CHAPTER THREE: **RESEARCH METHODOLOGY**

**3.1 Introduction**

Methodology in most research works refers to the general strategy followed by the researcher in gathering and analysing the data necessary for the work. In this regard this chapter presents the research design, conceptual framework, variables used in the model and source of data and model specification, data estimation procedure.

**3.2 Research Design**

Variables are selected upon the criteria and guidelines of theories, previous empirical evidences, and availability of the data. The author used the secondary data series of balanced data sets and collected the data mostly from one source in order to have data consistency. Even though there are still many factors that affect FDI and trade openness, the author excluded some of them because of data limitations and methodology constraints. Making decisions of choosing appropriate variables and methodology is one of the most challenging parts of this paper. In this study STATA 18 software is used for analyzing and estimating the results. In this investigation, annual panel data sources from 1980 to 2022 of BRI. In table 3.1, the author summarized scope of the study, time frame, sources of data, and methodology for this research.

**Table 3. 1 Summary of Study Scope, Time Frame, Sources and Method**

|  |  |
| --- | --- |
| Scope | BRI (150) Countries, T= 150 |
| Time Frame | 1980- 2022(42) years, N= 6300 observations |
| Source of data | World Bank, IMF, World Governance Indicators, and Statistical Years Books (various issue) ,World Economic Outlook(various issue) , WTO and UNCTAD |
| Method | Panel ARDL approach to Cointegration (PMG, MG DFE, ECT, Hausman Test and Unit Root Test.) |

Source: Author Design

In this research, the author used the variables that are available and fitted to the model and methodology. In fact, there have many variables that the relationship between trade openness, the effect of FDI on CO2 emission in BRI countries and many ways to conduct the research. Moreover, many studies have been conducted related to this field with different countries groups. In this study, the author applied Panel ARDL approach to cointegration to investigate long run and short run relationships. Then, the author divided the analysis and estimations into two groups by using PMG, MG and DFE estimators: (1) panel level analysis that means all BRI countries as a group for both long run and short run estimations and (2) individual level analysis that means each individual country of BRI Countries for short run estimations with long run panel level estimation. Firstly, the author collected the secondary data for the study period of 1980-2022 which has 6300 observations. Secondly, the author tested the collected data with panel unit root tests which are Levin-Lin-Chu (LLC 2002), Im-Peasaran (2003) and Maddala (1999). Based on the tests results of panel unit root tests, the author tried to test with different panel data analysis. Among them, Panel ARDL criteria are matched with the results. For the panel level, the author conducted to test the eight elected variables with eight different models. As the nature of the Panel ARDL approach, many variables cannot be tested in one model as this creates multi-co linearity between independent variables which have high correlation between them. Therefore, the author tests them with eight different models for panel level, selects one model for individual level to match with the objectives of the study and then uses Hausman test to choose for the best estimator between PMG, MG and DFE.The research process is as shown in Fig 3.1.

**3.3 Conceptual Framework**

In this section, author present about the theoretical framework that how they have the relationship between trade openness, FDI, CO2 emission and macroeconomic factor of the Developing Asian countries by applying Pooled Mean Group and Mean Group estimators to estimate the relationship and coefficients of variables.

Environment Quality (CO2)

Trade Openness (TO) Exchange Rate (EXR)

FDI Inflow (FDI) Economic Growth (GDP)

Market Size (POP)

Economic Stability (INF)

Urbanization (UR)

Infrastructure (EP)

Panel Unit Root Test

LLC Test

IPS Test

Fisher Type

Fisher Type

IP

Panel ARDL

Approach to  
Cointegration

Choose the best estimator by applying Hausmen Test

Interpretation

**Figure 3.1: The Research Method and Framework for the Study**

**Source: Author’s presentation**

For this reason, the author investigates the relationship of CO2 emission, FDI inflows and macroeconomic and political factors for both long run and short run by using Panel ARDL approach. The conceptual framework for this research is shown as in Fig 3.2.

