Chapter 3: Concepts, Database and Methodology

3.1 Introduction

The present chapter defines the concepts used in the study and outlines the methodologies applied for the measurement of determinants of FDI in EMEs. Besides, it also presents the nature, period of the study, selection criteria for the models and sources of data used in the study.

Figure 3.1: Research Methodology

Figure 3.1 explains the flow of research methodology used in this study. This study reviews the existing theoretical as well as empirical literature with respect to determinants of FDI. After critical review of literature this study carefully identifies the possible determinants of FDI. Hypothesis for this study has been developed through the existing literature. After development of the hypothesis, this study has developed the theoretical and econometric model to empirically test the hypothesis. This study collects data from various sources for analyzing country-level and sector-level determinates of
FDI. This data is further analyzed by using different panel data techniques to understand the impact of each independent variable on dependent variable and finally draws the conclusions and policy implications from the results.

3.2 Theoretical Model of FDI

The choice of location for investment depends on the stream of returns in the long-run which is largely determined by location specific potential to convert the investment into returns. This means that, in order to derive the theoretical model on the determinants of FDI inflows, it is important to define an appropriate functional form of conversion of investment into returns (Chatterjee et.al 2013). Based on Griffith and Webster (2004), we define the expected return from investment (ERi) as,

\[ ER_i = \Phi_i [\alpha_i \ln(FDI_i) + \beta] \]  

Here, FDI\(_i\) is the amount of FDI inflows into country \(i\), \(\Phi_i\) is the return per unit of realized output from FDI in the country/sector \(i\), and \(\alpha_i\) represents the country/sector specific potential to convert FDI into expected returns, given \(\beta\) as the threshold output which is constant across the countries/sectors. Therefore, the present value of expected return from investment (PVR\(_i\)) will be,

\[ PVR_i = \frac{\Phi_i [\alpha_i \ln(FDI_i) + \beta]}{(x_i + r)} \]

Here, \(x_i\) stands for the risks of investing in country/sector \(i\) and \(r\) for the rate of discount constant for all the countries/sectors. Similarly, the present value of the recurring expenses of investment will be,

\[ RE_i = \frac{\gamma FDI}{(x_i + r)} \]

Here, \(\gamma\) stands for the proportional factor and it is assumed to be constant across the countries/sectors. If the investors decide to make investment to the maximum amount of FDI\(_0\) in EMEs, investment in a particular country/sector FDI\(_i\) be either less than or equal to FDI\(_0\),
\[ FDI_0 \geq FDI_i \quad \text{or} \quad FDI_0 - FDI_i \geq 0 \]

Therefore, objective of the investors is to decide \( FDI_i \) so that the net expected return,

\[
NER_i = \frac{\phi_i [\alpha_i \ln(\text{FDI}) + \beta_i]}{(x_i + \gamma)} - \frac{\gamma FDI_i}{(x_i + \gamma)} \quad \text{is maximum subject to} \quad FDI_0 - FDI_i \geq 0
\]

Hence, the problem of the potential investors can be written in Lagrange expression as follows:

\[
L = \frac{\phi_i [\alpha_i \ln(\text{FDI}) + \beta_i]}{(x_i + \gamma)} - \frac{\gamma FDI_i}{(x_i + \gamma)} + \lambda (FDI_0 - FDI_i) \quad \text{..................4)}
\]

\[
\text{Where } \lambda \text{ is Lagrange Multiplier}
\]

By applying Kuhn-Tucker conditions of constrained optimization,

\[
\frac{\partial L}{\partial FDI} = 0
\]

\[
FDI = \frac{\phi_i \alpha_i}{(x_i + \gamma)} \quad \text{..........................5)}
\]

Assuming that the rate of discount ‘\( r \)’ is uniform across the countries/sectors, (5) can be expressed as the following functional relationship,

\[
FDI_i = f(\alpha_i, \phi_i, x_i) \quad \text{..........................6)}
\]

\[
\text{Here, } \frac{\partial FDI_i}{\partial \alpha_i} > 0; \frac{\partial FDI_i}{\partial \phi_i} > 0; \frac{\partial FDI_i}{\partial x_i} < 0
\]

This means that the optimum investment in country/sector \( i \) varies directly with the country/sector-specific potential to convert FDI into output, return per unit of realized output, but inversely with the risks of investment in the particular country/sector. But, the potential of a country/sector to convert FDI into return is likely to depend on various factors discussed in the literature review e.g. economic factors like Market Size, Trade openness, Exchange Rate; Policy variables like business environment, trade policies and
political and Institutional factors like institutional quality, better governance which is also empirically tested by various literatures discussed in chapter 2.

