RESEARCH METHODOLOGY

Although the FDI inflows into India have increased steadily, it was still less than one third of FDI into China in the year 2008. The two Asian giants India and China have made considerable change in the area of trade and foreign investment. Both the countries have population in excess of a billion and ability to become superpowers. Therefore, a comparison of the two is of interest in itself. However, most of the studies have compared India and China on the basis of some demographic, social and prosperity indicators like population, birth rate, death rate, life expectancy and adult literacy rate etc. But none of the studies have found the impact of FDI on macroeconomic indicators like exports, Foreign Exchange Reserves, GDP (Gross Domestic Product), Gross capital formation (GCF) and Employment in India and China. So, this study endeavours to find out the impact of FDI on these macro economic parameters in India and China.

3.1 Objectives of Study

This research work aims at accomplishing the following specific objectives:

1. To study the global trends of FDI.
2. To compare the recent trends of FDI in India and China.
3. To examine the factors which influence the inflows of FDI in India and China.
4. To analyse the impact of FDI on related macroeconomic indicators in India and China.
5. To investigate the causal links between FDI and trade in both the countries.

3.2 Study Period, Variables and Data Collection

The present study is based on secondary data, which has been collected from published documents like World Investment Report, Economic Survey, China Statistical Yearbook, SIA Newsletter and Handbook of Statistics of Indian Economy. In order to study the trends of Foreign Direct Investment, the annual data for the period 1991-2008 for India as well as China have been used. Quarterly data for the period 1990-91 to 2008-09 and annual data for the period 1976 to 2008 has been taken for examining the determinants of FDI in India and China respectively. Various explanatory variables used to achieve this objective are Gross domestic product
(GDP), Long term debt (LTD), Inflation (INF), Exchange rate (EXCH), Openness (OP) and Foreign exchange reserves (RES). Moreover to analyze the impact of FDI on macroeconomic indicators such as Gross domestic product (GDP), Gross capital formation (GCF), Exports (EXP), Foreign exchange reserves (RES) and Employment (EMP) and to investigate the causal links between FDI and trade i.e Imports (IMP) and Exports (EXP) annual data for the period 1976-2008 for both the countries have been used.

3.3 Hypotheses taken in the Study

This section describes the various hypotheses to be tested for achieving the above mentioned objectives of the study. These are categorised as under keeping in view the specific objectives.

(a) For examining the factors which influence the flow of FDI in India

- Impact of GDP (Gross Domestic Product) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

- Impact of Long term debt (LTD) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

- Impact of Inflation (INF) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

- Impact of Exchange rate (EXCH) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

- Impact of Openness (OP) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

- Impact of Foreign exchange reserves (RES) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

(a) For examining the factors which influence the flow of FDI in China

- Impact of Long term debt (LTD) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.
- Impact of Exchange rate (EXCH) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

- Impact of Foreign exchange reserves (RES) on Foreign Direct Investment (FDI) is insignificant against the alternative hypothesis of significant impact.

(b) To analyze the impact of FDI on related macroeconomic indicators in India and China.

- Impact of FDI on Gross domestic product (GDP) is insignificant against the alternative hypothesis of significant impact.

- Impact of FDI on Gross capital formation (GCF) is insignificant against the alternative hypothesis of significant impact.

- Impact of FDI on Exports (EXP) is insignificant against the alternative hypothesis of significant impact.

- Impact of FDI on Foreign exchange reserves (RES) is insignificant against the alternative hypothesis of significant impact.

- Impact of FDI on Employment (EMP) is insignificant against the alternative hypothesis of significant impact.

