CHAPTER 3

Methodology

3.1 Conceptual Framework

Based on the previous studies that analyzed the relationship between tourism expansion and economic growth all over the world, this paper studies the impact of tourism expansion on economic growth in Myanmar for the period 1985 to 2015 by using ARDL co-integration approach. With reference to the linkages, proxy variables to capture the tourism sector and economic growth, and theories from the recent papers; the author sets the variables, theories, and research methodology to investigate the linkages between tourism development and economic growth in Myanmar.

The causal relationship between tourism development and economic growth in Myanmar will be explored through the use following variables; the co-integration analysis will be used.

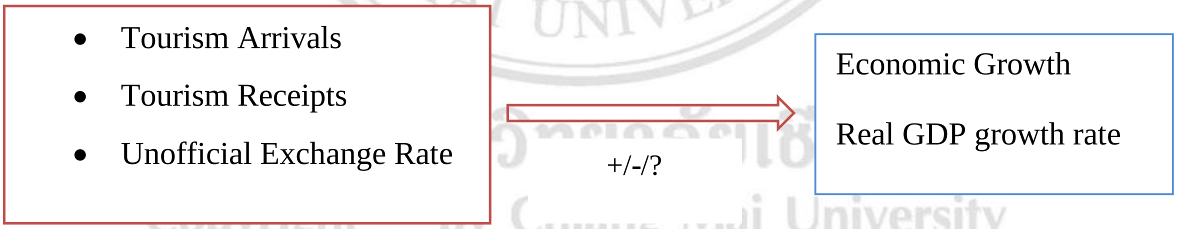


Figure 3.1: Conceptual Framework

The Keynesian multiplier model describes that how much income will change by multiplier times based on changes in autonomous expenditures. Increase in the number of tourist arrivals will lead to an increase in tourism receipts. Income earned from tourism sector in Myanmar will invest on the development of tourism sector and other sectors directly and indirectly; this will lead to the increase in government income and employment opportunities through multiplier effects based on changes in autonomous expenditures. The reason of putting real GDP growth rate as a proxy of economic growth while investigating the impact of tourism expansion on economic growth in Myanmar is that it can reveal the actual economic condition of the country based on previous researches like Katircioglu, S. (2009), Risso, W. A., & Brida, J. G. (2009), Milanovic, M., Stamenkovic, M. (2012), etc.