Therefore,

\[ \alpha_i = \emptyset (EV_i, PV_i, IV_i) \ldots \ldots \ldots (7) \]

Here, \( EV \) is economic variables like market size, exchange rate, trade openness, macroeconomic stability, \( PV \) represent variable like business environment and trade policies, \( IV \) represent institutional variables like Government effectiveness, rule of law, corruption. Substituting 7 into 6

\[ FDI_i = f (\emptyset (EV_i, PV_i, IV_i), \phi_i, x_i) \ldots (8) \]

Where,

- \( FDI_i \) = FDI inflow in country/sector \( i \)
- \( EV = \) Economic Variables which includes Market Size, Trade openness, Macroeconomic stability, Inflation Rate.
- \( IV = \) Corruption, Rule of Law, Voice and Accountability; Government Effectiveness, Regulatory Quality, Political Stability No Violence
- \( PV = \) Policy Variables which includes business environment, trade policies.
- \( \phi_i = \) Profitability measured by growth of the country/sector
- \( X_i = \) Risk for country/sector \( i \) in year \( t \) is measured as the standard deviation of profitability during previous three years.

This study is only concerned about country/sector specific potential to convert FDI into expected returns, so \( \phi_i \) and \( X_i \) are considered as constant.

The functional relationship (8) is, therefore, reduced to

\[ FDI_{it} = f (EF_{it} + PV_{it} + IV_{it}) \]
Assuming that there exists linearity in the relationships, the above functional relationship can be rewritten as

\[ FDI_{it} = \alpha + \beta I(Economic\ Variables)_{it} + \phi(Policy\ Variables)_{it} + \mu(Institutional\ Variables)_{it} + u_{it} \]

### 3.3 Sources and Types of Data

The study is based on secondary data. The study makes use of a panel dataset comprising 24 EMEs defined by IMF namely, Bulgaria, Brazil, Chile, China, Colombia, Estonia, Hungary, Indonesia, India, Lithuania, Latvia, Mexico, Malaysia, Pakistan, Peru, Philippines, Poland, Romania, Russian Federation, South Africa, Thailand, Turkey, Ukraine and Venezuela over the period 2002 to 2012 for macro level analysis and for major three sectors over the period 2003-2012. Data on aggregate FDI inflows for selected period is taken from World Bank. Data on FDI flows classified into primary, secondary, and tertiary sector were from investment map collected by International Trade Centre. Real GDP, Inflation Rate, Exchange Rate, Trade to GDP ratio, Natural Resource Availability and Infrastructure quality data for the entire study period were collected from World Bank. The data on Governance indicator were procured from the Worldwide Governance Indicator Publications of the World Bank. The data on Labor Freedom, Trade barriers, Business Environment and Freedom from Corruption were collected from Index of Economic Freedom published by Heritage Foundation. Definition of all dataset is given in Annexure C.

### 3.4 Methodology

#### 3.4.1 Panel Data Analysis

The present study finds out the determinants of FDI in EMEs at macro level and sector level which covers ten time periods (2002-2012) and twenty four EMEs. Time-series analysis for such a short period is not appropriate. Study based on simple cross-sectional data at country level also becomes ineffective due to limited number of observations (Baltagi, 2005). Since FDI is a dynamic process, panel data is more appropriate for a systematic and efficient analysis of determinants (Dunning, 1993). Panel data model allows us to study dynamic as well as cross-sectional aspects of a problem. As it takes
average over subjects, the statistics becomes more reliable and also we require fewer
time-series observations to estimate dynamic patterns. A panel data set possesses several
other advantages over cross-sectional or time-series data set. Panel data model usually
gives a larger number of data points, increasing the degrees of freedom, reducing the
collinearity among the explanatory variables, and hence improving the efficiency of
econometric estimates. It allows analysis of number of important economic questions that
cannot be tackled by cross section or time-series data, while becoming more informative.
In general, the panel data model can be written as:

$$ Y_{it} = \alpha_{it} + \sum_{k=1}^{K} \beta_{kit} X_{kit} + u_{it} $$

Where,

- $i=1,2,...,N$ (refers to cross-sectional units)
- $t=1,2,...,T$ (refers to a given time period)
- $k=1,2,...,K$ (refers to number of explanatory variables)

Thus, $Y_{it}$ represents the values of the dependent variable for the individual $i$ at time $t$ and
$X_{kit}$ represent the values of the $k^{th}$ non-stochastic explanatory variables for the individual $i$
at time $t$. The stochastic term $u_{it}$ is assumed to have mean zero [ $E(u_{it}) = 0$ ] and constant
variance [$E(u_{it}^2) = \sigma^2$], are unknown parameters of response coefficients and for the
most general case they can be different for different individuals in different time periods.
$\beta_{kit}$ are unknown parameters or response coefficients. The appropriate estimation
procedure for the model depends upon whether the $a i$ is assumed to be random or fixed.
Fixed assumption for $ai$ lead to fixed-effects (FE) model and Seemingly Unrelated
Regression (SUR) model, while the random assumption for $ai$ leads to error component
model or random effects (RE) model and Swamy Random Coefficient model.

