made to address the cointegration relationship between the macroeconomic series like Hubrich, Lutkepohl and Saikkonen and many others. The study adopts the Johansen's cointegration (1988) under Likelihood Ratio (LR) approach based on Gaussian assumption procedure. This is more efficient than univariate Engle-Granger cointegration tests and other tests mentioned above having severe shortcomings in some situations (Lutkepohl, 2004), to decipher the long-run equilibrium relationship among the variables under study. Under this approach of the cointegration test, Trace test (or Likelihood ratio test), as well as Maximum Eigen value test are applied to understand the existing long-term dynamics among the variables. This test is based on the following vector autoregressive model :

$$Y_t = \mu + A_1 Y_{t-1} + A_2 Y_{t-2} + A_3 Y_{t-3} + \dots + A_p Y_{t-p} + u_t$$

Where Y_t is a vector containing n variables, all of which are integrated of same order and the subscript t denotes the time period. μ is an $(n \times 1)$ vector of constants, A_p is an $(n \times n)$ matrix of coefficient where p is the maximum lag included in the model and u_t is an $(n \times 1)$ vector of error terms. The above VAR equation can be written in the form of the error correction framework as:

$$\Delta Y_t = \mu + \Pi Y_{t-p} + \Gamma_1 \Delta Y_{t-1} + \Gamma_2 \Delta Y_{t-2} + \dots + \Gamma_{p-1} \Delta Y_{t-p+1} + u_t$$

Where, $\Gamma_i = - \sum_{j=i+1}^p A_j$ represents the dynamics of the model in the short run; and $\Pi = \sum_{i=1}^p A_i - I$ represents the long run relationship between the variables included in the vector Y_t , and I is the identity vector. Ascertaining the rank of the matrix Π is the principal idea of the Johansen's approach, where rank of the matrix Π represents the number of co-integrating vectors. Thereafter, the study proceeds to estimate two principal test statistics, namely, Trace statistics and the Maximum Eigen Value statistics.

4.2.1.4 Auto Regressive Distributed lag (ARDL) Model

Cointegration is an econometric formulation that helps to modeling time series in order to measure the existence of a long-run equilibrium that converges over time. The ARDL cointegration technique (bound test for cointegration), developed by Pasaran & Shin (1999) and Pasaran et.al. (2001), does not demand pretests for unit roots unlike other cointegration techniques. This ARDL method is imperative when dealing with time series that are integrated of different order or $I(0)$, provided that nonexistence of integration of order 2 or more, and also providing robust results when there presents only one cointegrating relationship between the underlying series in a small sample size. The ARDL cointegration technique(bound test for cointegration) is attached with a few crucial advantages : i) it even allows testing to

check the presence of a cointegrating relationship between series in levels irrespective of whether the underlying regressors are $I(0)$ or $I(1)$; ii) it is considered to be more apposite than the Johansen & Juselius multivariate cointegration technique (1992) for testing the long-run relationship amongst time series of a small sample size (Mah1995; Tang & Nair 2002); iii) it also provides a consistent short-run parameters and super consistent long-run parameters of the estimators in small sample sizes (Pasaran and Shin 1999).

The cointegrating relationship or long-run association of the underlying series is captured through the F-statistic (Wald Test). Its asymptotic distribution is non-standard under the null hypothesis of no cointegration. Pasaran et al. (2001) have prescribed two different set of critical values for given significance level as the F statistics employed for this test have a non-standard asymptotic distribution. The first set of critical values assumes that all series are integrated of order zero, $I(0)$ and the second set assumes all series are integrated of order one, $I(1)$. If the calculated F statistic is overstepped the critical value band which is depicted in Passaran table, then the null hypothesis is rejected i.e. there may be exists a long-run co-movement between the series. In contrast, if the calculated F statistic is below the lower critical bound value then the null hypothesis is accepted. Here noteworthy to mention that, the test becomes inconclusive if computed F statistic lies between bound values.

