CHAPTER–III
DATA BASE AND METHODOLOGY

The purpose of this chapter is to discuss the data base and methodology used in the present study.

Data Base

The data used in this study were obtained from various authentic publications and websites as per details given below:

(1) Data on balance of payments of India and its various components were culled out from various issues of *Handbook of Statistics on Indian Economy*, Reserve Bank of India, *Economic Survey*, Government of India and *EPW* Research Foundation.

(2) Data on reserve money supply of India, fiscal deficit, gross national product, net barter terms of trade, nominal exchange rate in terms of $, foreign exchange reserves, foreign currency assets, infrastructure index, inflation rate in India, net domestic savings and net domestic capital formation were obtained from various issues of *Handbook of Statistics on Indian Economy*, Reserve Bank of India.

(3) Data on industrial sector, gross domestic product index of India, India’s GDP in financial sector, unit value index of exports, wholesale price index were also obtained from *Handbook of Statistics on Indian Economy*, Reserve Bank of India. All these variables were expressed with base 1999-00=100.

(4) Data on external debt and consumer price index (CPI) were obtained from various issues of *Economic Survey*, Government of India.

(5) Data on export promotion expenditure, export credit and returns in domestic market were obtained from *National Income Statistics and Capital Market Issues*, Center of Monitoring Indian Economy (CMIE) Report, Economic Intelligence Service.

(6) Data on consumer price index abroad and inflation rate abroad were obtained from *International Financial Statistics Yearbook*, I.M.F.
(7) Data on debt service ratio and debt interest were obtained from various issues of *World Bank Tables, Global Development Finances* and *World Development Indicators*, World Bank.

(8) Data on long term interest rate of U.S. economy was obtained from online source: www.eh.net.in (European Economic History Service).

**Period of Study**

Attempts were made to collect data for maximum number of independent variables depending upon their availability since 1980. The study covered the period 1980-81 to 2005-06. However, to make the comparative study of balance of payments position and to know the impact of liberalization and globalization on balance of payments since 1991, the whole period has been bifurcated into two sub-periods i.e. pre-liberalization period (1980-81 to 1989-90) and post-liberalization period(1992-93 to 2005-06). The two years i.e. 1990-91 and 1991-92 have been excluded from the analysis due to economic crisis of 1991 and abrupt changes in 1991-92. However, to know the cointegration relationship between the various components of balance of payments and economic growth via GDP, the data has been taken for period 1973-74 to 2005-06 as the application of technique requires minimum 30 years time series data.

In order to access real growth in international transactions, attempts have been made to estimate value figures of transactions at constant prices. Since it has been difficult to get the appropriate deflators for all the components of balance of payments, so the whole data has been deflated with wholesale price index (WPI) with 1999-00 as base. However, to deflate the external factors affecting the balance of payments, suitable deflators have been used.

**Methodology**

In order to meet the specified objectives, several statistical techniques were applied to analyze the data as discussed below:

(i) **Tabular Analysis**

Tabular analysis technique was largely used. Ratios, percentages, averages etc. were also worked out.
(ii) Rate of Change

To know the annual changes in balance of payments and its various components, annual rate of change was calculated as following:

\[ X_t = \frac{X_t - X_{t-1}}{X_{t-1}} \times 100 \]

\[ X_t = \text{item in current year} \]
\[ X_{t-1} = \text{item in previous year} \]

(iii) Trend Analysis

To know the trends in balance of payments over a period of 26 years (1980-81 to 2005-06) and for two sub-periods (1980-81 to 1989-90) and (1992-93 to 2005-06), trend analysis has been calculated:

\[ y = a + bt \]

Where \( t \) is time period, ‘a’ is constant and ‘b’ is the linear time trend
\( y \) is the dependent variable.

These trend values were also tested for significance by applying \( t \)-test.

(iv) Compound Annual Growth Rate

To analyze the changes over time in balance of payments and its components, compound annual growth rates have been calculated by using the following equation:

\[ y_t = a b^t e^{ut} \]

Transforming the above equation in log linear form

\[ \log Y = \log a + t \log b + u_t \]

Where \( Y_t \) is value of dependent variable in year \( t \), \( t \) is trend variable, \( U_t \) is stochastic disturbance term and ‘a’ and ‘b’ are constants.

From the estimated regression coefficients ‘b’, the rate of growth ‘r’ can be calculated as follows:

\[ r = \{\text{antilog}(\hat{b} - 1)\} \times 100 \]
(v) **Correlation Analysis**

In order to know the inter-correlation of various independent variables or determinants of balance of payments, correlation coefficients have been calculated. This will help us to know the most important variables in explaining variations in balance of payments position of India. To test the significance of correlation coefficients, t-test has been applied.

\[ t = \frac{|r_{ij}|}{\sqrt{1 - (r_{ij})^2}} \sqrt{n - 2} \]

Where \( n \) is number of observations (years) and \( r_{ij} \) is the correlation coefficient between \( i^{th} \) and \( j^{th} \) variables.