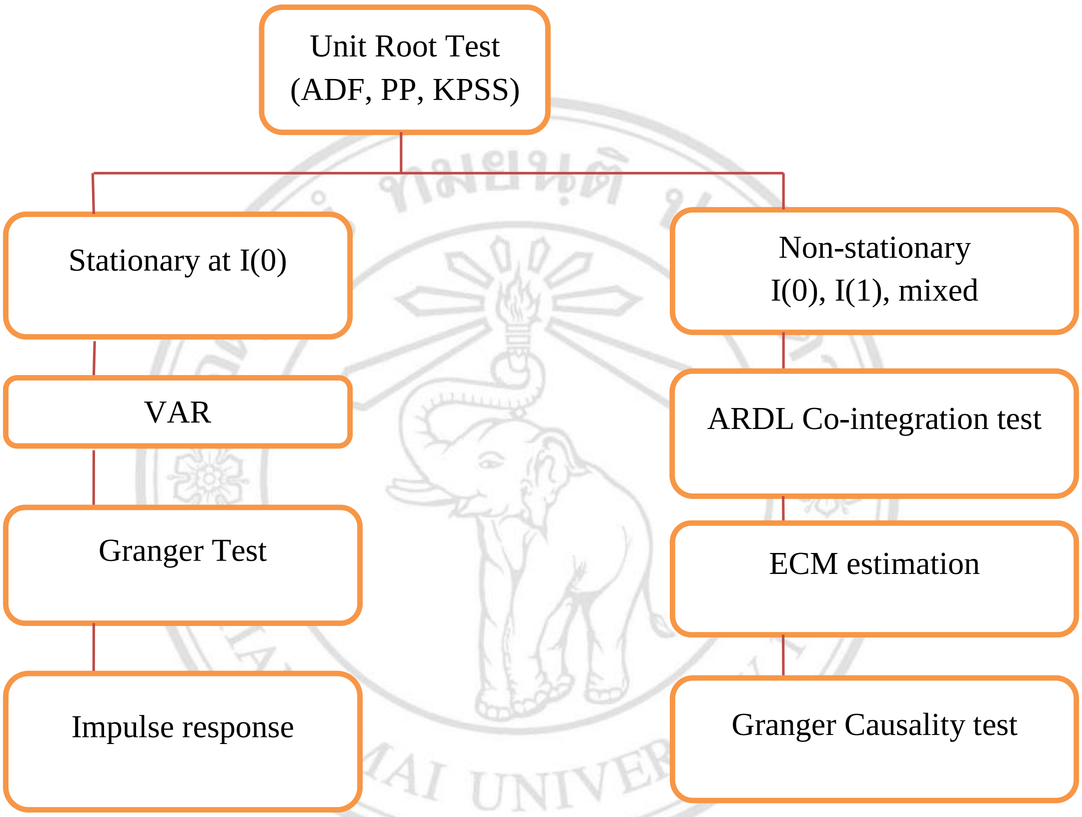


Figure 3.2: The Method of Estimation for the Research

In this research, the real GDP growth rate (RGDP) is set as a proxy of economic growth in Myanmar. To determine the tourism sector, tourism arrivals (TA), tourism receipts (TR) will be used. Due to the Keynesian multiplier effect, tourism sector development will be advantageous to national income and job opportunities in the country through multiplier effect. Besides, most of the studies tested the linkage between tourism and economic growth by using real exchange rate as a controlled variable. Because of lack of data availability, unofficial exchange rate (UER) will be denoted as the proxy for real exchange rate.

Theoretical framework of tourism sector for the function of Myanmar economic growth can be written as follows:

$$\boldsymbol{RGDP_t = f(Tourism\ Sector)} \qquad \textbf{(3.1)}$$

$$\boldsymbol{RGDP_t = f(TA_t, TR_t, UER_t, D_t)} \qquad \textbf{(3.2)}$$

Where $RGDP_t$ denotes Growth Rate of Real Gross Domestic Product

TA_t denotes Tourism Arrivals

TR_t denotes Tourism Receipts

UER_t denotes Unofficial Exchange Rate in the market

D_t denotes structural break dummy variables for 1989 and 2008

To be more multiplicative in interpretation of change in a variable, the variables used in paper except RGDP are taken natural logarithm form. So the model will be written as follows:

$$\boldsymbol{RGDP_t = \beta_0 + \beta_1 lnTA_t + \beta_2 lnTR_t + \beta_3 lnUER_t + \beta_4 D_t + \varepsilon_t} \qquad \textbf{(3.3)}$$

Where t is the time series respectively

$\beta_0, \beta_1, \beta_2, \beta_3, \beta_4$ = Parameters

$RGDP_t$ = Growth Rate of Real Gross Domestic Product

$lnTA_t$ = Natural logarithm of Tourism Arrivals

$lnTR_t$ = Natural logarithm of Tourism Receipts

$lnUER_t$ = Natural logarithm of Unofficial Exchange Rate

D_t = Structural breaks dummy variables for 1989 and 2008

3.2 Sources of Data and Variables used in the Model

The study of the long run and short run relationship of tourism expansion and economic growth in Myanmar was conducted in yearly from 1985 to 2015. Indeed, all the data of the variables employed was collected from various sources: real GDP growth rate from international monetary fund (IMF) and world economic outlook; tourism variables such as tourism arrivals, and tourism receipts from MoHT (Ministry of Hotel and Tourism); and unofficial exchange rate from CIA factbook, IMF and world bank.