Unlike the Johansen and Juselius(1990) cointegration technique, ARDL countegration assists in capturing the cointegrating vector(s). However, existence of a single long run cointegrating equation further plunge to re-parameterized the cointegrating vector into Error Correction Term (ECM) for better understanding of short-run dynamics between the series. Simply, distributed lag technique includes unrestricted lag of the

repressors in a regression equation. The ARDL equations for two variables, namely X and Y, is as follows:

$$\Delta Y_t = \alpha_0 + \alpha_1 Y_{t-1} + \alpha_2 X_{t-1} + \beta_i \sum_{i=1}^p \Delta Y_{t-i} + \delta_i \sum_{i=0}^q \Delta X_{t-i} + \varepsilon_t$$

$$\Delta X_t = \alpha_0 + \alpha_1 X_{t-1} + \alpha_2 Y_{t-1} + \beta_i \sum_{i=1}^p \Delta X_{t-i} + \delta_i \sum_{i=0}^q \Delta Y_{t-i} + \varepsilon_t$$

Here, p and q are automatically model specified maximum lag orders for dependent and independent variables respectively. α_1 and α_2 are the coefficients of the lagged variables and $(\alpha_1 - \alpha_2)$ represents the long-run relationship between variables in the model. Another two coefficients, β_i and δ_i correspond to the short-run dynamics of the model.

4.2.1.5 Vector Error Correction Model (VECM)

It is quite possible that in the long-run there exists equilibrium relationship between the components of foreign investments and economic development as measured by real GDP, but in the short-run observes the existence of disequilibrium relationship. The nature of the relationship among the variables in the short-run can be answered by considering the Vector Error Correction Mechanism (VECM) or Auto Regressive Distributed Lag Estimation (ARDL) or Vector Autoregressive (VAR) Estimation. If there do not exists any cointegrating vectors among variables, ARDL model or VAR model will be used to capture short-run dynamic relationship among the variables.

However, if there exist one or more cointegrating vectors under Johansen technique, the Vector Error Correction Model is employed for estimating short-run dynamics. The error correction procedure is a way that reconciles short-run and long-run behavior through a series of partial short-run adjustments. More precisely, in a two variable setting where X and Y are integrated of order one or I (1), VECM can be formulated as

$$\Delta X_t = \delta_i + \sum_{i=1}^p \alpha_i \Delta X_{t-i} + \sum_{i=1}^p \beta_i \Delta Y_{t-i} + \gamma_1 \hat{\varepsilon}_{1t-1} + u_{1t}$$

$$\Delta Y_t = \lambda_i + \sum_{i=1}^p d_i \Delta X_{t-i} + \sum_{i=1}^p c_i \Delta Y_{t-i} + \gamma_2 \hat{\varepsilon}_{2t-1} + u_{2t}$$

Where, $\hat{\varepsilon}_{1t-1}$ and $\hat{\varepsilon}_{2t-1}$ are the error correction terms gained from the long run model, which can be interpreted as the deviation of X and Y from their long run equilibrium values respectively. The error correction terms is seen to understand the short-run dynamics which is in fact required to reach the long run equilibrium. The coefficient γ_i helps us to identifying the convergence rate of previous period disequilibrium of the system, i.e., the speed of the adjustment towards the long-run equilibrium relationship between variables used. β_i Measures the short run impact of changes in Y on X, d_i measures the short run impact of changes in X on Y, and u_{it} represent the standard error term.

4.2.1.6 Granger Causality Test

Correlation does not always necessarily imply causation in any meaningful sense. Sometimes the correlation might be magnificent correlations, which are simply spurious or meaningless. The study will apply Granger Causality test to determine the direction of causal relationship between GDP and different form of cross border investments. It can be conducted in two different ways, depending on the results of the long-run analysis. The Granger test (1969) is suitable for analyzing the short-run causal relationship if no cointegration exists among the variables. However, if the variables are co-integrated we apply the Engle and Granger (1987) test in place of the standard Granger test. It should be noted that the concept of causality in the Granger test does not mean that changes in one variable cause changes in another variable. The Granger test only tests whether predictability exists among the variables. The test examines whether the lagged values of one variable, say X, along with the lagged or

past values of Y, have better predictive power than that of the lagged or past values of Y only. In other words, the test expounds that X causes Y if Y can be better prognosticate by including past values of X in the model rather than using only Y's past values. The Granger test (1969) is appropriate when the long-run analysis implies there is no long-run relationship between variables that are integrated in the same order, that is, X and $Y \sim I(1)$. The following models have been estimated in order to measure the direction of causality.