(vi) **Multiple Regression Analysis**

Linear regression equations were fitted by regressing dependent variable on each of the independent variables and including time variable separately as:

\[ Y = \hat{\alpha}_0 + \hat{\beta}_1 t + \hat{\beta}_2 X_i \]

(\( \text{where } i = 1, 2, 3 \ldots \ldots n \))

Where \( Y \) is dependent variable.

\( X_i \) = independent variable.

\( t \) = time variable.

Where \( \hat{\alpha}_0 \) is estimate of \( \alpha_0 \) and \( \hat{\beta}_1 \) is estimate of \( \beta_1 \).

The statistical significance of \( \beta' \)'s were tested by applying student’s t-test and \( r^2/R^2 \) was computed to see how much percentage variations a particular independent variable explains in the dependent variable.

(vii) **Step Down Regression Analysis**

In order to know the effect of various variables on balance of payments position of India, a disaggregated structural model of India’s BOP was set up. The specified variables which were considered as determinants of different components of balance of payments, are given below:
(1) Reserve Money Supply (ResMs)
(2) Fiscal Deficit (FD)
(3) GNP at constant prices (GNP)
(4) Net Barter Terms of Trade (TOT)
(5) Nominal Exchange Rate in terms of $ (ER)
(6) Foreign Exchange Reserves (FER)
(7) World GDP Index (WGDP_idx)
(8) External Debt (ED)
(9) Foreign Currency Assets (FCA)
(10) Export Promotion Expenditure (XP_e)
(11) Export Credit (ExCr)
(12) Infrastructure Index (IIndx)
(13) Relative Consumer price Index (RCPI)
(14) India’s Government Expenditure Index (IGEindx)
(15) Oil Prices ($ per barrel) (OP)
(16) Industrial Sector GDP index of India (IGDP_idx)
(17) Growth rate of India’s GDP in financial sector (GGDP_fs)
(18) India’s Industrial Production Index (IIPI)
(19) Growth Rate of Net Foreign Currency Assets (FCA_gt)
(20) India’s Inflation rate in Percent (IInf)
(21) Wholesale Price Index (WPI)
(22) Long Term U.S. Interest rate in Percent (ROI_Abroad)
(23) Gross Domestic Product at factor cost at constant Prices (GDP_FC)
(24) Inflation rate prevailing abroad(U.S. economy) in percent (Inf_abroad)
(25) Domestic Rate of Interest (ROI)
(26) Average Effective Interest rate on external debt in percent (Debt Int)
(27) Export Unit Value Index (EUI): This variable is computed by deflating the unit value index of exports by nominal exchange rate.
(28) Export Profitability (EP): The ratio of export unit value index to domestic wholesale price index captures the profitability of exports. Export unit value indices and domestic wholesale price indices were taken a base 1999-00=100.
(29) Gross Fiscal Deficit Ratio (GFD_ratio) : Ratio of gross fiscal deficit to gross domestic product at factor cost
Openness Ratio (OPEN): This variable shows the degree of openness of the economy. This is defined as \( \frac{\text{Exports} + \text{Imports}}{\text{GDP}_{FC}} \)

Returns in Domestic market (RBSE): Returns in BSE Sensex taken as a proxy variable.

Returns in Foreign Market (RSP): Returns in composite S & P 500 index taken as a proxy variable.

Interest Rate Differential (Int Diff): This is computed by taking the difference of domestic interest rate and interest rate prevailing abroad.

Saving-Investment gap (S-I gap): This is computed by taking the difference between the net domestic savings and net domestic capital formation with 1999-00=100 as base.

Relative Price Ratio (RPR): This is computed by taking the ratio of import price index to wholesale price index.

Reserve Money supply as percent of GDP (ResMs/GDP)

Fiscal Deficit as percent of GDP (FD/GDP)

Growth rate of GNP (GGNP)

Foreign Exchange Reserves as percent of GDP (FER/GDP)

External Debt as percent of GDP (ED/GDP)

Lagged Current Account Balance as percent of GDP (LCAB/GDP)

Current Account Balance as a percent of GDP (CAB/GDP)

In addition to these independent variables, dynamics need also be included in the equations to account for lagged effects, such as previous receipts earned by country of destination. It is expected that demand is not only be influenced by current, but also by lagged values since changes in income may take some time to affect the tourism demand, transportation demand or demand of any goods and services. U.S. economy is chosen as a foreign country to model capital inflows in India, because US is our major trading partner and accounts for largest proportion of foreign institutional inflows to India. Hence US could be safely used as proxy for rest of the world to study these inflows to India (Rai et al., 2007).