Table 3.1 Source of Data and Variables used in the Model

Concept	Variables	Symbol	Unit	Sources
Economic Growth	Real GDP Growth Rate	RGDP	Growth rate (percentage change)	IMF, World Economic Outlook, September 2006 IMF, World Economic Outlook, April 2015
Tourism Sector	Tourism Arrivals	TA	Visitors	Ministry of Hotel and Tourism (MOHT)
	Tourism Receipts	TR	Million in US($)	Kyaw Min Oo (2007) (CSO statistical Yearbook 2015)
Exchange Rate	Unofficial exchange rate	UER	Local Currency Unit (LCU),Kyat	CIA World Factbook World Bank IMF (Sean Junnell)

Source: Author

3.3 Hypotheses of the Study

The four verifiable main hypotheses to examine the impact of tourism expansion on Myanmar economic growth can be viewed as follows:

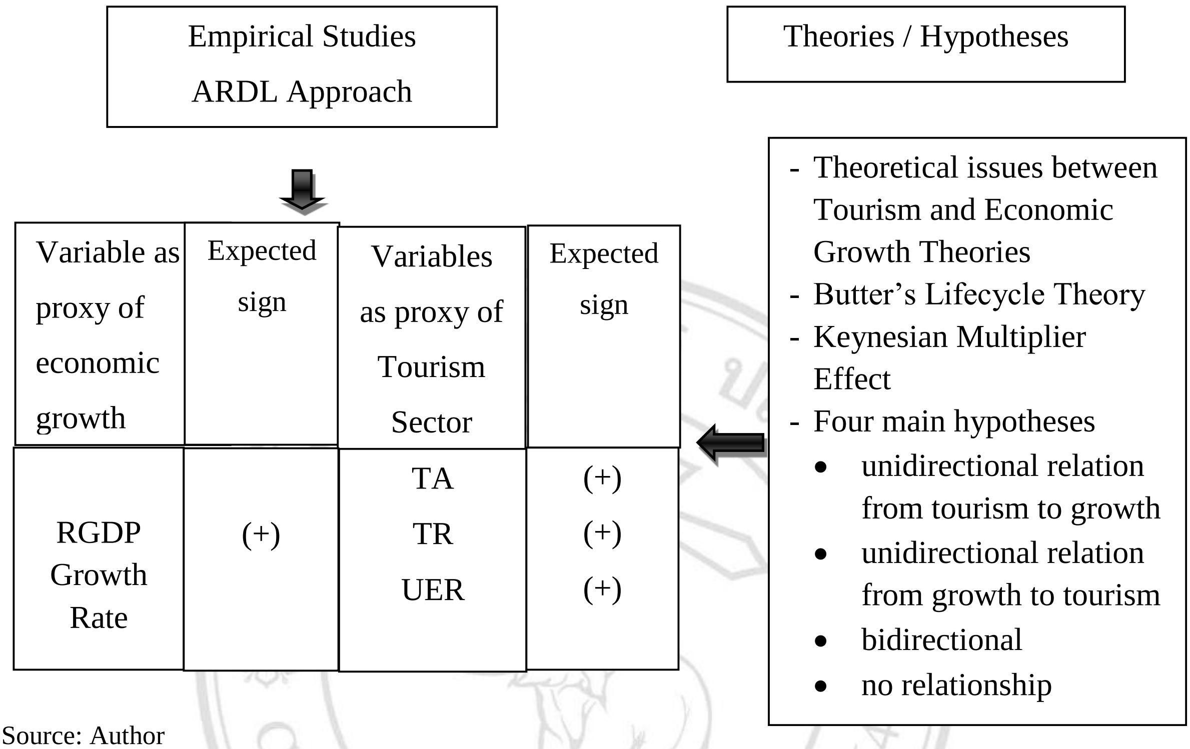
Hypothesis 1 : Tourism-led growth hypothesis which means tourism sector development drives to obtain economic growth

Hypothesis 2 : Growth- led tourism hypothesis that refers to the situation where the growth of the real economy contributes to the tourism sector development

Hypothesis 3 : Tourism expansion and economic growth lean on each other

Hypothesis 4 : Tourism expansion and economic growth do not rely on each other for both parties' development

The following figure shows the captured theories, proxy variables and expected signs based on other research papers.



Source: Author

Figure 3.3: Theories, Variables, Proxies and their Expected Signs

3.4 Research Methodology

3.4.1 Stationary / Non-stationary of time series data

Before the regression, the first step to do is to analyze whether there is stationary or non-stationary of time series data. Time series data is accepted to be weakly stationary if the series' mean, variance and covariance are constant over time. On the other hand, it is assumed to have non-stationary in the time series data. Non-stationary time series have different values of mean, variance and covariance at different time points. If the time series data are non-stationary, the simple ordinary least square (OLS) regression cannot be applied. That causes spurious regression problem. The non-stationary time series variables can be changed into stationary by taking differences in variables. The non-stationary time series can be changed into stationary by taking difference in variables. To define the degree of stationary of time series data, unit root testing will be undertaken through ADF Test, PP Test and KPSS Test.