### 3.4.2 Specification of Panel Data Model

The model adopted for the present study is one where varying intercept terms are
assumed to capture differences in behavior over individuals and where the coefficients
are assumed to be constant. This model, in general, can be written as:

$$ Y_{it} = \alpha_{it} + \sum_{k=1}^{K} \beta_k X_{kit} + u_{it} $$
Thus, Y_{it} is the value of FDI inflow for the $ith$ country/sector at time $t$. $X_{kit}$ is the value of such determinants as market size, trade openness, macroeconomic stability etc. The appropriate estimation procedure for this model depends upon whether $\alpha_i$ are assumed to be random or fixed. Fixed assumption for $\alpha_i$ leads to FE model and SUR model, while the random assumption for $\alpha_i$ leads to random effect model.

3.4.2.1 Fixed Effect Model

The generalization of constant intercept and slope model for panel data is to introduce dummy variable to allow for the effects of those omitted variables that are specific to country cross-sectional units but stay constant over time and the effects that are specific to each time-period but are the same for all cross-sectional units. In the present study, no time-specific effects are assumed and the focus is only on individual-specific effects. Thus, the value of dependent variable for the $ith$ unit at time $t$, $Y_{it}$, depends on $K$ exogenous variables ($X_{1it}, X_{2it}, \ldots, X_{kit}$) that differ across individuals at a given point in time $t$ and also shows variation through time as well as on variables that are specific to the $ith$ units and these variables stay constant over time.

In Fixed Effect model, since $\alpha_i$ are assumed to be fixed (invariant over time), we can write the model as:

$$Y_i = e_i \alpha_i + \beta X_{it} + u_i$$

$\beta_i (1*K)$ is a constant vector and $\alpha_i (1*1)$ scalar is constant representing the effects of those variables peculiar to the $ith$ country cross-sectional units that stay constant over time. $e_i$ is the dummy variable for $ith$ country, and $u_i$ represents the effects of the omitted variables that are peculiar to both individuals and time. We assume $u_i$ to be uncorrelated with $X_{it}$ and are independently and identically distributed random variables with mean zero and variance $\sigma^2 u$

Writing in the vector form, we have

$$Y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_n \end{bmatrix} = \begin{bmatrix} e \\ 0 \\ \vdots \\ 0 \end{bmatrix} \alpha_1 + \begin{bmatrix} e \\ 0 \\ \vdots \\ 0 \end{bmatrix} \alpha_2 + \ldots + \begin{bmatrix} e \\ 0 \\ \vdots \\ 0 \end{bmatrix} \alpha_n + \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix} \beta + \begin{bmatrix} u_1 \\ u_2 \\ \vdots \\ u_n \end{bmatrix}$$
Where,

\[
Y = \begin{bmatrix}
Y_{i1} \\
Y_{i2} \\
\vdots \\
Y_{iT}
\end{bmatrix} \quad \text{and} \quad X_i = \begin{bmatrix}
x_{1i1} & x_{2i1} & \cdots & x_{ki1} \\
x_{1i2} & x_{2i2} & \cdots & x_{ki2} \\
\vdots & \vdots & \ddots & \vdots \\
x_{1iT} & x_{2iT} & \cdots & x_{kiT}
\end{bmatrix}
\]

\[e^i = (1, 1, \ldots, 1), \quad u^i = (u_{i1}, u_{i2}, \ldots, u_{iT})\]

\(y_{i1}\) is \((T \times 1)\) matrix; \(x_{1}\) is \((T \times k)\) matrix; \(e^i\) is \((1 \times T)\) matrix; \(u^i\) is \((1 \times T)\) matrix and \(I_t\) denotes the \(T \times T\) matrix.