To know the effect of these independent variables on BOP and its components, step-down multiple regression analysis has been used. The four set of equations i.e. balance of payments, current account, trade balance and capital account are free of units of measurement as variables included are standardized variables.
Balance of Payments

\[ \frac{BOP}{GDP} = a_0 + b_1 ResM_s/GDP + b_2 FD/GDP + b_3 GGNP + b_4 TOT + b_5 ER \]
\[ + b_6 FER/GDP + b_7 WDP_{max} + b_8 ED/GDP + b_9 D + b_{10} OPEN + b_{11} OP \]

It is hypothesized that \( b_1, b_5, b_4, b_6, b_7, b_9, > 0 \).

while \( b_2, b_3, b_8, b_{11} < 0 \) and \( b_{10} \leq 0 \).

Current Account

\[ \frac{CA}{GDP} = a_0 + b_1 TOT + b_2 ER + b_3 FD/GDP + b_4 GGNP + b_5 FCA/GDP + b_6 D \]
\[ + b_7 OPEN + b_8 ResMs/GDP + b_9 ED/GDP + b_{10} OP + b_{11} LCAB/GDP \]

It is hypothesized that \( b_1, b_2, b_5, b_6, b_8, b_{11} > 0 \).

while \( b_3, b_4, b_9, b_{10} < 0 \) and \( b_7 \leq 0 \)

Trade Balance Account

\[ \frac{TB}{GDP} = a_0 + b_1 TOT + b_2 GGNP + b_3 FER/GDP + b_4 FD/GDP + b_5 ER + b_6 WDP_{ind} \]
\[ + b_7 ED/GDP + b_8 OP + b_9 D \]

It is hypothesized that \( b_1, b_3, b_5, b_6, b_9 > 0 \).

while \( b_2, b_4, b_7, b_8 < 0 \).

Capital Account

\[ \frac{Cap Ac}{GDP} = a_0 + b_1 (S - I \ gap) + b_2 CA/GDP + b_3 GGNP + b_4 FCA + b_5 RCPI + b_6 D \]
\[ + b_7 WDP_{ind} + b_8 INF + b_9 ER + b_{10} ROI_{Abroad} \]

It is hypothesized that \( b_1, b_3, b_4, b_5, b_6, b_7 > 0 \).

while \( b_2, b_8, b_9, b_{10} < 0 \).

Given the objectives of the study, not only aggregated analysis of balance of payments has been done but a disaggregated analysis of the components of BOP has also been carried out. Allowing for further full disaggregation, the current account has been arrived at as a summation of the trade account and invisible account. The trade balance account comprises of two equations: merchandise
exports and merchandise imports. On the other side, on the receipt side, the invisible account has been broken into travel receipts, transportation receipts, insurance receipts, investment receipts, private transfer receipts and other investment receipts (include government not included elsewhere, official transfer receipts and miscellaneous account receipts). Invisible payments comprise of travel payments, transportation payments, insurance payments, investment income payments, private transfer payments and other invisible payments (which include government not included elsewhere, official transfer payments and miscellaneous account payments).

**DETERMINANTS OF CURRENT ACCOUNT**

(1) **Merchandise Trade**

(A) **Merchandise Exports (Exp)**

\[
Exp = \alpha_0 + b_1 \text{WGDP}_{\text{Indx}} + b_2 \text{ER} + b_3 \text{Exp}(-1) + b_4 \text{DEX Price} + b_5 \text{XP}_e + b_6 D \\
+ b_7 \text{EP} + b_8 \text{TOT} + b_9 \text{ER}(-1) + b_{10} \text{WGDP}_{\text{Indx}}(-1) + b_{11} \text{ExCr} \\
+ b_{12} \text{I}_{\text{Indx}}
\]

It is hypothesized that \( b_1, b_2, b_3, b_6, b_7, b_8, b_9, b_{10}, b_{11}, b_{12} > 0 \) while \( b_4 < 0 \).

(B) **Merchandise Imports (Imp)**

\[
\text{Imp} = \alpha_0 + b_1 \text{RPR} + b_2 \text{GNP} + b_3 \text{TOT} + b_4 \text{OP} + b_5 \text{ER} + b_6 \text{Imp}(-1) + b_7 D \\
+ b_8 \text{GNP}(-1) + b_9 \text{ER}(-1) + b_{10} \text{WGDP}_{\text{Indx}} + b_{11} \text{WGDP}_{\text{Indx}}(-1)
\]

It is hypothesized that \( b_2, b_6, b_7, b_8, b_{10}, b_{11} > 0 \) while \( b_1, b_3, b_4, b_5, b_9 < 0 \).