3.4.2 Unit Root Testing (ADF Test)

Based on several econometric assumptions and hypotheses, several unit root tests exist whether to identify whether our time series data are stationary and integrated of the same order or not. In time series data analysis, the Augmented Dicky-Fuller (ADF) Test is the most widely used. The three different forms of ADF test can be viewed as follows:

When the time series is flat that means it has no trend and potentially slow – turning around Zero, the test equation can be written as:

$$\boldsymbol{X_t} = \boldsymbol{\beta_0 X_{t-1}} + \boldsymbol{\beta_1 \Delta X_{t-1}} + \boldsymbol{\beta_2 \Delta X_{t-2}} + \cdots + \boldsymbol{\beta_p \Delta X_{t-p}} + \boldsymbol{\varepsilon_t} \quad \textbf{(3.4)}$$

Where number of lags (p) is determined by using lag criterion such as Akaike Information Criterion (AIC) or Schwartz Bayesian Information Criterion (SC) till the last Lag (p) is statistically significant.

When the time series is flat and potentially slow turning around a non-zero value (i.e. a random walk with drift), the test equation can be described as:

$$\Delta \boldsymbol{X_t} = \propto + \boldsymbol{\beta_0 X_{t-1}} + \boldsymbol{\beta_1 \Delta X_{t-1}} + \boldsymbol{\beta_2 \Delta X_{t-2}} + \cdots + \boldsymbol{\beta_p \Delta X_{t-p}} + \boldsymbol{\varepsilon_t} \quad \textbf{(3.5)}$$

Where the test has intercept term and no trend. Choosing maximum number of lags will be chosen in the same way to the first situation.

When the time series is potentially slow turning around the trend line (i.e. it has intercept and time trend), the test equation can be viewed as follows:

$$\Delta \boldsymbol{X_t} = \propto + \boldsymbol{\beta t} + \boldsymbol{\beta_0 X_{t-1}} + \boldsymbol{\beta_1 \Delta X_{t-1}} + \boldsymbol{\beta_2 \Delta X_{t-2}} + \cdots + \boldsymbol{\beta_p \Delta X_{t-p}} + \boldsymbol{\varepsilon_t} \quad \textbf{(3.6)}$$

Where $\propto$ = the constant term

β_0 = the coefficient of X_{t-1}

X_{t-1} = value of X_t with the lag in order one

ε_t = the error term

p = maximum number of lags

The null hypothesis of ADF test is that time series variable has a unit root whereas the alternative one is that the variable has a stationary process. It can be shown as:

H_o: $\beta_0 = 0$ (the data has unit root and differenced to make stationary)

H_1: $\beta_0 < 0$ (the data is stationary and no need to be differenced)

The author testify whether the proxy variables in this paper are stationary at level, lag of order one I(1) or lag of order two I(2). The following table shows the Augmented Dicky fuller (ADF) test process of the used variables of this research.

The ADF unit root test has main advantage that it can overcome the auto correlation problem of the error terms in order to obtain valid result. Maximum number of lags is determined by AIC or SC criterion. Using too few lags will cause auto correlation in the errors while many lags will affect badly to the power or validity of the test.

3.4.3 Unit Root Testing (PP Test)

The Phillips-Perron (PP) test is used to take correction in serial correlation and autocorrelation in the case of autocorrelation and heteroskedastic regression residuals. For PP test, the Newey-West (1987) heteroskedasticity and autocorrelation-consistent convarience matrix estimator is used to overcome these problems. The Phillips-Perron test involves fitting the regression. The test equation can be seen as follows:

$$\mathbf{S^2n} = \frac{1}{n-k} \sum_{i=1}^{n} \hat{\mu}_i^2 \quad \textbf{(3.7)}$$

Where μ_i denotes the OLS residual

k denotes the number of covariates in the regression

q denotes the number of Newey-West lags to use

The null hypothesis of PP test is that the variable contains a unit root and is rejected by a stationary process in alternative hypothesis.