Given the properties of \(uit\), the ordinary least squares (OLS) estimator is the Best Linear Unbiased and Efficient (BLUE) but we are faced with two problems. First, there is loss of degrees of freedom, since a large number of parameters are estimated. Second, with lot of dummies, there could be multicollinearity problems. Hence, we take deviation from the individual means to get

\[y_{it} - \bar{y}_i = \beta (x_{kit} - \bar{y}_k) + (u_{it} - \bar{u}_i)\]

Running OLS on this equation will be BLUE. This is called the least square dummy variable estimator. \(\alpha_i\) is estimated to see the individual-specific effect as follow:

\[y_{it} = \alpha_{it} + \sum_{k=1}^{K} \beta_k x_{kit} + u_{it}\]

Averaging over time gives

\[\bar{y}_i = a + \alpha_{it} + \sum_{k=1}^{K} \beta_k \bar{x}_{ki} + \bar{u}_it\]

Averaging over all observations gives:

\[\bar{y}_i = a + \sum_{k=1}^{K} \beta_k \bar{x}_i + \bar{u}_i \quad (\alpha_i \text{ sums to zero})\]

Where \(\bar{y} = \frac{\sum_i \sum t y_{it}}{NT}\)

So, \(y_{it} - \bar{y}_i = \beta (x_{kit} - \bar{y}_k) + (u_{it} - \bar{u}_i)\) and \(\alpha_i\) has been estimated as follows:
\[ \hat{\alpha}_i = (y_i - y) - \hat{\beta}_k (x_{ki} - x_k) \]

### 3.4.2.2 Random Effect Model

In fixed effect model, the effects of omitted variables (individual specific) are considered as fixed constants over time, whereas in Random Effect model, the individual-specific effects are treated as random variables. Thus, \( \alpha_i \) is assumed random. So in this case, \( \alpha_i \) is distributed independently and identically with mean zero and constant variance. Here, the model is

\[
Y_{it} = \mu + \beta' X_{iit} + v_{it}
\]

\[
v_{it} = \alpha_i + u_{it}
\]

Where \( \alpha_i \) is the individual-specific time invariant variable and \( u_{it} \) represents the effects of the omitted variables that vary with both individuals and time. The properties of \( \alpha_i \) are as follows:

\[
E(\alpha_i) = E(u_i) = 0
\]

\[
E(\alpha_i \alpha_j) = \begin{cases} \sigma^2 \alpha & \text{if } i = j \\ 0 & \text{if } i \neq j \end{cases}
\]

\[
E(u_i u_{jt}) = \begin{cases} \sigma^2 \alpha & \text{if } i = j, t = s \text{ or } o \text{ (otherwise) } \\ 0 & \text{otherwise} \end{cases}
\]

and \( E(\alpha_i) = E(u_i) = 0 \)

It exhibits serial correlation over time between disturbances of the same individual because \( v_{it} \) and \( v_{is} \) both contain \( \alpha_i \), so the residuals are correlated. To get efficient estimate, we have to use generalized least squares (GLS) estimator. The BLUE in the latter case is GLS estimator.

Rewriting in vector form, we have

\[
\begin{bmatrix} y_i \\ \vdots \end{bmatrix} = \begin{bmatrix} X_{i1} & \ldots & X_{iK} \end{bmatrix} \begin{bmatrix} \delta \\ \vdots \end{bmatrix} + \begin{bmatrix} v_i \\ \vdots \end{bmatrix}
\]

Where \( i = 1, 2, \ldots, N \)

\[
\begin{bmatrix} y_1 \\ \vdots \end{bmatrix} = (e, X_1), \delta = (\mu, \beta'), v_i = (v_{i1}, \ldots, v_{ik}), \text{ and } v_{it} - \alpha_i + u_{it}
\]

The variance-covariance matrix of \( v_i \) is
\[ E(v_i v_j') = \sigma^2 u_i + \sigma^2 \alpha^2 ee' = V \]

Its inverse is \[ V^{-1} = \frac{1}{\sigma^2} [I_1 + \sigma^2 \alpha^2 / (\sigma^2 + T \sigma^2 \alpha^2) ee'] \]

GLS estimator is \[ (\sum_{i=1}^{N} \hat{X}_i' V^{-1} \hat{X}_i) \delta GLS = \sum_{i=1}^{N} \hat{X}_i' V^{-1} y_i \]

3.4.2.3 Model Selection

This study conducts different three tests to identify the appropriate statistical model. We would like to test the joint significance of the intercepts capturing the unobserved individual effects of the Fixed Effect model. The pooled regression model is used as the baseline for the comparison. The joint significance test of these intercepts is performed with an F test

F-test : \[ H_0 : \alpha_1 = \alpha_2 = \alpha_3 = \ldots = \alpha_N \]
\[ H_a : \text{All possible alternatives} \]
\[ F = \frac{(RRSS - URSS) / (N - 1)}{URSS / NT - N - k} \]

Where, RRSS = restricted residual sum of squares is of the pooled OLS model, URSS = unrestricted residual sum of squares being the least square dummy variable.