$$X_t = \mu_1 + \sum_{i=1}^p \alpha_i X_{t-i} + \sum_{j=1}^p \beta_j Y_{t-j} + \epsilon_{1t}$$

$$Y_t = \mu_2 + \sum_{i=1}^p \gamma_i Y_{t-i} + \sum_{j=1}^p \delta_j X_{t-j} + \epsilon_{2t}$$

In the models above μ_1 and μ_2 are constants, the subscripts t and p signifies time period and optimum number of lag used in the model. ϵ_{1t} and ϵ_{2t} are the error terms observed by the model and it is assumed that the error terms are independent from each other. Based on the OLS coefficient estimates the null hypotheses can be tested to find out the direction of the relationship between the macro economic variables. If $\sum_{j=1}^p \beta_j = 0$ and $\sum_{j=1}^p \delta_j = 0$, it can be concluded that X and Y do not boost to predict one another. There exists a complementary relationship between the two variables X and Y, which we call both way causality, when $\sum_{j=1}^p \beta_j$ and $\sum_{j=1}^p \delta_j$ are both significantly different from zero. In the case where $\sum_{j=1}^p \beta_j = 0$ but $\sum_{j=1}^p \delta_j \neq 0$, unidirectional Granger causality exists from X to Y, but not vice versa. In other words, changes in X can help to forecast future values of Y, but Y cannot help to predict future values of X. Finally, the reverse relationship is true when $\sum_{j=1}^p \beta_j \neq 0$ and $\sum_{j=1}^p \delta_j = 0$, where changes in Y can help to estimate future values of X but not the other way around.

The study has tested the long-term causality by using the significance of error correction term of VECM and the short-term causality among the variables are tested through VEC Granger causality test or Block Exogeneity Wald test. On the other hand, if the variables are not cointegrated, then the study uses VAR Granger causality test to decipher the direction of short-run causality.

4.2.1.7 Impulse Response Function Analysis (IRF)

So far we discuss that Granger Causality test infer about the direction of causality among the study variables, but it fails to decipher the magnitude and the direction of causality at different time points. Where, Impulse Response Function (IRF) explains the reactions on present and future values of endogenous variable of one standard deviation shock to one of the innovations. The estimated impulse response of the VAR system enables us to examine how each of the variables responds to innovations from other variables in the system. To be specific, impulse response functions essentially map out the dynamic response path of a variable owing to a one standard deviation shock to another variable. The impulse response analysis is a useful tool for determining the degree, direction, and the duration of time that the variables in the system are affected by a shock to another variable. Therefore, IRF trace the dynamic effects of structural shocks in the VAR system. To The estimation of impulse response functions needs the VAR model to be transformed into the vector moving average (VMA) representation. The VAR model of X and Y is as follows:

$$Y_t = \sigma + \sum_{i=1}^k \beta_i Y_{t-i} + \sum_{j=1}^k \varphi_j X_{t-j} + u_{1t}$$

$$X_t = \alpha + \sum_{i=1}^k \beta_i Y_{t-i} + \sum_{j=1}^k \varphi_j X_{t-j} + u_{2t}$$

Where, u's denote the stochastic error terms commonly called innovations.

However, the residuals generated by the VAR models are usually contemporaneously correlated. The impulse responses derived from the initial estimates of the VAR model can be affected such that any adjustment to the order in which the variables are entered in the system could produce different results. Thus, it is highly required to order some restrictions while estimating the VAR model for identifying the IRFs. In order to impose some restrictions, consider a common approach namely Cholesky decomposition of recursive ordering, which was originally applied by Sims (1980). The Cholesky decomposition overcomes the problem of contemporaneous relationships among the VAR generated innovations.