The asymptotic distribution of PP test is the same as ADF test and the hypotheses to be tested are as follows:

H_o: the time series data is non-stationary

H_1: the time series data is stationary

It PP test > Mackinnon statistics, it can conclude that the data is stationary or rejected null – hypothesis.

If PP test < Mackinnon statistics, it can conclude that the data is non-stationary as accepted the null hypothesis.

It is preferable to check the stationary process of the proxy variables through various approaches of unit root testing. Despite that, this Phillip-Perron approach may have low power against the stationary near the unit root testing.

3.4.4 Unit Root Testing (KPSS Test)

The first two unit root tests such as ADF test and PP test set the null hypothesis as the situation that it has a unit root and seek to reject this null hypothesis.

Unlike other unit root tests for time series analysis, Kwiatkowski et al. (1992) provides the test of null hypothesis of trend stationary against the alternative of unit root (i.e. non – stationary). The KPSS statistics is based on the residuals from OLS regression Y_t on the exogenous variables X_t. After taking this residual to calculate in the KPSS statistics:

$$\mathbf{KPSS} = \mathbf{T2S}\boldsymbol{S_t^2} \ /_{[\boldsymbol{S}^2\ (\boldsymbol{L})]} \qquad \mathbf{(3.8)}$$

The hypothesis to be tested in the KPSS test can be described as follows:

H_o: $s^2 = 0$ (X_t ~ I (0)): stationary data

H_1: $s^2 < 0$ (X_t ~ I (1)): non-stationary data

The KPSS statistics > Quantities of distribution of KPSS statistics table, it can conclude that the data is non-stationary or rejected the null hypothesis.

The KPSS statistics < Quantities of distribution of KPSS statistics table, it can conclude that the data is stationary or accepted the null hypothesis.

After testing the stationarity of proxy variables by using all these tests such as ADF test, PP test and KPSS test, the test statistics of each test can be compared to obtain validity of unit root testing result. In some cases, the stationary time series data has significance in ADF statistics whereas it has no significance in KPSS statistics. At that point, each variable can take first difference and investigate whether the variable is integrated of order on I (1).

3.4.5 Autoregressive Distributed Lag (ARDL) Co-integration Analysis

After unit root testing whether the proxy time series variables are integrated or stationary of level I (0) or order one I (1), the ARDL co-integration approach can be applied to examine the co-integration relationship between tourism expansion and economic growth of Myanmar.

Looking back to the similar research papers, although the different techniques can be found in co-integration analysis between tourism expansion and economic growth, the main advantage of ARDL co-integration approach is it allows the used variables in the regression to have mix order of integration with I(0) and I(1). The co-integration approach like Johansen co-integration techniques requires the variables at the same order of regression. The author cannot run the simple ordinary least square (OLS) regression when the used variables are not integrated at the level. In this regard, for ARDL approach, this co-integration technique can be applied whether all used variables are stationary at level I(0)or all are at first difference I(1) or at the mixed stationary with I(0) and I(1). It is important to note that the author can't run the ARDL model in time series analysis if the system contains I(2) variables.

Second, the ARDL approach has benefit like the significant outcomes can occur although the sample size or number of observations is small by avoiding some biasness problems. Furthermore, unlike Johansen cointegration technique, the dummy variables can be added to the regression of ARDL cointegration equation. The ARDL approach is the least square regression which contains lags of dependent and explanatory variables. Therefore, lags selection should be suitable.

Pesaran and Shin (1997, 1999) and Pesaran, Shin, and Smith (2001) propose two-step approach in ARDL co-integration regression to determine the long run co-integration relationship between variables in time series analysis. First, the existence of long run relationship among variables needs to be estimated. Second, the long run and short run coefficients of the regression model to be forecasted if the existence of co-integration relationship is found in the first step of ARDL co-integration approach.