(c) For investigating the causal links between FDI and trade in India and China

- Imports do not cause FDI

- FDI does not cause Imports

- Exports do not cause FDI

- FDI does not cause Exports

- Imports do not cause Exports

- Exports do not cause Imports

3.4 Data Analysis

Various Statistical and Econometric techniques have been employed to analyse the collected data such as Compound Annual Growth Rate (CAGR),
Cointegration technique and Granger Causality tests. Augmented Dickey fuller (ADF) test has also been used to check the stationarity of data series and to find out the optimum lag structure. Computer softwares such as SPSS version 18 and Eviews 6 have been used to apply these techniques. The brief description of the techniques applied is as follows:

3.4.1 Compound Annual Growth Rate

The Compound Annual Growth Rate (CAGR) has been calculated by using the semi log equation and applying the method of ordinary least square (OLS) as explained below:

Let \( Y = ab^t \) \hspace{1cm} (3.1)

\[ \log Y = \log(ab^t) = \log a + t \log b \]

Or \( \log Y_t = A + Bt \) (Here \( A = \log a, B = \log b \))

Or \( \log Y_t = A + Bt + u_t \) \hspace{0.5cm} (t = 1,2,...........,n) \hspace{0.5cm} (3.2)

Where \( Y_t \) = \( t \) th observation on the variable \( Y \)

\( t \) = Time variable taking \( n \) values \( 1,2,......n \),

\( u_t \) = A random disturbance (or error) term at time \( t \) satisfying the usual assumptions of OLS

\( r \) = compound annual rate of growth.

\( a \) and \( b \) are parameters of the original model (3.1) and \( A \) and \( B \) are the parameters of transformed model (3.2)

From (3.2), \( a \) and \( b \) are calculated as follows:

\( \log a = A \), \( a = \text{Antilog} \ A \)

\( \log b = B \), \( b = \text{Antilog} \ B \)

Since \( (1+r) = b \),

Since rate of growth is calculated in % terms,

\[ r = (b-1) \times 100 = (\text{Antilog} B) -1 \times 100 \] (Gujrati, 2004).
3.4.2 Stationarity of Time-Series

Stationarity is the first fundamental statistical property that should be tested in time series analysis, because most statistical models require that the underlying generating processes be stationary. A time series is covariance stationary when it has the following three characteristics:

1) Exhibits mean reversion in that it fluctuates around a constant long-run mean.
2) Has a finite variance that is time-invariant.
3) Has a theoretical correlogram that diminishes as the lag length increases.

In its simplest terms a time series \( y_t \) is said to be stationary if:

a) \( E(y_t) = \text{constant for all } t; \)

b) \( Var(y_t) = \text{constant for all } t; \)

c) \( Cov(y_t, y_{t+k}) = \text{constant for all } t \text{ and all } k \neq 0, \text{ or if its mean, variance and covariance remain constant over time. Stationarity is important because if the series is non-stationary then all the typical results of the classical regression analysis are not valid. Regression with non-stationary series may have no meaning and are therefore called spurious. In stationary time series, Shocks will be temporary and over time their effects will be eliminated as the series revert to their long-run mean values. There are some formal tests for checking the stationarity of time series on the basis of existence of unit roots. The problems regarding the existence of unit roots in the regression models is as follows:}

Consider the auto regressive model of order one:

\[
y_t = \phi y_{t-1} + u_t
\]

Where \( u_t \) is a white noise process and the stationarity condition is \( |\phi| < 1 \).

In general we have three possible cases:

**Case 1**: if \( |\phi| < 1 \), the series is stationary.

**Case 2**: if \( |\phi| > 1 \), in this case the series explodes.

**Case 3**: if \( |\phi| = 1 \), in this case the series contains a unit root and is non-stationary.
Testing for the order of integration

A test for the order of integration is a test for the number of unit roots, and it follows the steps described below:

**Step 1**: Test $y_t$ to see if it is stationary. If yes, then $y_t \equiv I(0)$; if no then $y_t \equiv I(n)$; $n>0$.

**Step 2**: Take first differences of $y_t$ as $\Delta y_t = y_t - y_{t-1}$, and test $\Delta y_t$ to see if it is stationary. If yes then $y_t \equiv I(1)$; if no then $y_t \equiv I(n)$; $n>0$.

**Step 3**: Take second differences of $y_t$ as $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$, and test $\Delta^2 y_t$ to see if it is stationary. If yes then $y_t \equiv I(2)$; if no then $\Delta^2 y_t = \Delta y_t - \Delta y_{t-1}$; $n>0$. Etc... till we find that it is stationary and then we stop. So, for example if $\Delta^3 y_t \equiv I(0)$, then $\Delta^2 y_t \equiv I(1)$ and $\Delta y_t \equiv I(2)$, and finally $y_t \equiv I(3)$; which means that $y_t$ needs to be differenced three times in order to become stationary.