**3.4 Variables Used in the Model and Sources of Data**

Economic growth is presented by gross domestic product per capita growth in US$. It is taken from the WDI. Countries with better living standards and income levels are fancied by the investors because their citizens have purchasing ability to buy quality products and afford valued

**3.4.1 Dependent Variable: Economic Growth**

GDP growth rate means the increasing capacity of the economy to satisfy the needs and wants of the society. Economic growth can be achieved by increasing the productivity of the economy. GDP growth means that economy is growing and developing, technically it means the increase or decrease in the GDP compared with previous years. It is also most important indicator of the economy

**3.4.2 Controlling/Independent Variables**

In this study seven independent variables are taken, in which Foreign Direct Investment, market size, trade liberalization; economic stability, Exchange Rate, infrastructure and Institutional Quality are controlling variables while trade liberalization is taken as the main explanatory variable. All the independent variables utilized in the study are elucidated below, in order to show their relative importance.

**3.4.2.1 Foreign Direct Investment**

Foreign direct investment inflows is one of the stimulator for the economic growth, in past studies, FDI was applied that an exogenous variable for the economic growth.

**3.4.2.2 Market Size**

Market size is one of the vital factors in attracting FDI. Larger markets provide opportunities for

the economies of scale. In FDI literature market size is mostly considered as one of the important

determinant of FDI inflows. Population (total) is taken as a proxy for market size from the World

development indicators for each Developing Asian-20 countries.

**3.4.2.3 Economic Stability**

Economic stability is considered a proxy for the prevailing inflation rate. The level of economic stability is used as an indirect indicator to gauge the expected movements or behaviour of inflation within the economy. This approach relies on the premise that a stable economic environment, characterized by consistent growth, low unemployment, and steady consumer and business confidence, often corresponds to a controlled and predictable inflation rate.

**3.4.2.4 Trade Openness**

Trade Openness data is used as a proxy for trade as a percentage of GDP ratios provided by the World Bank. The metrics related to trade openness are employed as an indirect measure to estimate the expected levels of trade as a share of GDP within the analysed regions or economies. This approach relies on the premise that higher degrees of trade liberalization, as captured through various indicators such as tariff reductions, removal of trade barriers, and open market policies, often correspond to elevated trade-to-GDP ratios, reflecting the magnitude of international trade relative to the economic output.

**3.4.2.5 Infrastructure**

Infrastructure data is used as a proxy for total electricity production. The level and quality of infrastructure serve as an indirect measure to gauge the expected capacity and output of electricity production within the region or sector under consideration. This approach relies on the premise that a well-developed and robust infrastructure, encompassing factors such as power generation facilities, grid connectivity, and distribution networks, often correlates with higher levels of electricity production and supply.

**3.4.2.6 Institutional Quality**

In the context of assessing the control of corruption, institutional quality is utilized as proxy data. The overall quality of institutions, including governance structures, legal frameworks, and regulatory effectiveness, is employed as an indirect measure to gauge the anticipated level of control over corrupt practices within a given socio-economic or political environment. This approach relies on the principle that robust and transparent institutions often correlate with a higher capacity to combat corruption, enforce the rule of law, and uphold ethical standards across public and private sectors.

**3.4.2.7 Exchange Rate**

Exchange rate is also one of the location pull factors of FDI inflows. Weaker currency mostly seems attractive to the foreign investors as it provides better purchasing power. Froot and Stein (1991) found negative relation between exchange rate and FDI. Here the local currency’s exchange rate per US$ is taken as a proxy for exchange rate.

Table 3.2. Expected Signs of the Variables Effecting Trade Openness and FDI on CO2 emission in BRI Countries

|  |  |
| --- | --- |
| Economic Growth (GDPP) | Positive |
| Trade Liberalization (TO) | Positive |
| Foreign Direct Investment | Positive |
| Infrastructure (EP) | Positive |
| Market Size (POP) | Positive |
| Urbanization (UR) | Positive |
| Environmental Quality (CO2) | Negative |
| Economic Stability (INF) | Negative |
| Exchange Rate | Negative |

