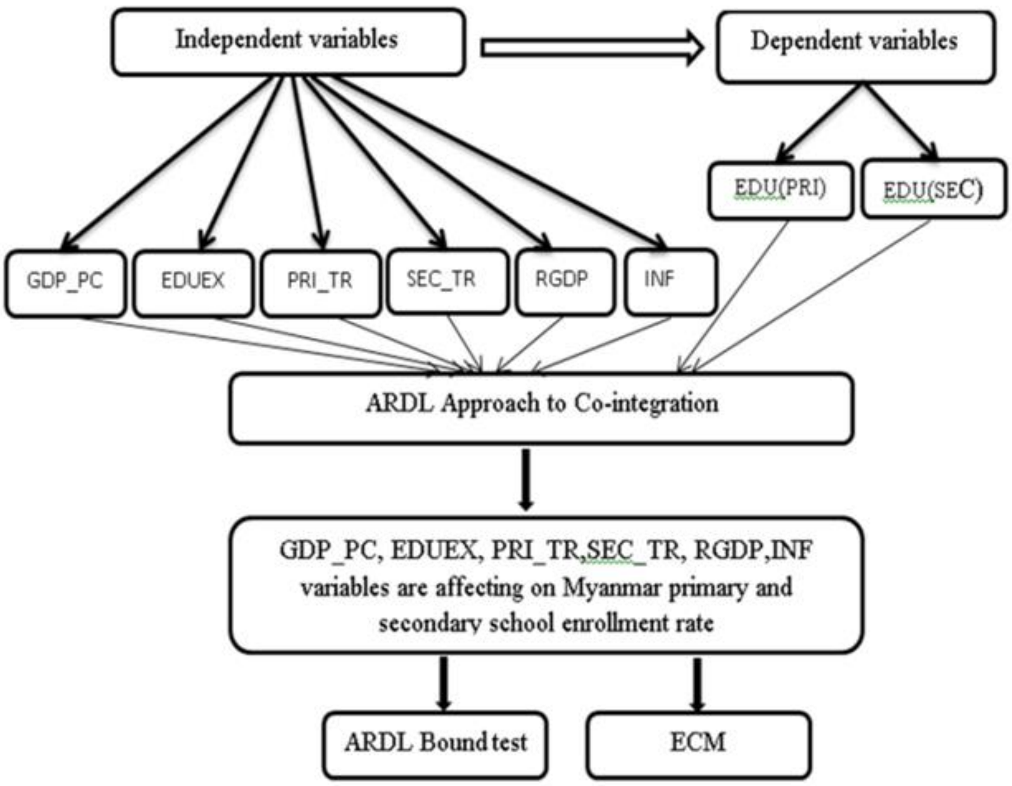
CHAPTER 3

Methodology

3.1 Conceptual Framework

This paper examines the long-run relationship between the macroeconomics variables namely, gross domestic product per capita ($US) as the proxy for personal income, real GDP growth rate as the proxy for Myanmar economy growth rate, inflation as the proxy for macroeconomics stability, education expenditure (% of GNI), primary school teachers and second school teachers are the proxy of educational variables that affected to the school enrollment rate on gross primary school enrollment, gross secondary school enrollment, in Myanmar.



Source: Author

Figure0.1: Research Methodology

In this study, Autoregressive Distributed lagged model is used to analyze the long-run co-integration between the variables and also the error correction model is used in order to examine the short run relationship.

$$EDU(pri)_t = f\left(GDPPC_{(t)}, EDUEX_t, PRI\_TR_{(\mathrm{t})}, INF_{(\mathrm{t})}, RGDP_{(\mathrm{t})}, D_t\right) \quad (3.1)$$

$$EDU(sec)_t = f(GDPPC_{(t)}, EDUEX_t, SEC\_TR_{(\mathrm{t})}, INF_{(\mathrm{t})}, RGDP_{(\mathrm{t})}, D_t) \quad (3.2)$$

$$lnEDU(pri)_t = \beta_0 + \beta_1 lnGDPPC_t + \beta_2 lnEDUEX_t + \beta_3 lnSECTR_t + \beta_4 INF_t + \beta_5 RGDP_{t+} + \beta_6 D + \mu_1 \quad (3.3)$$

$$lnEDU(sec)_t = \propto_0 + \propto_1 lnGDPPC_t + \propto_2 lnEDUEX_t + \propto_3 lnSEC_{TR_t} + \propto_4 INF_t + \propto_5 RGDP_{t+} + \propto_6 D + \mu_1 \quad (3.4)$$