The **Augmented Dickey-Fuller (ADF) test for unit root**

Dickey and Fuller extended their test procedure suggesting an augmented version of the Dickey-Fuller test which includes extra lagged terms of the dependent variable in order to eliminate autocorrelation. The lag length on these extra terms is either determined by the Akaike Information Criterion (AIC) or Schwartz Bayesian Criterion (SBC), or more usefully by the lag length necessary to whiten the residuals (i.e. after each case we check whether the residuals of the ADF regression are autocorrelated or not through LM tests and not the DW test).

The three possible forms of the ADF test are given by the following equations:

$$\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t$$

$$\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t$$

$$\Delta y_t = a_0 + \gamma y_{t-1} + a_2 t + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t$$

The difference between the three regressions again concerns the presence of the deterministic elements $a_0$ and $a_2 t$. The ADF-test statistic is the t statistic for the lagged dependent variable. If its statistical value is smaller in absolute terms than the critical
value then we reject the null hypothesis of a unit root and conclude that $y_1$ is a stationary process. This procedure is the most sensible way to test for unit roots when the form of the data-generating process is unknown (Chan, 2000).

### 3.4.3 Cointegration Analysis

If two or more than two variables are integrated of the same order $d$ where $d > 0$, and there exists a stationary linear combinations of these variables, the variables are said to be cointegrated. The notion of cointegration was introduced by Granger (1981, 1983) and Engle and Granger (1987) to explain stationary equilibrium relationship among the non-stationary variables. This revolutionary notion has radically changed the way empirical models of economic relationships are formulated today.

The test of cointegration basically tests whether there exists a stationary linear combination of non-stationary variables. If such combination is found, it is inferred as an equilibrium relationship between the variables. This equilibrium relationship is then used to construct an Error Correction Model (ECM) which is a statistical specification of economic dynamics through which the pull and push forces restore the equilibrium relationship whenever a disequilibrium takes place. ECM captures both the short-term and the long-run dynamics of cointegrating variables. The Granger representation theorem shows that if two (or more than two) variables are cointegrated their relationship can appropriately be represented as an ECM. Therefore, it will not be appropriate if we use differenced variables in a regression model without including ECM, when the variables at the level are cointegrated.

#### Testing for Cointegration

There are two approaches to test the cointegration in a given equation:

1) **Cointegration in single equations: the Engle-Granger approach**

Engle and Granger further formalized concept of cointegration by introducing a very simple test for the existence of cointegrating (i.e. long-run equilibrium) relationships in bi-variate analysis.

**Engle-Granger Test for Cointegration:**

i. Test whether both the variables $x_t$ and $y_t$ follow I(1) process.

ii. Estimate a regression equation using variables at level, such as

$$y_t = \alpha + \beta x_t + u_t \quad \text{------------------- (3.7)}$$
iii. Test whether $u_t$ is stationary. We can use ADF test for this purpose. But the critical values for this test will change. The modified critical values of ADF-test in context to cointegration are estimated by Engle and Yoo (1987).

iv. If $u_t$ is stationary, the above regression between $x_t$ and $y_t$ is a cointegrating equation. In this situation $\beta$ is called cointegration parameter and $(1, -\beta)$ is called cointegration vector.

Engle-Granger Error Correction Model (ECM):

In its simplest form the bi-variate Engle-Granger Error Correction Model (ECM) for above variables $x_t$ and $y_t$ can presented as follows:

\[
\Delta y_t = \theta_1 u_{t-1} + \epsilon_t \\
\Delta x_t = \theta_2 u_{t-1} + \epsilon_t
\]

(3.8)  
(3.9)

Since $u_{t-1}$ represents unexpected move in $y_{t-1}$ to restore the equilibrium either $y_t$ will move in opposite direction of $u_{t-1}$ or $x_t$ will move in the direction of $u_{t-1}$. Therefore, the expected sign of $\theta_1$ is negative and that of $\theta_2$ is positive.