Source: Author`s Design

**3.5 Model Specification**

To achieve our objectives of the study, we specified a model which is a process of constructing logical thinking and abstraction of economic reality. The specification of our model is based on the variables adopted for trade openness, Foreign Direct Investment and CO2 emission in the study. The following model is specified in this study to analyse the impact of trade openness along with the market size, Economic stability, Institutional Quality economic growth and Infrastructure, Exchange rate on inward FDI in BRI countries. Based on the literature reviews of the previous authors, the functional form can be derived as follows:

(CO2)it= f(TOit, FDIit, GDPit, INFit, URit, EPit, POPit, ) (1)

(LnCO2)it= + (LnTOit) + FDIit)+ (LnGDPit)+ INFit)+ URit) (2)

+ (LnEPit) + POPit) + EXRit )+

The above equation (2) is for panel level where i represent cross-section data (countries) and t represents time-series data. These variables are replaced by suitable proxies and equation 1 is log linearized, which gives us equation II. Taking log of the variables helps in removing the expected heteroskedasticity (Resmini, 2000). Where:

lnCO2 = log of CO2 emission is taken as a proxy for environmental quality. lnFDI = log of net FDI inflows in US$ is taken as a proxy for FDI. α = constant. β₁β₂β₃β₄β₅β₆ β7 and β8 = parameters used to link the dependent variable with independent variables. lnPOP = market size is represented by log of total population. LnINF= natural log of inflation rate employed as a proxy for economic stability. lnUR = Urbanization is represented by log of total urban population. lnXrate = log of exchange rate per US$ is used for exchange rate. lnEP = log of total electricity production is employed as a proxy for infrastructure. lnGDP = log of gross domestic product per capita growth is taken as a proxy for economic growth. lnTO = log of trade (% of GDP) is taken as a proxy for trade opennessthe error term..

**3.6 Data Estimation Procedure**

The advantage of using panel data is that it can include large no. of observations both individuals and time caused. Panel unit root tests were derived from time series unit root tests. The estimates are more consistent and efficient for panel unit root tests. In addition, it is suitable for both macro and micro level data with time-series and crosssection dimensions. In this research, the author tested the selected data with panel unit root tests in order to identify the appropriate methodology to apply for the estimation process. Also, selection of methodology can be varied upon the size of data which means number of countries and time lengths. This work used the application of STATA-View version 18.0 for its estimation procedure.

This particular software will adopt the following procedures:

**3.6.1 Panel Unit Root Tests**

Panel Unit Root Tests were derived from time series unit root testing. Time series unit root tests lacks power in testing the difference of the unit root from stationary alternatives. By using panel data unit root tests, power of tests can be increased substantially. There are four most widely used panel unit root tests which were developed by Levin, Lin and Chu (2002), Im, Pesearan and Shin (1997, 2003), Fisher type of ADF and PP tests (Maddala and Wu (1999).

**3.6.1.1 Levin, Lin and Chu (2002) Test**

The nature of panel data has both cross-section and time-series dimensions. Most of the panel unit root tests are similar but not identical. Levin et al (2002) considered a stochastic term () for i=1,...,n and t=1,…,t, this test is one of the suitable test to apply to test the panel data with the assumption of 𝜌𝑖 is same for all series under the alternative hypothesis. Normally all panels share a common autoregressive parameter and LLC augment the test with additional lags of the dependent variable. The following equation is to LLC test’s regression model:

= + +

In the above equation, ∆𝑌𝑖𝑡 is the difference term of 𝑌𝑖𝑡 and 𝑌𝑖𝑡−𝑗 represents panel data and ∗ is exogenous variable with individual time trend or country fixed effects. The assumption of LLC test is that 𝜀𝑖𝑡, the error term is distributed independently across panel and follows a stationary invertible autoregressive moving- average process for each panel. The null and alternative hypotheses are as below;

: = 0 for all i which means panel data has unit root

: or all i which means panel data has no unit root.