(2) **Invisibles Account**

(A) **Nominal Invisible Receipts(Inv Rec)**

\[
\text{Inv Rec} = \alpha_0 + b_1 \text{WGDP}_{\text{Indx}} + b_2 \text{ER} + b_3 \text{RCPI} + b_4 D + b_5 \text{Inv Rec}(-1) \\
+ b_6 \text{WGDP}_{\text{Indx}}(-1) + b_7 \text{ER}(-1) + b_8 \text{I}_{\text{Indx}}
\]

It is hypothesized that \( b_1, b_2, b_4, b_5, b_6, b_7, b_8 > 0 \) while \( b_3 < 0 \).
Components of Invisible Receipts

(i) Travel Receipts (Trav Rec)

\[ Trav \, Rec = \alpha_0 + b_1 WGDPI_{\text{idx}} + b_2 IGE_{\text{idx}} + b_3 Trav \, Rec(-1) + b_4 D + b_5 ER \]
\[ + b_6 GNP + b_7 WGDPI_{\text{idx}}(-1) + b_8 ER(-1) + b_9 GNP(-1) + b_{10} I_{\text{ndx}} \]

It is hypothesized that \( b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10} > 0. \)

(ii) Transportation Receipts (Trans Rec)

\[ Trans \, Rec = \alpha_0 + b_1 Exp + b_2 I_{\text{ndx}} + b_3 D + b_4 Trans \, Rec(-1) + b_5 ER + b_6 GNP \]
\[ + b_7 WGDPI_{\text{idx}} + b_8 WGDPI_{\text{idx}}(-1) + b_9 ER(-1) + b_{10} GNP(-1) \]

It is hypothesized that \( b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10} > 0. \)

(iii) Insurance Receipts (Ins Rec)

\[ Ins \, Rec = \alpha_0 + b_1 Exp + b_2 WGDPI_{\text{idx}} + b_3 ER + b_4 GNP + b_5 D + b_6 GNP(-1) \]
\[ + b_7 WGDPI_{\text{idx}}(-1) + b_8 ER(-1) + b_9 Ins \, Rec(-1) \]

It is hypothesized that \( b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9 > 0. \)

(iv) Investment Income Receipts (Inv Inc Rec)

\[ Inv \, Inc \, Rec \]
\[ = \alpha_0 + b_1 FCA + b_2 D + b_3 WGDPI_{\text{idx}} + b_4 ER + b_5 GNP \]
\[ + b_6 WGDPI_{\text{idx}}(-1) + b_7 ER(-1) + b_8 GNP(-1) + b_9 Inv \, Inc \, Rec(-1) \]

It is hypothesized that \( b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9 > 0. \)

(v) Private Transfer Receipts (PTR)

\[ PTR = \alpha_0 + b_1 OP + b_2 ER + b_3 WGDPI_{\text{idx}} + b_4 GNP + b_5 ER(-1) \]
\[ + b_6 WGDPI_{\text{idx}}(-1) + b_7 PTR(-1) + b_8 GNP(-1) + b_9 D \]

It is hypothesized that \( b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9 > 0 \) while \( b_1 < 0 \)

(vi) Other Invisible Receipts (Ot Inv Rec)

\[ Ot \, Inv \, Rec = \alpha_0 + b_1 Exp + b_2 WGDPI_{\text{idx}} + b_3 IGDPI_{\text{idx}} + b_4 GNP + b_5 ER + b_6 Ot \, Inv \, Rec(-1) + b_7 WGDPI_{\text{idx}}(-1) + b_8 ER(-1) + b_9 GNP(-1) + b_{10} D \]

It is hypothesized that \( b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_8, b_9, b_{10} > 0 \)
(B) Nominal Invisible Payments

Invisible Payments (Inv Pay)

\[ Inv\ Pay = a_0 + b_1 GNP + b_2 ED + b_3 ER + b_4 Imp + b_5 D + b_6 Imp(-1) + b_7 ER(-1) + b_8 GNP(-1) \]

It is hypothesized that \( b_1, b_2, b_4, b_5, b_6, b_7, b_8 > 0 \) while \( b_3 \) and \( b_7 < 0 \).

Components of Invisible Payments

(i) Travel Payments (Trav Pay)

\[ Trav\ Pay = a_0 + b_1 GNP + b_2 FER + b_3 D + b_4 WGDPI_{ndx} + b_5 ER + b_6 ER(-1) + b_7 GNP(-1) + b_8 Trav\ Pay(-1) + b_9 WGDPI_{ndx}(-1) \]

It is hypothesized that \( b_1, b_2, b_4, b_5, b_7, b_8, b_9 > 0 \) while \( b_5 \) and \( b_6 < 0 \).

(ii) Insurance Payments (Ins Pay)

\[ Ins\ Pay = a_0 + b_1 FER + b_2 GDP_{PS} + b_3 Imp + b_4 GNP + b_5 WGDPI_{ndx} + b_6 ER + b_7 D + b_8 Ins\ pay(-1) + b_9 ER(-1) + b_{10} WGDPI_{ndx}(-1) + b_{11} GNP(-1) \]

It is hypothesized that \( b_1, b_2, b_3, b_4, b_5, b_7, b_8, b_{10}, b_{11} > 0 \) while \( b_6 \) and \( b_9 < 0 \).

(iii) Transportation Payments (Trans Pay)

\[ Trans\ Pay = a_0 + b_1 Imp + b_2 ER + b_3 ER(-1) + b_4 D + b_5 WGDPI_{ndx} + b_7 GNP + b_8 GNP(-1) + b_9 Trans\ Pay(-1) \]

It is hypothesized that \( b_1, b_2, b_3, b_5, b_7, b_8, b_9 > 0 \) while \( b_2, b_3 < 0 \).