If we also accommodate the short-term dynamics of the variables in ECM, it can be presented as follows:

\[
\Delta y_t = \alpha_{10} + \theta_1 u_{t-1} + \sum_{i=1}^{p} \gamma_{1i} \Delta y_{t-i} + \sum_{j=1}^{p} \delta_{1j} \Delta X_{t-j} + \epsilon_{1t} \\
\Delta x_t = \alpha_{20} + \theta_2 u_{t-1} + \sum_{i=1}^{p} \gamma_{2i} \Delta y_{t-i} + \sum_{j=1}^{p} \delta_{2j} \Delta X_{t-j} + \epsilon_{2t}
\]

(3.10)  
(3.11)

This formulation is essentially a VAR model augmented by error correction term (Cromwell et al., 2007)

2) Cointegration in multiple equations and the Johansen approach

If there are more than two variables in the model, then there is the possibility of having more than one cointegrating vector. By this we mean that the variables in the model might form several equilibrium relationships governing the joint evolution of all the variables. In general for $n$ number of variables we can have only up to $n-1$
Having n=2 and assuming that only one cointegrating relationship exists, where there are actually more than one, is a very serious problem that cannot be resolved by the EG single equation approach. Therefore, an alternative to the EG approach is needed and this is the Johansen approach for multiple equations.

In order to present this approach, it is useful to extend the single equation error correction model to a multivariate one, let’s assume that we have three variables, \( Y_t \), \( X_t \) and \( W_t \) which can all be endogenous, i.e. we have that (using matrix notation for \( Z_t = [Y_t, X_t, W_t] \))

\[
Z_t = A_1 Z_{t-1} + A_2 Z_{t-2} + \ldots + A_k Z_{t-k} + u_t \tag{3.12}
\]

It can be reformulated in a vector error correction model (VECM) as follows:

\[
\Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \Gamma_2 \Delta Z_{t-2} + \ldots + \Gamma_{k-1} \Delta Z_{t-k-1} + \Pi Z_{t-1} + u_t \tag{3.13}
\]

Where \( \Gamma_i = (I - A_1 - A_2 - \ldots - A_k) \) and \( \Pi = -(I - A_1 - A_2 - \ldots - A_k) \).

Here we need to carefully examine the \( \Pi \) matrix of order \( 3 \times 3 \). The \( \Pi \) matrix contains information regarding the long run relationships. We can decompose \( \Pi = \alpha \beta' \) where \( \alpha \) will include the speed of adjustment to equilibrium coefficients while \( \beta' \) will be the long run matrix of coefficients. Therefore the \( \beta' Z_{t-1} \) term is equivalent to the error correction term \( (Y_{t-1} - \beta_0 - \beta X_{t-1}) \) in the single equation case, except that now \( \beta' Z_{t-1} \) contains up to \( (n-1) \) vectors in a multivariate framework.

The steps of the Johansen approach in practice are as follows:

**Step 1: Testing the order of integration of the variables**

As with EG approach, the first step in the Johansen approach is to test for the order of integration of the variables under examination.

**Step 2: Setting the appropriate lag length of the model**

The issue of finding appropriate lag length is very important because we want to have Gaussian error terms. The most common procedure in choosing the optimal lag length is to estimate a VAR model including all our variables in levels. This VAR model should be estimated for a large number of lags, then reducing down
by re-estimating the model for one lag less until we reach zero lags. Model that minimizes AIC and SBC is selected as the one with the optimal lag length.

**Step 3: Choosing the appropriate model regarding the deterministic components in the multivariate system.**

Another important aspect in the formulation of the dynamic model is whether an intercept and/or a trend should enter in either the short run or the long run model, or both models. The general case of the VECM including all the various options that can possibly happen is given by the following equation:

\[
\Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_k \Delta Z_{t-k-1} + \alpha \begin{pmatrix} \beta \\ \mu_1 \\ \delta_1 \end{pmatrix} \begin{pmatrix} Z_{t-1} & 1 & t \end{pmatrix} + \mu_2 + \delta_2 t + u_t \quad \text{---------- (3.14)}
\]

And for this equation we can see the possible cases. We can have a constant (with coefficient \( \mu_1 \)) and/or a trend (with coefficient \( \delta_1 \)) in the long run model (the cointegrating equation (CE)), and a constant (with coefficient \( \mu_2 \)) and/or a trend (with coefficient \( \delta_2 \)) in the short run model (the VAR model). In general five distinct models can be considered. The first and the fifth are not that realistic therefore the problem reduces to a choice of one of the three remaining models.