LLC test requires 𝛽𝑖 to be homogeneous across i for this hypothesis and this homogeneity requirement becomes a disadvantage of the LLC test. This implies that if the autoregressive parameters are same across panel, HA is restrictive, and t-statistics relied on pooled estimation can be described as:

=

Where 𝛼̂ has standard normal distribution, is standard t-statistics for 𝛼 = 0;

se ( ) = standard error of

= error term

= average standard deviation ratio

= adjustment term of the mean

𝜎𝑚𝑇∗ = adjustment term of the mean

**3.6.1.2 Im, Pesaran and Shin (2003) Test**

Im et al. (2003) suggests a t-bar statistic which is more flexible to analyse the unit root hypothesis on panel data which is relied on the average of individual ADF t-statistics. IPS test is more flexible, accurate and powerful than LLC test because it is based on the assumption of cross-sectional independence and allows hetrogeneity of coefficients across groups. For a sample having n groups and t time periods where i = 1,….,n and t = 1,…..,t , the regression model of the conventional ADF test for panel unit root is as follow.

∆𝑌𝑖𝑡 = 𝛼𝑖 + 𝛽𝑖𝑌𝑖𝑡−1 +

The null hypothesis and alternative hypothesis are as below;

𝐻𝑜: 𝛽𝑖 = 0 for all i which means panel data has unit root.

𝐻𝐴: 𝛽𝑖 < 0 for at least one cross-section which means some of the individual series allow having unit root. Two alternatives are specified and tested as unit root with an intercept and as unit root with trend and intercept. The test statistic can be written as follow:

𝑍𝑡𝑏𝑎𝑟 =

Where 𝑡𝑏𝑎𝑟𝑁𝑇 = the average ADF t-statistic:

E(𝑡𝑇𝑖) and 𝑉𝑎𝑟(𝑡𝑇𝑖) are mean and variance and calculated based on Monte-Carlo simulated moments. They depend on time dimension, nature, lag order and structure of ADF test. IPS test is known as one-sided lower tail test approached to standard normal distribution. Only balanced panel data is applicable according to the theory but in reality, when unbalanced data is applied, more simulations are required to get critical values.

**3.6.1.3 Fisher Type test (Maddala and Wu 1999) Augmented Dickey- Fuller (ADF) and Phillips-Perron (PP) Tests**

Test statistic discussed by Maddala and Wu (1999) is based on Fisher (1932) and combining p-values of t statistics for each unit root of each cross section. Fisher tests do not need to use the same unit root test in each cross section. This test permits different first-order autoregressive coefficients and tests stationary of null hypothesis and is similar to IPS. The Fisher test statistics is written as below:

P(𝜆) = −2 ln(𝑝𝑖)

Where P(𝜆) = Fisher’s panel unit root test, N= all N cross section:

𝑝𝑖 is P-value of ADF test for cross-section i and the test follows chi- square distribution with 2N degree of freedom. Fisher test is more flexible, accurate and powerful than LLC test and also having an advantage over IPS. This test allows panel unit root test with no intercept or trend. Maddala and Wu (1999) stated that “Fisher test is simple, straight forward and better than LLC and IPS.

Inverse normal test (Z) equation and logit test is

𝑍 = (

∅−1 is the inverse of standard normal cumulative distribution function 0 ≤≤ 1 and ∅−1(𝑝𝑖)~𝑁(0,1) , therefore 𝑍~𝑁(0,1)

The null and alternative hypotheses are as below;

𝐻𝑜: 𝑃𝑖 = 1, which means that all the time series have unit root.

𝐻𝐴: 𝑃𝑖 < 1, which means that some cross-section is non-stationary and has no unit root.

There are some advantages of applying Fisher-type test: they are (1) data are not necessary to be balanced, (2) they have size adjusted power and (3) different lag lengths are allowed for individual ADF regression. After tests are conducted by the above four unit root tests, we made decision of methodology selection based on their results and choose the method that is most suitable and applicable to the objective of this study. Also, the quality of data is very much important for the quality of the results. If the variables are non-stationary at level I(0), then we can take first difference of the variables at level I(1). Normally the acceptable significant levels ranged from 1% to 10% level of confidence. The decision process for unit root tests can be summarized as follow (see Table):