Where,

$lnEDU_{\mathrm{pri}(t)}$ = natural logarithmic of gross primary school enrollment rate (%)

$lnEDU_{sec(t)}$=natural logarithmic of gross secondary school enrollment rate (%)

$lnEDUEX_t$ =natural logarithmic of education expenditure as percentage of GNI

$lnPRI\_TR_t$ = natural logarithmic of number of primary school teachers

$lnSEC\_TR_t$ =natural logarithmic of number of secondary school teachers

$INF_{(t)}$ = inflation rate

$RGDP_t$ = real GDP growth rate

D = the structural break dummy variable.

Many researchers analyzed the relationship between the economics variables and the school enrollment rate. Emmanuel Carsamer and Eric Ekyem (2015) analyzed the relationship between the educational expenditure and the primary and secondary school level in Africa, and the other the researcher, Risikat Oladoyin S.Dauda, examined the effect of public educational spending and macroeconomic uncertainty on schooling outcomes in Nigeria. This paper followed by the hypothesis of the above papers. However, this paper focuses on which macroeconomics variables are the most affecting on school enrollment of Myanmar.

In this paper, the first hypothesis is that the primary school enrollment rate is a function of the gross domestic product per capita, education expenditure as the percentage of GNI, primary school teachers, inflation rate, and the real GDP growth rate. The second hypothesis is that the secondary school enrollment rate is the function of the gross domestic product per capita, education expenditure as the percentage GNI, secondary school teacher, inflation rate, and the real GDP growth rate.

3.2 Sources of the Data and the variables description

This paper used the annual time series data from 1985 to 2015 of Myanmar. Most of the data was collected from the World Bank data. Some was collected from the UNESCO, IMF, and UNDP Human Development Report. In this paper, most of the variables were used in the natural logarithmic form before the regression. However two variables were not used in the logarithmic form because these variables are real GDP growth rate and the inflation rate. The rest variables are the gross enrollment ratio, number of teachers, the percentage of the education expenditure, and the current exchange rate (US dollars). Table 3.1 shows the variables index and the source of the variables.

Table 0.1: Review of variables

Variables	Indicators	Measures	Symbols	Excepted Sign	Sources
Primary School Enrollment rate	Gross primary school enrollment ratio	Primary Gross enrollment ratio	EDU_PRI	(+)	World Bank Data (UNESCO Institute for Statistics)
Secondary School Enrollment rate	Gross secondary school enrollment ratio	Secondary Gross enrollment ratio	EDU_SEC	(+)	World Bank Data (UNESCO Institute for Statistics
Gross Domestic Product per capita	Gross Domestic Product per capita($US)	Gross Domestic Product is divided by population($US)	GDPPC	(+)	World Economic Outlook (International Monetary Fund)
Education Expenditure	Public expenditure on education	Education expenditure as the percent of GNI (%)	EDUEX	(+)	UNDP annual report
Primary School teachers	Number of primary school teachers	Total number of primary school teachers in public and private institutions	PRI_TR	(+)	World Bank Data (UNESCO Institute for Statistics)
Secondary School teachers	Number of secondary school teachers	Total number of secondary school teachers in public and private institutions	SEC_TR	(+)	World Bank Data (UNESCO Institute for Statistics)
Inflation rate	Annual Inflation rate (consumer price index)	Inflation rate measured by annual percentage change of the consumer price index	INF	(-)	International Monetary Fund
Real Gross GDP rate	real GDP growth rate	GDP growth rate (%)	RGDP	(+)	World Economic Indicators

Source: Author

3.3 Econometric Methods

3.3.1 Augmented-Dickey Fuller Unit Root Test

Unit root test is used to test if time series data is non-stationary or stationary. For time series, y_t, unit root test can be done by examining one or all of the following three regression models:

Intercept and Trend: $\Delta Y_t = \beta_1 + \beta_2 t + \rho Y_{t-1} + \beta \sum_{i=1}^{m} \Delta Y_{t-1} + \varepsilon_t$ (3.5)

Intercept: $\Delta Y_t = \beta_1 + \rho Y_{t-1} + \beta \sum_{i=1}^{m} \Delta Y_{t-1} + \varepsilon_t$ (3.6)

No Intercept and No Trend: $\Delta Y_t = \rho Y_{t-1} + \beta \sum_{i=1}^{m} \Delta Y_{t-1} + \varepsilon_t$ (3.7)

Where,

β_1 = the constant term

β_2 = the coefficient of time trend

ρ = the coefficient of Y_{t-1}

Y_{t-1} = the values of the lag in order one Y_t

ΔY_{t-1} = changes the values of the lag

ε_t = white noise error term.

m = the lags order of the autoregressive process.