(iv) Investment Income Payments (Inv Inc Pay)

\[ Inv\ Inc\ Pay = a_0 + b_1 IPI + b_2 ED + b_3 ER + b_4 FER + b_5 GNP + b_6 WGDPI_{ndx} + b_7 D + b_8 ER(-1) + b_{10} WGDPI_{ndx}(-1) + b_{11} Inv\ In\ Pay(-1) \]

It is hypothesized that \( b_1, b_2, b_4, b_5, b_6, b_7, b_8, b_{10}, b_{11} > 0 \) while \( b_3 \) and \( b_8 < 0 \).

(v) Private Transfer Payments (PTPay)

\[ PTPay = a_0 + b_1 PTPay(-1) + b_2 WGDPI_{ndx} + b_3 WGDPI_{ndx}(-1) + b_4 D + b_5 ER + b_6 ER(-1) + b_7 GNP + b_8 GNP(-1) \]
It is hypothesized that $b_1, b_2, b_3, b_4, b_7, b_9 > 0$ while $b_5$ and $b_6 < 0$.

(vi) **Other Invisible Payments (Ot Inv Pay)**

$$\text{Ot Inv Pay} = a_0 + b_1 \text{Ot Inv Pay}(-1) + b_2 \text{WGDP}_{\text{indx}} + b_3 \text{WGDP}_{\text{indx}}(-1) + b_4 \text{D}$$
$$+ b_5 \text{ER} + b_6 \text{ER}(-1) + b_7 \text{GNP} + b_8 \text{GNP}(-1)$$

It is hypothesized that $b_1, b_2, b_3, b_4, b_7, b_9 > 0$ while $b_5$ and $b_6 < 0$.

**DETERMINANTS OF CAPITAL ACCOUNT**

Given the limitations of the availability of comparable disaggregated time series data since 1980-81, the capital account has been modeled by taking into account as much disaggregation as possible since 1991. The capital account is disaggregated into following items:

(A) Capital Account Receipts
(B) Capital Account Payments
(C) Net inflows under Non-Resident Indian Deposits [NRINET]

(A) **Components of Capital Account Receipts**

(i) **Foreign Direct Investment Inflows (FDIIN)**

$$\text{FDIIN} = a_0 + b_1 \text{WGDP}_{\text{indx}} + b_2 \text{GFD}_{\text{RATIO}} + b_3 \text{ED} + b_4 \text{Inf} + b_5 \text{FER}$$
$$+ b_6 \text{FDIIN}(-1) + b_7 \text{II}_{\text{indx}} + b_8 \text{ER} + b_9 \text{GDP} + b_{10} \text{GDP} + b_{11} \text{Exp}$$

It is hypothesized that $b_1, b_2, b_3, b_4, b_5, b_6, b_7, b_9, b_{10}, b_{11} > 0$ while $b_2, b_3, b_4, b_8 < 0$

(ii) **Foreign Institutional Investment Inflows (FIIN)**

$$\text{FIIN} = a_0 + b_1 \text{FIIN}(-1) + b_2 \text{WGDP}_{\text{indx}} + b_3 \text{ER} + b_4 \text{Inf}_{\text{Abroad}} + b_5 \text{RSP} + b_6 \text{Inf}$$
$$+ b_7 \text{II}_{\text{indx}} + b_8 \text{RSP}(-1)$$

It is hypothesized that $b_1, b_2, b_3, b_4, b_7 > 0$ while $b_3, b_5, b_6, b_8 < 0$

(iii) **External Assistance Inflows (EXTASS(I))**

$$\text{EXTASS(I)} = a_0 + b_1 \text{EXTASS(I)}(-1) + b_2 \text{GNP} + b_3 \text{DSR} + b_4 \text{Inf} + b_5 \text{Inf}_{\text{Abroad}}$$

It is hypothesized that $b_1, b_3, b_4 > 0$ while $b_2$ and $b_5 < 0$
(iv) External Commercial Borrowings Inflows [ECB(I)]

\[ ECB(I) = \alpha_0 + b_1 ECB(I)(-1) + b_2 DSR(-1) + b_3 Imp + b_4 ROI \]

It is hypothesized that \( b_1, b_2, b_3, b_4 > 0 \)

(B) Components of Capital Account Payments

(i) Foreign Direct Investment Outflows (FDIOUT)

\[ FDIOUT = \alpha_0 + b_1 ER + b_2 OP N + b_3 INF + b_4 FER + b_5 GNP + b_6 INF_{Abroad} + b_7 ED + b_8 FDIOUT(-1) \]

It is hypothesized that \( b_1, b_2, b_3, b_4, b_5, b_8 > 0 \) while \( b_6, b_7 < 0 \)

(ii) Foreign Institutional Investment Outflows (FIIOUT)

\[ FIIOUT = \alpha_0 + b_1 FIIOUT(-1) + b_2 GFD_{RATIO} + b_3 RBSE + b_4 INF + b_5 GNP + b_6 INF_{Abroad} + b_7 RBSE(-1) \]