4.2.1.8 Variance Decomposition Test (VDC)

Despite the importance of conducting causality tests, the empirical inferences based on the causality test is unable to determine neither the strength of the causal relationships between the variables nor do they describe the relationship between these variables over time. Moreover, the Granger causality test described above can only indicate the existence or non-existence of Granger causality within the sample period. It is incapable to gauge the strength of the causal relationship between the variables, and it does not provide an enlightenment of the dynamic properties of the system beyond the sample period. Variance decomposition determines the percentage of the forecast error of the real GDP is explained by cross border investment in the system and vice versa. This test is used to explore the degree of exogeneity of the associated variables. It explains the portion of the forecast error of one variable as a result of changes in the other variables over time. Hence, the relative significance of each variable can be measured, which causes oscillations in the other variable.

For any variable, short-run variations are due to its own shocks, but over time, shocks in other variables contribute to these changes as well. Forecast error variance

decomposition is a method available to examine this interesting phenomenon. In a point of fact, IRFs forecast the dynamic behavior of the model variables due to unrehearsed shocks within a VAR model; whereas, variance decompositions forecast the magnitude of relative importance of each innovation to the variables in the system. That is, variance decompositions can be considered similar to R^2 values associated with the dependent variables in different horizons of impulses. The study considers Choleski algorithm to estimate forecast error variance decomposition as follows:

The h-step forecast error for the y_t variables in terms of structural innovations $\varepsilon_t = (\varepsilon_{1t}, \dots, \varepsilon_{Kt})'$ can be shown to be

$$\Psi_0 \varepsilon_{T+h} + \Psi_1 \varepsilon_{T+h-1} + \dots + \Psi_{h-1} \varepsilon_{T+1},$$

so that the k_{th} element of the forecast error vector is

$$\sum_{n=0}^{h-1} (\psi_{k1,n} \varepsilon_{1,T+h-n} + \dots + \psi_{kK,n} \varepsilon_{K,T+h-n})$$

where $\psi_{ij,n}$ denotes the ij th element of ψ_n (see Lütkepohl (1991)). Because, by construction, the ε_{kt} are contemporaneously and serially uncorrelated and have unit variances, the corresponding forecast error variance is

$$\sigma_k^2(h) = \sum_{n=0}^{h-1} (\psi_{k1,n}^2 + \dots + \psi_{kK,n}^2) = \sum_{j=1}^K (\psi_{kj,0}^2 + \dots + \psi_{kj,h-1}^2)$$

The quantity $(\psi_{kj,0}^2 + \dots + \psi_{kj,h-1}^2)$ is interpreted as the contribution of variable j to the h-step forecast error variance of variable k . This interpretation is justified if the ε_{it} can be viewed as shocks in variable i . The percentage contribution of variable j to the h-step forecast error variance of variable k is obtained by dividing the above terms by $\sigma_k^2(h)$,

Therefore, $\omega_{kj}(h) = (\psi_{kj,0}^2 + \dots + \psi_{kj,h-1}^2) / \sigma_k^2(h)$.

The corresponding estimated quantities are often reported for various forecast horizons.

Besides, to ensure the models are not misspecified, the study applies serial correlation test, normality test, heteroskedasticity test and stability test using cumulative sum of recursive residuals (CUSUM) proposed by Brown et al. (1975). Notably, Microsoft Office Excel 2007, Eviews-7 and Eviews-10 package are used for econometric analyses.

4.2.1.9 Recursive Residuals and the CUSUM Test

Visual examination of the graphs of the recursive parameter estimates can be useful in evaluating the stability of the model. It would be useful to have a formal statistical test that we could apply to test the null hypothesis of model stability. The CUSUM test, which is concentrated on the residuals, obtained from the recursive estimates.

As an example, we calculate a statistic, called the CUSUM statistic, for each t . Under the null hypothesis, the statistic is drawn from a distribution, called the CUSUM distribution. If, the calculated CUSUM statistics appear to be too large to have been drawn from the CUSUM distribution, we reject the null hypothesis (of model stability). Technically:

Let $e_{t+1,t}$ denote the one-step-ahead forecast error associated with forecasting Y_{t+1} based on the model fit for over the sample period ending in period t . These are called the **recursive residuals**.

$$\begin{aligned} e_{t+1,t} &= Y_{t+1} - \hat{Y}_{t+1,t} \\ &= Y_{t+1} - [\hat{\alpha}_{0,t} + \hat{\alpha}_{1,t}(t+1) + \dots + \hat{\alpha}_{s,t}(t+1)^s + \hat{\phi}_{1,t}Y_t + \dots + \hat{\phi}_{p,t}Y_{t-p+1}] \end{aligned}$$

where the t subscripts on the estimated parameters refers to the fact that they were estimated based on a sample whose last observation was in period t .