**Step 4: Determining the rank of \( \Pi \) or the number of cointegrating vectors**

According to Johansen and Juselius (1990), there are two methods for determining the number of cointegrating relations. The first test statistic is based on the maximum eigenvalue and because of that is called the maximum eigenvalue statistic. The second method is based on a likelihood ratio test about the trace of the matrix because of that it is called the trace statistic.

**Step 5: Testing for linear restrictions in the cointegrating vectors**

An important feature of the Johansen approach is that it allows us to obtain estimates of the coefficients of the matrices \( \alpha \) and \( \beta \), and then test for possible linear restrictions regarding those matrices. Especially for matrix \( \beta \), the matrix that contains the long run parameters, this is very important because it allows us to test specific hypothesis regarding various theoretical predictions from an economic theory point of view (Harris and Sollis, 2006).
3.4.4 Granger Causality Test

The concept of the causality was initially defined by Granger (1969). In the current study, Granger Causality tests have been conducted to examine possible casual relationships among three variables: FDI, Exports and Imports in India as well as in China separately. The present study employs data that consist of annual observations during the period 1976-2008. The definition of Granger Causality is based on the assumption of stationarity. The econometric methodology examines the stationarity properties of the data series. The Granger Causality test has been performed to analyse the causal links between above mentioned variables. Then firstly following unrestricted equation with n lags is estimated by using OLS.

\[ F_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i F_{t-i} + \sum_{j=1}^{l} \beta_j E_{t-j} + \varepsilon_t \]  \hspace{1cm} \text{(3.15)}

\[ E_t = \omega + \sum_{i=1}^{l} \gamma_i E_{t-i} + \sum_{j=1}^{l} \theta_j F_{t-j} + \varepsilon_t \]  \hspace{1cm} \text{(3.16)}

Equation 3.15 postulates that current exports is related to past values of exports itself as well as of FDI and equation 3.16 postulates a similar behaviour for FDI similarly in order to study the granger causality between FDI and imports; exports and imports following sets of equations can be framed.

\[ F_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i F_{t-i} + \sum_{j=1}^{l} \beta_j I_{t-j} + \varepsilon_t \]  \hspace{1cm} \text{(3.17)}

\[ I_t = \omega + \sum_{i=1}^{l} \gamma_i I_{t-i} + \sum_{j=1}^{l} \theta_j F_{t-j} + \varepsilon_t \]  \hspace{1cm} \text{(3.18)}

\[ I_t = \alpha_0 + \sum_{i=1}^{l} \alpha_i I_{t-i} + \sum_{j=1}^{l} \beta_j E_{t-j} + \varepsilon_t \]  \hspace{1cm} \text{(3.19)}

\[ E_t = \omega + \sum_{i=1}^{l} \gamma_i E_{t-i} + \sum_{j=1}^{l} \theta_j I_{t-j} + \varepsilon_t \]  \hspace{1cm} \text{(3.20)}

Where F stands for FDI, E for exports and I for Imports.
To test the null hypothesis $F$ test is conducted by comparing the respective residual sum of squares, which are given as:

$$RSS_i = \sum_i \mu_i^2; \quad RSS_0 = \sum_i \hat{\epsilon}_i^2$$

The test statistics given as:

$$F = \frac{\frac{(RSS_0 - RSS_i)}{P}}{RSS/(T=2P-1)} \cdot \frac{FPIT-2p-1}{FPIT-2p-1}$$

It follows the $F$ distribution with $m$ and $(n-k)$ degrees of freedom. If the computed value of $F$ is greater than the specified critical value then the null hypothesis is rejected (Gujrati, 2004).