F-test is carried out on the residual sums of squares for FE and pooled data model to show that there are variations in the unobserved country-specific and sector-specific effects. The basic idea behind the statistical test is as follows: If unobserved country-specific and sector-specific effects do not exist, pooled (OLS) estimators are BLUE and GLS estimators are inefficient and vice-versa (Judge et al. 1985; Hsiao, 2003; Baltagi, 2005). Then, the Lagrange multiplier (LM) test developed by Breusch and Pagan (1980) is designed to test RE model. The null hypothesis of the one-way random group effect model is that variances of groups are zero. If the null hypothesis is not rejected, the pooled regression model is appropriate. The LM is distributed as chi-squared with one degree of freedom. The two-way RE model has the null hypothesis that both cross sectional and time-series variance components are zero. The LM test combines two one way RE model for group and time. The LM is distributed as chi-squared with two degrees of freedom for two-way RE model.
Finally, Hausman specification test is used while comparing FE model with RE model. Hausman Test:

\[ H_0: \text{Random Effects} \]
\[ H_a: \text{Fixed Effects} \]

### 3.5 Panel Unit Root Test

As in the case of time series data, in estimating the panel data model, it may also be possible that the time-series properties of cross-section have an important influence on the specification of the econometric model. Thus, a stationarity test in the panel data is crucial. Due to the complex procedure in dealing with the panel data, the usual ADF test for unit root may result in inconsistent estimators. Therefore, we perform the panel unit root tests belonging to the first generation in panel data. In this study, we employ first generation tests provided by Levin, Lin and Chu (2002), and Im, Pesaran and Shin (2003) to confirm whether the underlying variables are stationary.

The general model of first generation panel unit root tests can be presented as follows. Let the model be:

\[
y_{it} = \delta_i y_{it-1} + z_{it}' K + \epsilon_{it} \quad i=1,2, \ldots, N \quad t=1,2, \ldots, T \quad (1)
\]

where \( \epsilon_{it} \) is the error term; \( z_{it}' \) is the deterministic component which could be zero, one, the fixed effects \( \mu_i \), or fixed effects as well as a time trend \( t \); and if \( \delta=1 \), it indicates the presence of unit root. Levine, Lin and Chu (2002) or (LLC, 2002) assume that meaning the persistence parameters are common across cross-section or all panel shares the same autoregressive parameters. Given this assumption, Eq. (1) can be rewritten as

\[
\Delta y_{it} = \rho y_{i,t-1} + z_{it}' K + \epsilon_{it} \quad i=1,2, \ldots, N \quad t=1,2, \ldots, T
\]

where, \( \rho = \delta - 1 \) Here the null hypothesis of \( \delta=1 \) is equivalent to \( \rho=0 \) and \( \Delta \) is the first difference operator. With additional lags of dependent variable, the augmented model of Levin, Lin and Chu (2002) can be represented as \( 1-\delta=\rho \)
\[ \Delta y_{it} = \rho y_{it-1} + z_i'K + \sum_{j=1}^{p} \beta_{ij}\Delta y_{it-j} + \epsilon_{it} \]

Where, \( p \) represents the maximum lags. The LLC test is based on testing the null hypothesis of \( H0: \rho = 0 \), against the alternative hypothesis of \( H1: \rho < 0 \). The null hypothesis of unit root is tested through a bias adjusted t-statistics for \( \rho \). This test assumes that the individual processes are cross-sectionally independent. The drawback of the LLC panel unit root test is that it considers the \( \rho \) to be constant across the panel. However, in practice such an assumption is tenuous. Im, Pesaran and Shin (2003) or (IPS, 2003) on the other hand relax such assumption by considering the null hypothesis of all cross-sectional entity contain unit roots against the alternative hypothesis of some (but not all) entities are stationary. The IPS (2003) panel unit root test can be specified as follows.

\[ \Delta y_{it} = \rho_i y_{it-1} + z_i'K + \sum_{j=1}^{p_i} \beta_{ij}\Delta y_{it-j} + \epsilon_{it} \]

Here, unlike the LLC test the null hypothesis is defined as \( H_0: \rho_i = 0 \) \( \forall i \) against the alternative hypothesis of \( H_0: \rho_i < 0 \) for \( i = 1,2,.....N1 \) and for \( i = N1 + 1, 2,.....N \) with \( o < N_1 \leq N \). The IPS test is based on the Augmented Dickey-Fuller (ADF) statistics taking simple average across groups. The null hypothesis of unit root in ith country is tested through \( t_{iNT} = \frac{1}{N} \sum_{i=1}^{N} iT(p_i, \beta_i) \) where, \( t_{iNT}(p_i; \beta_i) \) with \( \beta_{i,1} \ldots \beta_{i,p} \) denote the ADF t-statistic. They then standardized the statistics and show that the standardized t-bar statistics converge to a normal distribution as \( N \) and \( T \) tends to infinity. According to Im, Pesaran and Shin (2003) the t-bar test perform swell when the sample is small.