It is hypothesized that \( b_1, b_4, b_5 > 0 \) while \( b_2, b_3, b_6, b_7 < 0 \)

(iii) External Assistance Outflows [EXTASS(O)]

\[ EXTASS(O) = \alpha_0 + b_1 EXTASS(O)(-1) + b_2 GNP + b_3 INF_{Abroad} + b_4 DSR \]

It is hypothesized that \( b_1, b_2, b_3, b_4 > 0 \)

(iv) External Commercial Borrowings Outflows [ECB(O)]

\[ ECB(O) = \alpha_0 + b_1 ECB(O)(-1) + b_2 Imp + b_3 ER \]

It is hypothesized that \( b_1, b_3 > 0 \) and \( b_2 < 0 \).

(C) Non - Resident Deposits (NRINET)

\[ NRINET = \alpha_0 + b_1 NRINET(-1) + b_2 IntDiff + b_3 OP + b_4 ER \]

It is hypothesized that \( b_1, b_2, b_3 > 0 \) while \( b_4 < 0 \).

In order to identify the chief determinants of balance of payments and its components, step-down multiple regression analysis was used. For this purpose, all the independent variables were jointly regressed upon the dependent variable in the
first step. The least non-significant variable has been dropped out from the equation in the next iteration and again multiple linear regression was worked out for the remaining variables. This iteration procedure continued till only the significant variables are left. With the help of adjusted coefficients of determination (i.e. $R^2$) associated with these lines of regression, the line with the best set of explanatory variables was chosen.

$$\tilde{R}^2 = 1 - [(n - 1)/(n - k)(1 - R^2)]$$

Significance of the individual partial regression coefficients was tested through t-test and that of overall regression equation was tested through Snedecor’s F-Test:

$$F = \frac{R^2/k}{(1-R^2)/(n-k-1)} \sim F_{(k,n-k-1)}d.o.f$$

Where $R^2$ stands for the coefficient of multiple determination, ‘$k$’ for the number of explanatory variables appearing in the equation and ‘$n$’ for the number of observations used for making computations.

However, to consider the effect of time on the regression equation, the set of significant variables and variable time are jointly regressed upon dependent variables through multiple linear regression equation, expressed as:

$$Y = \alpha_0 + \alpha_1 t + b_1X_1 + b_2X_2 + b_3X_3 + \cdots \cdots + b_nX_n$$

Where $Y$ is the dependent variable, $t$ is time variable, $X_1, X_2, X_3, \ldots \ldots \ldots \ldots X_n$ are all independent variables and $b_1, b_2, b_3, \ldots \ldots \ldots \ldots b_n$ are the regression coefficients.

**Cointegration and Vector Error Correction Mechanism Approach**

An econometric model can be set out to test the long run relationship and causality between the variables. Most of the macroeconomics time series are characterized by a random walk so that their first differences are stationary [Engle and Granger,(1987); Nelson and Plosser,(1982)]. If statistical tests, like cointegration, establish co-movements in these time series, the residuals from the regression can be used as error correction terms in the dynamic first-difference equation [Ahmed and Harnhirun,(1995)]. Therefore, for two integrated I(1) and cointegrated time series, there must exist causality in at least one direction in the I(0) variables[Engle and
Granger,(1987)] and hence a VAR model can be prepared with an error correction term for two cointegrated I(1) time series to capture the short run dynamics and to decrease the chance of observing ‘spurious regression’ in terms of the levels of data or their first differences.

Thus the first step is to test the order of integration, i.e. the stationarity, of the natural logarithm of the levels of the Current Account Balance (LnCA), Trade Balance (LnTB), Invisibles Account (LnINV) and Capital Account (LnCAP).

**Iteration Towards Time Series Stationarity**

Engle and Granger (1987), define a non-stationary series as a series which become stationary after being differentiated ‘d’ times. This notion is usually denoted by $X_t \sim I(d)$. Hence, all series are tested for the probable order of difference stationarity by using the augmented Dickey-Fuller (ADF) tests. ADF is a standard unit root test, it analyzes order of integration of the data series. These statistics are calculated without a constant, with a constant and a constant plus time trend respectively, and has a null hypothesis of non-stationarity against an alternative of stationarity.

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

\[
\Delta y_t = \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t \tag{3.1}
\]

\[
\Delta y_t = \alpha_0 + \gamma y_{t-1} + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t \tag{3.2}
\]

\[
\Delta y_t = \alpha_0 + \gamma y_{t-1} + a_2 t + \sum_{i=1}^{p} \beta_i \Delta y_{t-i} + u_t \tag{3.3}
\]

Where $u_t$ is a pure white noise error term and $\Delta y_t = (y_t - y_{t-1}), \Delta y_{t-1} = (y_{t-1} - y_{t-2})$

These tests determine whether the estimates of $\gamma$ are equal to zero. Fuller (1976) provided the cumulative distribution of the ADF statistics, if the calculated $t$-ratio of the coefficient $d$ is less than the critical value from the Fuller table, then $y_t$ is said to be stationary. Having established that all the series are integrated of order (d), that is I(d), the study then proceeds to determine the long run behavioral relationship among these variables. For examining the long run relationship among the variables, they must be co-integrated.
Thus, the second step is to test for cointegration applying Johansen’s Maximum Likelihood estimation approach[Johansen,(1988); Johansen and Juselius,(1990); and Johansen,(1991)].