To determine the long run co-integration between variables, the simple formula of bound testing of the ARDL approach can be seen as follows;

$$\boldsymbol{Y_t = c + \emptyset_t + \omega_0 Y_{t-1} + \cdots + \omega_p Y_{t-p} + \delta_0 X_{t-1} + \cdots \delta_q Y_{t-q} + \gamma D_1 + \varepsilon_t} \qquad \textbf{(3.9)}$$

Where X_t denotes the explanatory variables
Y_t denotes the dependent variable
c denotes the intercept term
t denotes the time trend,
D denotes to the dummy variable and
ε_t denotes the error term
p and q denotes lag of dependent and explanatory variables respectively

After replacing first difference of each variable ($Y_t = Y_{t-1} + \Delta Y_t$) and ($X_t = X_{t-1} + \Delta X_t$), Error Correctional model can be derived as follows:

$$\Delta \boldsymbol{lnY_t} = \boldsymbol{\alpha_{0Y}} + \sum_{i=1}^{n} \boldsymbol{\beta_{iY}} \,\Delta \boldsymbol{lnY_{t-i}} + \sum_{i=1}^{n} \boldsymbol{\gamma_{iY}} \,\Delta \boldsymbol{lnX_{t-i}} + \boldsymbol{\delta_{1Y} lnY_{t-1}} + \boldsymbol{\delta_{2Y} lnX_{t-1}} + \boldsymbol{\mu_t} \quad \textbf{(3.10)}$$

Where α_{0Y} denotes intercept

Δ refers to the first difference operator

μ_t denotes the error term

lnX_t and $ln\, Y_t$ denote the natural logarithms of the explanatory variables and the dependent variable respectively

In this equation, δ_{1Y} and δ_{2Y} reflects the long run relationships among variables and β_{iY} and γ_{iY} reflects the short run relationships. The null hypothesis of the bound test refers to the condition that the existence of long run cointegration relationship isn't found against the alternative hypothesis.

H_0: $\delta_{1Y} = \delta_{2Y} = 0$ (The absence of Long run cointegration)

H_1: $\delta_{1Y} \neq \delta_{2Y} \neq 0$ (The existence of Long run cointegration)

In the interpretation of bound testing, the critical values of upper bound and lower bound of the coefficient restriction test (F-test) supported by Pesaran et-al, 2001 are used to determine the long run co-integration of the equation. If the resulted F value is lower than the upper and lower bound critical values, it means there is no co-integration among variables as accepted the null hypothesis. Unlike this, if the resulted F value is higher than those critical values, it can be assumed that the co-integration relationship exists among variables as it is rejected the null hypothesis.

Looking forward to the next step of examining the long run and short run coefficients of the regression model of the research paper, the vector error correction mechanism (VECM) will be tested. The error correction mechanism is used to measure the speed of adjustment towards the long run equilibrium in the future or not.

$$\Delta \boldsymbol{lnY_t} = \boldsymbol{\alpha_{0Y}} + \sum_{i=1}^{n} \boldsymbol{\beta_{iY}} \,\Delta \boldsymbol{lnY_{t-i}} + \sum_{i=1}^{n} \boldsymbol{\gamma_{iY}} \,\Delta \boldsymbol{lnX_{t-i}} + \emptyset \boldsymbol{ECT_{t-1}} + \boldsymbol{\mu_t} \quad \textbf{(3.11)}$$

Where ECT_{t-1} expresses the short run dynamic mechanism

$\emptyset$ denotes the adjustment coefficient for ECM

α_{0Y} denotes intercept

β_{iY} and γ_{iY} denotes the coefficient terms

Δ refers to the first difference operator

μ_t denotes the error term

lnX_t and $ln\ Y_t$ denote the natural logarithms of the explanatory variables and the dependent variable respectively

As the author use the ADF test, PP test and KPSS test for unit root testing of the used proxy variables to examine the cointegration relationship between tourism expansion and economic growth of Myanmar, the stationary outcomes of the time series variables will be resulted depending on the level of integration. Additionally, the ARDL cointegration approach is applied to investigate the long run cointegration of the regression. What's more, the ECM mechanism is taken to know the long run and short run coefficients of the regression.