### 3.6 Specification of Panel Data Model (SUR)

To identify and compare the determinants of the FDI for selected EMEs, the current study used Seeming Unrelated Model (SUR) model. Panel estimator is a standard where elasticity coefficients are assumed constant and the intercept varies over individuals capturing the effects of those omitted variables that are specific to individual cross-sectional units but they stay constant over time. However, it is quite possible that
different attributes over the countries, in the present study, will be reflected in different
elasticity coefficients (Judge et al. 1985). This may be due to the changing economic
structure or due to different socioeconomic and demographic factors that allow the
response parameters to vary over time and/or they may be different for different cross
sectional units (Hsiao, 2003). In SUR model, the response parameters are allowed to vary
from one unit to another, while remaining invariant over time and the errors are allowed
to be contemporaneously correlated and heteroscedastic between individuals. When
elasticity coefficients are treated as invariant over time, but varying from one unit to
another, the model can be expressed as:

\[ Y_i = \beta_i^T X_{ki} + u_i \]

Or,

\[ Y_{it} = \sum_{k=1}^{K} (\beta_{k} + \alpha_{kt}) x_{kit} + u_{it} \]

Where \( \beta = (\beta_1, \ldots, \beta_K) \) is the common mean coefficient vector and \( \alpha_i = (\alpha_{i1}, \ldots, \alpha_{ik}) \) is
the individual deviation from the common mean \( \bar{\beta} \). Here individual observations are
heterogeneous, hence \( \alpha_i \) are treated as fixed constants. When \( \beta_i \) are treated as fixed and
different constants, we can stack the N*T observations in the form of Zellner’s (1962)
SUR model. SUR model in the vector form can be written as:

\[ Y_i = X_i \beta_i + u_i \]

Where, \( y_i \) and \( u_i \) are of dimension (T*1), \( X_i \) is (T*K_i) and \( \beta_i \) is (K_i*1)

Here though the errors are allowed to be contemporaneously correlated and heteroscedastic between individuals (cross-section), we still assume serial independence as well as homoscedasticity within individuals. If we are interested in only one equation, say the \( i^{th} \), then least square estimator is the minimum variance unbiased estimator. SUR model is better than lest square estimator because it allows for the correlation between \( ui \) and the other disturbance vectors and it uses information on explanatory variables that are included in the system but excluded from the \( i^{th} \) equation (Judge et al. 1985).
covariances between different cross-sectional units are not zero \((E u_i u_j = 0)\) the GLS estimator \((\beta_i, \beta_j \ldots \beta_n)\) is more efficient than the single equation estimator of \(\beta\) for each cross-sectional unit. If \(X_i\) are identical for all \(i\)'s or \(E u_i u_{ij} = \sigma^2_i\) and \((E u_i u_j = 0)\) for \(i \neq j\), the GLS estimator for \(i\) \((\beta_i, \beta_j \ldots \beta_n)\) is the same as applying least squares separately to time-series observations of each cross-sectional unit.

With a view to examine whether the response parameters vary significantly from one state to another and invariant over time, the test procedure has used three main steps:

1. Test whether or not slopes and intercepts simultaneously are homogenous in different times.

That is: \(y_{it} = \alpha + \sum_{k=1}^{K} \beta_k x_{ikt} + u_{it}\)

2. Test whether regression slopes collectively are the same.

That is: \(y_{it} = \alpha_i + \sum_{k=1}^{K} \beta_k x_{ikt} + u_{it}\)

3. Test whether or not regression intercepts are the same.

That is: \(y_{it} = \alpha + \sum_{k=1}^{K} \beta_{ik} x_{ikt} + u_{it}\)

The null hypotheses postulated for the present model are:

\(H_1\): Both slope and intercept coefficients are the same for all countries.

\[\alpha_1 = \alpha_2 = \ldots \ldots \alpha_N\]

That is: \(H_1: \beta_1 = \beta_2 = \ldots \ldots \beta_N\)

\(H_2\): Regression elasticity coefficients are identical, and intercepts are not.