**Cointegration and the VECM**

The desire to have models which combine both short-run and long-run properties, and which at the same time maintain stationarity in all of the variables, has led to a reconsideration of the problem of regression using variables that are measured in their levels. Non-stationary variables in a regression model give spurious results (Harris, Sollis, 2006). So, if $Y_t$ and $X_t$ are both I(1), then regression equation is:

$$Y_t = \beta_1 + \beta_2 X_t + u_t \quad (3.4)$$

This will not give satisfactory estimates of $\beta_1$ and $\beta_2$. One way of resolving this is to difference the data in order to ensure stationarity of the variables. Therefore, after differencing [$\Delta Y_t \sim I(0)$ and $\Delta X_t \sim I(0)$], the regression model will be:

$$\Delta Y_t = \tilde{a}_1 + \tilde{a}_2 \Delta X_t + \Delta u_t \quad (3.5)$$

In this case, the regression model may give us correct estimates of the $a_1$ and $a_2$ parameters and the spurious equation problem is resolved. However, equation (3.5) only reveals the short run relationship between the two variables. However, in the long run:

$$Y_t^* = \beta_1 + \beta_2 X_t \quad (3.6)$$

So $\Delta Y_t$ is bound to give us no information about the long run behaviour of our model. In order to resolve this problem, the concept of cointegration and Error Correction Mechanism(ECM) are very useful. As $Y_t$ and $X_t$ are both integrated of order one i.e. $Y_t \sim I(1)$ and $X_t \sim I(1)$. In the special case that there is a linear combination of $Y_t$ and $X_t$, that is $I(0)$, then $Y_t$ and $X_t$ are cointegrated. Thus, if this is the case, the regression of equation (3.5) is no longer spurious, and it also provides us with the linear combination:

$$\tilde{a}_t = Y_t - \hat{\beta}_1 - \hat{\beta}_2 X_t \quad (3.7)$$

That connects $Y_t$ and $X_t$ in the long run.
The Error-Correction Mechanism (ECM)

If, then $Y_t$ and $X_t$ are cointegrated by definition $u_t \sim I(0)$. We can express the relationship between $Y_t$ and $X_t$ with an ECM specification as:

$$
\Delta Y_t = a_0 + b_1 \Delta X_t - \Pi \hat{u}_{t-1} + Y_t
$$

(3.8)

This will now have the advantage of including both long run and short run information. In this method, $b_1$ is the impact multiplier (the short run effect) that measures the immediate impact that a change in $X_t$ will have on a change in $Y_t$. On the other hand, $\Pi$ is the feedback effect, or the adjustment effect, and shows how much of the equilibrium is being correlated, i.e. the extent to which any disequilibrium in the previous period affects any adjustment in $Y_t$.

$$
\hat{u}_{t-1} = Y_{t-1} - \hat{\beta}_1 - \hat{\beta}_2 X_{t-1}
$$

From this equation we also have $\beta_2$ being the long run response (it is estimated by equation 3.5).

Equation 3.8 emphasizes the basic approach of cointegration and error correction models. The spurious regression problem arises because of the use of non stationary data but in equation (3.8) everything is stationary, the change in $X$ and $Y$ is stationary because they are assumed to be $I(1)$ variables and the residual from the levels regression (3.7) is also stationary by the assumption of cointegration. So equation (3.8) fully conforms to our set of assumptions about the classical linear regression model and OLS should perform well (Granger, 1995).

Testing for Cointegration

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

1) Cointegration in single equations: The Engle-Granger Approach
2) Cointegration in multiple equations: The Johansen Approach

Here we will concentrate only on second approach as we have more than two variables in the model.

Cointegration in Multiple Equations and the Johansen Approach

If we have more than two variables in the model, then there is the possibility of having more than one cointegrating vector. In general for $n$ number of variables
we can have only upto n-1 cointegrating vectors. Therefore, when n=2 and if cointegration exists then the cointegrating vector is unique.

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 there are three variables, Y_t, X_t and W_t which can all be endogenous.

\[ Z_t = A_1Z_{t-1} + A_2Z_{t-2} + \ldots + A_kZ_{t-k} + u_t \]  \hspace{1cm} (3.9)

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

\[ \Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \Gamma_2 \Delta Z_{t-2} + \ldots + \Gamma_{k-1} \Delta Z_{t-k-1} + \Pi Z_{t-1} + u_t \]  \hspace{1cm} (3.10)

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

(i=1, 2,\ldots k-1)

Here we need to carefully examine the \( \Pi \) matrix of order 3\times3. 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.