Using the proxy variables of the research paper, the simple formula of bound testing of the ARDL approach can be expressed as follows:

$$\boldsymbol{RGDP_t = \alpha_0 + \sum_{i=1}^{n} \beta_i RGDP_{t-i} + \sum_{i=1}^{n} \gamma_i lnTA_{t-i} + \sum_{i=1}^{n} \delta_i lnTR_{t-i} + \sum_{i=1}^{n} \theta_i lnUER_{t-i} + \emptyset_5 D_1 + \emptyset_6 D_2 + \mu_t} \qquad \textbf{(3.12)}$$

Where $RGDP_t$ = Real Gross Domestic Product (growth rate)

TA_t= Tourism Arrivals

TR_t= Tourism Receipts

UER_t = Unofficial Exchange Rate

D_1 And D_2 = Structural break dummy variables for 1989 and 2008

α_0 = Constant term

$\beta\ _i$ = The coefficient of real growth domestic product

γ_i = The coefficient of tourism arrivals

δ_i = The coefficient of tourism receipts

θ_i = The coefficient of unofficial exchange rate

μ_t = The error term

The hypotheses of bound testing for the research can be viewed as follows:

H₀: $\beta_1 = \gamma_2 = \delta_3 = \theta_4 = \emptyset_5 = \emptyset_6 = 0$ (absence of long run)

H₁: $\beta_1 \neq \gamma_2 \neq \delta_3 \neq \theta_4 \neq \emptyset_5 \neq \emptyset_6 \neq 0$ (existence of long run)

If the resulted outcome value is above the critical values of the upper bounds, it is noted that the long run co-integration relationship occurs among the variables as it can reject the null hypothesis. Likewise, if the result value is below the lower limit, the outcome can explain there is no long run co-integration relationship and the null hypothesis is rejected in bound testing. If the result value occurs between the lower bound and upper bound critical values, the result is said to be inconclusive.

After testing the existence of the long run relationship in time series co-integration analysis, the coefficients of long run and short run regression can be estimated using the error correction mechanism.

The short run error correction model's equation can be written as follows:

$$\boldsymbol{\Delta lnRGDP_t = \alpha_{0Y} + \sum_{i=1}^{n} \beta_i \Delta lnRGDP_{t-i} + \sum_{i=1}^{n} \gamma_i \Delta lnTA_{t-i} + \sum_{i=1}^{n} \delta_i \Delta lnTR_{t-i} + \sum_{i=1}^{n} \theta_i \Delta lnUER_{t-i} + \emptyset_5 D_1 + \emptyset_6 D_2 + \rho ECT_{t-1} + \mu_t} \quad \textbf{(3.13)}$$

Where $RGDP_t$ = Real Gross Domestic Product (growth rate)

$LnTA_t$= Natural Logarithm of Tourism Arrivals

$LnTR_t$= Natural Logarithm of Tourism Receipts

$LnUER_t$= Natural Logarithm of Unofficial Exchange Rate

D_1 And D_2 = Structural break dummy variables for 1989 and 2008

α_{0Y} = Constant term

β_i = The coefficient of real growth domestic product

γ_i = The coefficient of tourism arrivals

δ_i = The coefficient of tourism receipts

θ_i = The coefficient of unofficial exchange rate

$\rho =$ The coefficient of the short run for ECM mechanism

$ECT_{t-1} =$ The error correction mechanism lagged by one year

μ_t = The error term

Autoregressive Distributive Lag (ARDL) Co-integration Approach employed by the author in this research has some basic assumptions:

- Data must be free from autocorrelation
- Data must be free from heteroscedasticity
- Data must be normally distributed
- Lags must be appropriate
- Errors must be serially independent
- Model must be dynamically stable

In this case, to check the stability of the parameters of the regression, the multiple breakpoint test will be applied to know whether the two time series sector has structural breaks or not. It is noted that the structural breaks can be found when there is the sudden change in the time series or the relationship between the two time series. (Rob J Hyndman's blog) In the time series regression, the structural breaks can be found through two tests_ Chow breakpoint test and multiple breakpoints test. In economic situation, the structural break can occur when the country suffers situations such as the Government change, war occurrence, financial and economic crisis, weather shock and other similar conditions. Out of the two breakpoint check tests, Chow breakpoint test is used for single break occurrence. It can also be tested if the author knows the date when the break occurs.