4.2.2 Scheme of Investigation

Given the nature of the problem and the quantum of data, we first study the data properties from an econometric perspective with the help of descriptive statistics and unit root test to show the nature and basic characteristics of the variables used in the analysis and to find out whether the data series are stationary or non-stationary. The study applies the commonly used ADF, PP, and KPSS unit root tests to determine the stationarity properties or integration order of the variables. Briefly stated, a variable is said to be integrated of order n , written $I(n)$, if it requires differencing n times to achieve stationarity. Thus, the variable is nonstationary if it is integrated of order 1 or higher. Classification of the variables into stationary and non-stationary variables is crucial for applying standard time series econometric tests. The final decision regarding the unit root property could be taken by considering the two popularly used unit root test results, namely, the ADF and PP tests. Any contradiction arises among the two results derived from two different unit root tests, then for this case, the decision regarding the unit root property can be taken with the help of the unit root result obtained from KPSS test.

As the autoregressive model is sensitive to the selection of appropriate lag length, the study ascertains the appropriate lag length prior to estimation. The study has determined the optimum lag length based on AIC, SIC, and HQC and finally, the study uses SIC criteria for optimum lag length selection in each case.

To determine the long-run relationship between the different forms of foreign investments and volume of country's GDP with sectoral specification, the study

considers Johansen cointegration test or Autoregressive Distributed Lag Model (ARDL). The study applies the VAR-based approach of cointegration test suggested by Johansen (1988) and Johansen and Juselius (1990) if the variables are non-stationery in level and are integrated of any same higher-order. Appropriately, the test provides us with information as to whether the variables are move together in the long run. But when some variables are stationery in level, that is, $I(0)$, and some are stationary at their first difference or both the variables are stationary at their level then we apply a bound test of cointegration under Autoregressive Distributed Lag Model framework to explore the relationship between them.

After that, the nature of the relationship between the study variables in the short-run is tested through the VECM or ARDL estimation. If there are no cointegrating vectors the variables, then the model will be used to capture short-run relationship between the variables. However, if one or more cointegrating vectors do exist in the VAR-based approach of cointegration test suggested Johansen and Juselius (1990) the VECM will be employed instead of the normal VAR model. The error correction term of VECM specification indicates the rate at which it corrects its previous period disequilibrium or speed of adjustment to restore the long-run equilibrium relationship. The study proceeds with a Granger causality test in the form of VECM, when the variables are found to be cointegrated, that is, the long-run relationship exists among variables. The Granger causality test is performed to identify the existence and nature of the causal relationship between the variables. VECM allows the modeling of both the short-run and long-run dynamics for the variables involved in the model. The error correction term of VECM indicates the direction of long-run causality, and short-term causality among the variables are tested through VEC Granger causality test or Block Exogeneity Wald test. Nevertheless, if the variables are not cointegrated, then the

study uses VAR Granger causality test to decipher the direction of short-run causality. Impulse Response Function and Variance Decomposition Analysis have been used to examine how the stock prices respond to a sudden change (i.e., shocks) or innovations in the macroeconomic variables considered in the study. The IRFs analyze the dynamic behavior of the target variables due to unanticipated shocks within a VAR model and variance decompositions determine the relative importance of each innovation in the variables in the system.

Finally, the investigation applies some diagnostic test namely, serial correlation test (based on Lagrange Multiplier Test of Residual Serial Correlation), normality test (based on a test of Skewness and Kurtosis of Residuals, Jarque-Bera test of Normality), and heteroskedasticity test (based on the White Heteroskedasticity Test with no Cross Terms Yields). Also, we apply CUSUM test (based on Cumulative Sum of Recursive Residuals) for stability analysis.