That is: \(H_2: \beta_1 = \beta_2 = \ldots \ldots \beta_N\)

\(H_3\): Regression intercepts are the same, and elasticity coefficients are not.

That is: \(H_3: \alpha_1 = \alpha_2 = \ldots \ldots \alpha_N\)

Under the assumption that \(u_{it}\) are independently and normally distributed over \(I\) and \(t\) with zero mean and variance \(\sigma^2 u\), F-test is conducted to test the null hypotheses \(H_1\), \(H_2\) and \(H_3\). Under \(H_1\), the F-statistic carried out would be:

\[F_1 = \frac{(S_3 - S) / [(N - 1)(k + 1)]}{S / [NT - N(k + 1)]}\]
where $S3$ is the residual sum of squares with common intercept and slope; $S$ is the residual sum of squares of within group with heterogeneous intercept and slope. If $F1$ is not significant, we pool the data and estimate a single equation. If the $F$ ratio is not significant, a further attempt is made to find out if the non-homogeneity is due to heterogeneous slopes or intercept. Under the null hypothesis of heterogeneous intercept and homogeneous slope (H2), the $F$-statistic would be:

$$F_2 = \frac{(S_2 - S)/[(N - 1)k]}{S/[NT - N(k + 1)]}$$

Where $S_2$ is the residual sum of squares with constant slope and heterogeneous intercept; is the residual sum of squares of within group with heterogeneous intercept and slope. $NT - N(k+1)$ degrees of freedom is significant, then the null hypothesis of heterogeneous intercept but homogenous slope is rejected. However, if $F2$ is not significant, we can then determine the extent to which non-homogeneity can arise in the intercepts (Hsiao, 2003). If $H2$ is accepted, we can apply a conditional test for homogeneous intercepts, as $H3$:

$$\alpha_1 = \alpha_2 = \ldots \ldots \ldots \alpha_N \text{ given } \beta_1 = \beta_2 = \ldots \ldots \ldots \beta_N .$$

3.7 GMM Estimation

Measuring the relationship between FDI flows and many of the institutional variables used in the study, especially Infrastructure, Governance, Business Environment, Trade Freedom and Control on Corruption raise some endogeneity concerns. To address this, present study estimated the sector-specific result using Generalized Method of Moments (GMM) dynamic estimator based on the Arellano-Bond methodology. Arellano and Bond (1991) have shown that GMM estimators allow controlling for unobserved individual effects which is present in the static model, endogeneity and simultaneity of explanatory variables and its lagged values and the use of the lagged dependent variables. The Arellano-Bond methodology specifies a dynamic model which allows for time-invariant sector-specific effects. This seems plausible in the case of FDI, where variables outside the analysis, such as rule of law, regulatory quality, trade freedom display little, if any, variation over the period of the analysis. Sector specific effects are taken care of by first differencing the variables. The use of time dummies for each year in the sample takes care of time effects. In principle, the simultaneity problem in the estimation of models
can be tackled by the use of instrumental variables. Consistent GMM estimation requires 
that the instruments used be uncorrelated with unobservable effects to the function since 
these effects may be included in the error term.

The equation estimated below is:

\[ y_{i,t} = \alpha + \lambda y_{i,t-1} + X^i_{i,t} \beta + \mu_i + \nu_{i,t} \]  

where \( y_{i,t} \) denotes FDI inflow in sector \( i \) at time period \( t \), \( X \) is the vector of 
institutional/qualitative variables, and \( \mu \) represents the time-invariant sector-specific 
effects. Relaxing the assumption that all explanatory variables are strictly exogenous (i.e. 
if it is uncorrelated with the error term at all leads and lags) and assuming weak 
exogeneity i.e. variables are assumed to be uncorrelated with future realizations of the 
error term (i.e. current explanatory variables may be affected by past and current 
observations, but not by future ones) of the explanatory variables, the joint endogeneity 
of the explanatory variables requires an instrumental variable procedure to obtain 
consistent estimates of the coefficients of interest.