Unlike chow test, the multiple break points test is more suitable for the condition when the time series data have more than one sudden change or when the researcher does not know about the exact breaks time. To check the maximum number of the breaks at which break levels, the multiple break points test consider the trimming percentage of the sample size, the breakpoint significant level and assumptions relating with computation of variance matrices used in structural break testing. In this connection, the small value of the trimming percentage estimates the coefficients and variances regarding few observations.

The most important thing to point out in structural break testing is that the researcher needs to be careful about using too many dummy variables for setting structural breaks in small sample size. It can lead to the false results of the regression. Therefore, notable point is that if the research does not enough sample size to run the regression, introducing too many variables in the equation should be avoided. That can cause several degrees of freedoms in model. Besides, as the several degrees of freedom dwindle, statistical inference becomes unreliable.

As ARDL co-integration approach has basic assumptions to follow, the overall stability of the model is residual diagnostic tests such as correlogram Q-statistics test, correlogram squared residuals test, normality test, serial correlation test and heteroscedasticity test; and the stability diagnostic test like cumulative sum of squares (CUSUM squares) test proposed by Brown et al. (1975).

3.4.6 Granger Causality Test

In regression, the granger-causality test is employed to check whether the tourism expansion and economic growth of Myanmar have causality or not. The granger -causality technique is selected in order to analyze causation in regression analysis. The four main hypotheses can be testified to know the causal direction from which sector to which. The Granger causality equations can be viewed as follows:

$$Y_t = \alpha_{10} + \Sigma_{j=1}^{k} \alpha_{1j} Y_{t-j} + \Sigma_{j=1}^{k} \beta_{1j} X_{t-j} + u_t \quad (3.14)$$

$$X_t = \alpha_{20} + \Sigma_{j=1}^{k} \alpha_{2j} Y_{t-j} + \Sigma_{j=1}^{k} \beta_{2j} X_{t-j} + v_t \quad (3.15)$$

Where Y_t is dependent variable

X_t is explanatory variable

u_t *and* v_t are the error term

If $\Sigma_{j=1}^{k}\alpha_{ij}$ and $\Sigma_{j=1}^{k}\beta_{ij}$ equal to zero for i=1, 2; it can be assumed that event X and event y do not help each other. If $\Sigma_{j=1}^{k}\alpha_{ij}$ and $\Sigma_{j=1}^{k}\beta_{j}$ are not equal to zero for i=1, 2; the situation can explain event X and event Y contribute to each other (i.e. bidirectional causality). If $\Sigma_{j=1}^{k}\alpha_{ij}$ equal to zero for i=1, 2 and $\Sigma_{j=1}^{k}\beta_{ij}$ is not equal to zero for i=1,2; past values of event X_{t-j} affects to event Y_t for equation 3.14 and event X_{t-j} affects to event X_t for equation 3.15. It can be concluded that unidirectional causality from X to Y exist. If $\Sigma_{j=1}^{k}\alpha_{ij}$ is not equal to zero for i=1,2 and $\Sigma_{j=1}^{k}\beta_{ij}$ equal to zero for i=1, 2: it can be supposed that there is the existence of causality form event Y_{t-j} to event X_t as past values of event Y can predict event X for equation 3.15 and event Y_{t-j} affects to event Y_t for equation 3.14.

However, it does not mean X cause Y or Y cause X. It refers to the situation that X might be causing Y or Y might be causing X.

Granger causality test is based on the calculation of F-statistics. If the F-statistics is greater than the critic value of the chosen significant level, it is considered to reject the null hypothesis, and vice versa.

The interpretation of the results of the Granger Causality Test should be careful. The result does not represent changes in one variable do cause changes in another variable. The outcomes of the test only provide the idea that the correlation between the current value of one variable and the previous values of another variable exist on this research.

The Granger causality method is approvable in causality analysis. This is because it takes into account the possibility bidirectional causality over time. However, sometimes the cause and effect between these two sectors can be linked indirectly each other; and the third party or factor which is also called confounding factor can be occurred in the causality regression. In this connection, the outcomes of the causation can be wrong. Therefore, the author needs to make ensure that there is no confounding omitted variable.

The choice of the lags in the causality regression is also essential. If the lag chosen is different from the real lag, the result can lead to biased or inefficient condition.

In this regard, the validity of the granger causality test depends on the right choice of the number of lags and the stationarity of the used variables in the research paper.