**Table 3.3: Panel unit root tests and hypotheses**

|  |  |  |  |
| --- | --- | --- | --- |
| Panel unit root tests | Levin, Lin and  Chu, LLC | Im, Pesaran and  Shin, IPS | Fisher Type: ADF  and PP tests |
| Null Hypothesis:H0 | Has unit root | Has unit root | Has unit root |
| Alternative  Hypothesis: HA | Has no unit root  (stationary) | Has no unit root  (stationary) | Has no unit root  (stationary) |
| Statistic test | t-statistics | w-statistics | Chi-square |
| Probability < 0.1 | 0.00-0.10 | 0.00-0.10 | 0.00-0.10 |

Source: Author’s illustrations

**3.6.1.4 Panel Granger Causality**

If we check the Y and X are whether cointegrated or not, we might need to examine the causality between these variables. The cointegration test which have been explained in the previous section that we identify the long-run relationship between these variables. In an applied econometrics the Granger (1969) has been widely used as an ideal econometric tool. It is a statistical hypothesis test for determining whether one time series is useful in forecasting another. According to the Granger causality test, in a stationary time series Yt is said to "cause‶ another stationary time series Xt if predictions of the value of Xt based on its own past values and on the past values of Yt are better than predictions of X t based only on its own past values. And it is important to note that ‶y Granger causes x" does not imply that x is the effect or the result of y

In general, the model of panel causality test takes the following form,

= + +…+ ++…+ +

= + +…+ ++…+ +

Where, t denotes the time period of the panel data and i denotes the cross- sectional dimension. There are two simplest approaches to causality test of STATA. The first one is to treat the panel data as one large stacked set of data and then perform the granger causality test in the standard way, with the exception of not letting data from one cross-section enter the lagged values of data from the next cross-section. This method assumes that all coefficients are same across all cross sections, i.e..:

= , = = , i,j

= ,…., i,j

The second approach of Hurlin (2012) makes an extreme opposite assumption and allows all coefficients to be different across cross-section.

,…, i,j

,…, i,j

This test is calculated by simply running standard Granger causality regression for each cross-section individually.

**3.6.2 Panel ARDL Approach**

Dynamic panel data with large N and large T are different from data with large N and small T. Normally; small T data are tested with fixed effect models and random effect models or combination of them with instrumental variable model. They needed the data for pooling individual groups and only intercepts are allowed to differ across cross-sections. For data with large n and large T, it is not consistent to use the assumption of homogeneity of slope parameters. Therefore, with the increase in large N and T dynamic panels, panel ARDL model is useful for the mixed of stationarity and nonstationarity of data at level I(0) and I(1). The following equation is Autoregressive Distributed Lag model (ARDL) with unrestricted specification,

Yit = Yi,t-j + Xi,t-j + +

where xit (k x 1) is vector of regressors for group i, 𝜆𝑖𝑗 represents for the coefficients of lagged dependent variables and 𝜇𝑖 is the fixed effect and they all are scalars and coefficient vectors are 𝛾𝑖𝑗 (k x 1). If the data has large T, we can apply ARDL for individual group estimations. We discuss about three estimators developed by Pesaran and Smith (1995) and by Pesaran et al. (1999) which are mean group (MG), pooled mean group (PMG) and dynamic fixed affect (DFE).

**3.6.3 Pooled Mean Group Estimator**

Pooled mean group (PMG) is estimation methods based on both pooling and averaging and estimate the long-run and short-run correlation with one equation. The estimator allows error variances, intercepts and short-run coefficients to differ freely across countries while long-run coefficients remained constant across groups. PMG estimator is used to apply the homogeneity restrictions on the long run coefficients and averaging across countries to get estimated means for error-correction coefficients and short-run parameters. The good point for using this estimator is that when we estimate for the parameters of dynamic models from panel data that include time series observations, so that individual equations can be estimated for each country especially for the short run coefficients. The main feature is that the responsiveness of co-integrated variables can have any deviation from long-rum equilibrium. Therefore, this deviation from equilibrium influenced on short-run dynamics of variables and thus error correction model needs to apply. This method is mainly relied on maximum likelihood methods and PMG estimators can be applied to any regressors which has stationarity at I(0) or I(1). Based on their asymptotic distributions of PMG estimators, the treatments to stationary regressors and nonstationary regressors are different. PMG with restriction specification is that 𝜃s are common across cross-section data, countries:

=(- ) + +++

𝜃𝑖 = −(1 − ij)

𝜃𝑖 = error correction speed of adjustment term

If 𝜃𝑖 is zero, there has no long-run cointegration.