In case unobserved effects are not present, we can apply GMM estimation in levels. 
Under the assumption that the error term \( \varepsilon_{i,t} \) is serially uncorrelated or at least follows a 
moving average process of finite order and also assume that future innovations of the 
dependent variable do not affect current values of the explanatory variables, the 
observations viz. \( (Y_{it-2}, Y_{it-3}, \ldots, Y_{in}) \) and \( (X_{it-2}, X_{it-3}, \ldots, X_{in}) \) can be used as valid 
instruments in the GMM estimation. This is the GMM level estimator. However, in the 
presence of unobserved individual effects, the GMM level estimator produces 
inconsistent estimates. Taking first difference of equation (1) eliminates the time-
invariant sector-specific effects, generating the following equation:

\[ y_{i,t} - y_{i,t-1} = \alpha + \lambda(y_{i,t-1} - y_{i,t-2}) + (X^i_{i,t} - X^i_{i,t-1}) \beta + (\nu_{i,t} - \nu_{i,t-1}) \]

To account for possible endogeneity between the explanatory variables \( X_{i,t} \), and the 
dependent variable \( y_{i,t} \), the equation is estimated using as instruments the lagged values of 
the left- and right-hand side variables in levels. These instruments are valid if the error 
term \( \nu \) is not serially correlated.
The instruments for GMM model have been selected in the way followed by Blundell, et al (1992). In this procedure, for the variables, which are found to be pre-determined, instruments dated t-1, and for the variables, which are exogenously determined, instruments dated t-3 should be chosen. In order to test for the possibility that \( x_{it} \) is predetermined with respect to \( u_{it} \), we start using the instrument-dated t-2 for each variable included in the instrument set. Ideally, the instrument set would include all the instruments dated t-2 and earlier. Then \( x_{it-1} \) is added to the existing instruments to investigate the potential biases, which arise from the correlation between \( x_{it-1} \) and the first differenced error term \( \Delta u_{it} \). In the presence of the measurement error the estimate of the coefficient of \( x \), is expected to fall, which suggests a downward bias due to the simultaneous determination of \( \Delta x_{it-1} \) and \( \Delta u_{it-1} \). To access the validity of these assumptions on which these estimators are biased, we consider Sargan (1958) test of over identifying restrictions, which tests the validity of the instruments. It is based on heteroskedasticity consistent two-step GMM estimator that tests for the validity of extra instruments in the equation.

There are some statistical shortcomings to a straightforward instrumental variables estimation of the above equation, namely that in a small sample with some persistent explanatory variables, lagged levels make weak instruments for the regression when run in differences. To address this weakness, Blundell and Bond (1988) developed the system GMM dynamic model, which combines the regression in first differences above with an estimation run in levels, using both lagged levels and lagged differences as instruments.

Summary of the Econometric models applied in this study is given below:
<table>
<thead>
<tr>
<th>Objective</th>
<th>Model Applied</th>
<th>Functional Form</th>
<th>Justification</th>
</tr>
</thead>
<tbody>
<tr>
<td>1) To find out country specific determinants of FDI in EMEs</td>
<td>Panel Data Model</td>
<td>Panel Least square model</td>
<td>It increases the efficiency of econometric estimates and captures the dynamic of both cross-sectional and time-series data.</td>
</tr>
<tr>
<td></td>
<td>(a. common intercept &amp; slope</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>b. constant slope &amp; heterogeneous intercept</td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Fixed Effect Model</td>
<td>$Y_{it} = \alpha_{it} + \sum_{k=1}^{k} \beta_{kit} X_{kit} + u_{it}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td>Random Effect Model</td>
<td>$Y_{it} = (\alpha + \mu_{i}) + X_{it} \beta + v_{it}$</td>
<td></td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification Tests:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Incremental F Test : OLS Model vs. Fixed Effect Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Breush-Pagan LM Test : OLS Model vs. Random Effect Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Hausman Test : Fixed Effect Model vs. Random Effect Model</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>2) To compare the determinants of FDI in Selected EMEs</td>
<td>Seemingly Unrelated Regression (SUR) model</td>
<td>SUR Model</td>
<td>SUR model with heterogeneous intercept and slope is applied to find the response of determinants and/or associated intercepts varies according to countries.</td>
</tr>
<tr>
<td></td>
<td>(heterogeneous slope &amp; intercept)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification Tests:</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>F Test : To study the heterogeneous intercept and slope within group</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>To account for possible endogeneity between the explanatory variables $X_{it}$ and the dependent variable $Y_{it}$</td>
<td>Generalized Method of Moments (GMM) dynamic estimator based on Arellano-Bond methodology</td>
<td>$y_{it} - y_{i,t-1} = \alpha + \lambda_1(y_{i,t-1} - y_{i,t-2}) + (X_{i,t} - X_{i,t-1})\beta + (v_{i,t} - v_{i,t-1})$</td>
<td>GMM allows for the time-invariant sector-specific effects. This is plausible in the case of FDI, where variables outside the analysis, such as institutional quality and governance display little, if any, variation over the period of the analysis.</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Specification Test</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Sargan P value : To study over identifying restrictions, which tests the validity of the instruments</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>