The steps of the Johansen approach are the following

Step 1: The first step in the Johansen approach is to test for the order of integration of the variables under examination.

Step 2: The second step is the setting up of appropriate lag length of the model. Lag length which we have chosen in our analysis is 1 [followed by Astreou and Hall (2003)] in which it has been stated that if we have annual data then we can take lag length 1.

Step 3: This step includes 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:

\[
\Delta Z_t = \Gamma_1 \Delta Z_{t-1} + \ldots + \Gamma_{k-1} \Delta Z_{t-k-1} + \alpha \begin{pmatrix} \beta \mu_1 \\ \delta_1 \end{pmatrix} (Z_{t-1} \ 1 \ t) + \mu_2 + \delta_2 t + u_t \quad (3.11)
\]

And for this equation there are five possible cases. One 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. Although the first and the fifth are not that realistic, one can present all of them for reasons of complementarity.

**Model 1: No intercept or trend in CE or VAR** \(\delta_1 = \delta_2 = \mu_1 = \mu_2 = 0\).

In this case there are no deterministic components in the data or in the cointegrating relations. However, this is quite unlikely to occur in practice, especially as the intercept is generally needed in order to account for adjustments in the units of measurements of the variables in \((Z_{t-1} \ 1 \ t)\).

**Model 2: Intercept (no trend) in CE, no intercept or trend in VAR.**

**Model 3: Intercept in CE and VAR, no trends in CE and VAR, \((\delta_1 = \delta_2 = 0)\).** In this case there are no linear trends in the levels of the data, but we allow both specifications to drift around an intercept. In this case, it is assumed that the intercept in the CE is cancelled out by the intercept in the VAR, leaving just one intercept in the short run model.

**Model 4: Intercept in CE and VAR, linear trend in CE, no trend in VAR \((\delta_1 = 0)\).** In this model we include a trend in the CE as a trend stationary variable in order to take into account exogenous growth (i.e. technical progress). We also allow for intercepts in both specifications while there is no trend in the short run relationship.
Model 5: Intercept and quadratic trend in the CE and linear trend in VAR. This model allows for linear trends in the short run model and thus quadratic trends in the CE. Thus, in this final model everything is unrestricted.

So the problem is which of the five different models is appropriate in testing for cointegration. As stated earlier, first and fifth are not that realistic, therefore, the problem reduces to a choice of one of the three remaining models. We have chosen third model which is most appropriate model by using Monte Carlo simulation (Hollis and Follis, 2000).

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, and both involve estimation of the matrix $\Pi$. The procedures are based on propositions about eigenvalues.

(a) One method tests the null hypothesis, that rank ($\Pi$) = $r$ against the hypothesis that the rank is $r+1$. So, the null in this case is that there is cointegrating vectors and that we have upto $r$ cointegrating relationships, with the alternative suggesting that there is ($r$+1) vectors. The test characteristics are based on the characteristics roots (also called eigen values) obtained from the estimation procedure. The test consists of ordering the largest eigen values in descending order and considering whether they are significantly different from zero. To understand the test procedure, suppose we obtained $n$ characteristic roots denoted by $\lambda_1 > \lambda_2 > \lambda_3 > \ldots > \lambda_n$. If the variables under examination are not cointegrated, the rank of $\Pi$ is zero and all the characteristic roots will equal zero. Therefore (1- $\hat{\lambda}_i$) will equal to 1 and since ln(1)=0, each one of the expressions will be equal to zero for no cointegration. On the other hand, if the rank of $\Pi$ is equal to 1, then 0< $\lambda_i$<1 so that the first expression (1- $\hat{\lambda}$) <0, while all the rest will be equal to zero. To test how many of the numbers of the characteristic roots are significantly different from zero, this test uses the following statistic:

$$\lambda_{max}(r, r + 1) = -2ln(1 - \hat{\lambda}_{r+1})$$  
(3.12)

As stated earlier, the test statistic is based on the maximum eigen value and because of that is called the maximum eigen value statistic (denoted by $\lambda_{max}$).
The second method is based on a likelihood ratio test about the trace of the matrix (and because of that it is called the trace statistic). The trace statistic considers whether the trace is increased by adding more eigen values beyond the rth eigen value. The null hypothesis in this case is that the number of cointegrating vectors is less than or equal to r. From the previous analysis it should be clear that when all \( \hat{\lambda}_i = 0 \), then the trace statistic is equal to zero as well. On the other hand, the closer the characteristic roots are to unity the more negative is the \( \ln(1 - \hat{\lambda}_i) \) term and, therefore, the larger the trace statistic. This statistic is calculated by:

\[
\lambda_{\text{trace}}(r) = -T \sum_{i=r+1}^{n} \ln(1 - \hat{\lambda}_i)
\]  

(3.13)

The usual procedure is to work downwards and stop at the value of r which is associated with a test statistic that exceeds the displayed critical value. Critical values for both statistics are provided by Johansen and Juselius(1990).