If 𝜃𝑖 > 0, [no long-run cointegration]

If 𝜃𝑖 < 0, [long-run cointegration].

t = 1,2,…., T (time series)

i = cross-section groups

𝑦𝑖𝑡 = dependent variables of BRI countries i at time

= coefficient k x 1 vecotr of independent variables

𝜆𝑖𝑡= scalars

𝜇𝑖 = fixed effect

𝛽𝑖 = / (1- ∑k),

𝜃𝑖 is the important vector which has long-run relationship between variables. PMG estimations allows for heterogeneous short-run dynamics and common long-run and for examining long-run homogeneity while it does not imposing parameter homogeneity in the short-run.

The advantage of using PMG is that short-run coefficients, speed of adjustment and error variances are allowed to differ across countries but there has a constraint for long-run coefficients to be the same. It also gives out the consistent estimates of long-run parameters and did not take into considerations of whether underlying regressors are stationary, non-stationary or mutually cointegrated. In this research, the author applied Panel ARDL approach as after tested the data with unit root tests, the results are given as stationarity at mixed levels. Panel ARDL model by using PMG for eight models constructed to meet the purpose of this study are discussed below:

= (- ) + +++

Model 1 by using (GDPPC, OPN, IR and LFPR) for independent variables:

=

++ +

+ ++

**3.6.4 Mean Group Estimator**

There are two common ways to apply for panel data which are estimating separate equations for each country and examining the distributions of estimated coefficients across countries (Peasaran 1995). Estimates referred to mean estimates, MG estimator. The estimation gives the averaged estimates of parameters but this does not considered that certain parameters can be in same group. MG estimator allows that intercepts, error variances and slope of coefficient vary across countries. It has the least restrictive process and allows heterogeneity of all parameters with no cross-country restrictive process and allows heterogeneity of all parameters with no cross-country restrictions. MG estimator was derived from the equation for ARDL:

= + 1+ +

where i = 1,2,….,N and t =1,2,…,T, estimation of long run parameter’s coefficient 𝜃𝑖 for country i can be written as: =

Mean Group (MG) estimator for the whole panel can be written as below:

= i

MG’s variance is: v (MG) = (i- MG) (i- MG)

**3.6.5 Dynamic Fixed Effect Estimator (DFE)**

The Dynamic Fixed Effect Model also has restrictions on the coefficients of cointegrating vector to be equal across all panels, short run coefficients and speed of adjustment coefficients to be equal. It also restricts on convergence coefficients and common variance. DFE allows for the greater heterogeneity of the parameters and imposes homogeneity on all slope coefficients. With DFE, all coefficients are resulted similar to PMG and MG estimators as it only allows individual intercepts to differ across countries.

**3.6.6 Error Correction Term**

In the dynamic model, the error correction terms (ECT) can be described as the speed of adjustment to reach the equilibrium. Its coefficient confirmed that how the variables converge or diverge to the equilibrium and it’s sign would be negative or positive. The statistically significant ECT shows that there has a stable long run relationship.

**3.6.7 Hausman Test**

Hausman test is used to determine the best estimation to choose among PMG and MG or PMG and DFE. The hypothesis of homogeneity for the long run parameters are not able to assume as priori. Hausman test determine the effect of heterogeneity on the coefficients’ means and if the parameters are homogenous, PMG’s results are more consistent and efficient than MG’s or DFE’s. It can be said that if null hypothesis is accepted, PMG is preferred upon choosing efficient estimator and if null is rejected, MG or DFE is preferred. The test statistic is described by.

H= (b - B) (b - B)

= (b - B), which is generalized inverse

𝑉is the variance of coefficient

H0: Accept PMG if p value of 𝜆2 > 0.05

HA: Reject MG or DFE if p value of 𝜆2 < 0.05