The value of the parameter ρ is the most important in the ADF test. We can test the null hypothesis and alternative:

$H_0 : \rho = 0$ (Indicating a stationary at level)

$H_1 : \rho < 0$ (Indicating a non-stationary at level).

The DF test statistics are

$$DF_\tau = \frac{\widehat{\beta}-1}{Estimated\ Standard\ error} \quad (3.8)$$

And the Augmented Dickey Fuller test statistics is

$$DF_\tau = \frac{\widehat{\beta}-1}{1-\widehat{\beta}-\ldots-\widehat{\beta}_p} \quad (3.9)$$

The ADF test procedure is similar to the DF test so we can use the same critical value of the DF test. However, ADF test overcomes the difficulties of the error terms are correlated by including the lagged difference values (Gujarati, 2008). By testing with ADF unit root test, the test critical value of the DF test is used. If the calculated critical value is above the DF test critical value, the null hypothesis of the ADF unit root test is rejected, which means there is no the unit root problem and the time series data are stationary. If the calculated critical value is lower than the DF test statistic, the null hypothesis is failed to reject which means that there is a unit root problem. The Phillips-Perron (PP) test is similar to the ADF unit root test. PP test is also taking care of the correlation between error terms.

3.3.2 PP Unit Root Test

The difference between the PP unit root test and ADF test is the ability to deal with the problem of serial correlation, and heteroscedasticity in the error term. Although ADF test and PP test are taking care about the error correlation, PP test is not adding lagged difference terms. The test regression of the PP test is as follows:

Intercept and Trend: $\Delta Y_t = \beta_1 + \beta_2 t + \rho Y_{t-1} + \varepsilon_t$ (3.10)

Intercept: $\Delta Y_t = \beta_1 + \rho Y_{t-1} + \varepsilon_t$ (3.11)

No Intercept and No Trend: $\Delta Y_t = \rho Y_{t-1} + \varepsilon_t$ (3.12)

The specific hypotheses are as follow:

$H_0 : \rho = 0$ (Indicating a stationary at level)

$H_1 : \rho < 0$ (Indicating a non-stationary at level)

To check if the used variables are stationary or non-stationary, we used the unit root testing through the developed Augmented Dickey-Fuller (ADF) test. We will conclude which order integrated for the variables, and it is tested with constant but no trend and with constant and trend.

$$\Delta lnEDU(pri)_t = a_1 + a_2 t + \rho lnEDU(pri)_{t-1} + a \sum_{i=1}^{m} \Delta\, lnEDU(pri)_{t-1} + \varepsilon_t \quad (3.13)$$

$$\Delta lnEDU(sec)_t = b_1 + b_2 t + \alpha lnEDU(sec)_{t-1} + b \sum_{i=1}^{m} \Delta\, lnEDU(sec)_{t-1} + \varepsilon_t \quad (3.14)$$

$$\Delta lnEDUEX_t = c_1 + c_2 t + \mu lnEDUEX_{t-1} + c \sum_{i=1}^{m} \Delta\, lnEDUEX_{t-1} + \varepsilon_t \quad (3.15)$$

$$\Delta lnPRI\_TR_t = d_1 + d_2 t + \gamma PlnRI\_TR_{t-1} + d \sum_{i=1}^{m} \Delta\, lnPRI\_TR_{t-1} + \varepsilon_t \quad (3.16)$$

$$\Delta lnSEC\_TR_t = e_1 + e_2 t + \delta lnSEC\_TR_{t-1} + e \sum_{i=1}^{m} \Delta\, lnSEC\_TR_{t-1} + \varepsilon_t \quad (3.17)$$

$$\Delta lnGDPPC_t = f_1 + f_2 t + \lambda GDPPC_{t-1} + f \sum_{i=1}^{m} \Delta\, GDPPC_{t-1} + \varepsilon_t \quad (3.18)$$

$$\Delta RGDP_t = g_1 + g_2 t + \phi RGDP_{t-1} + g \sum_{i=1}^{m} \Delta\, RGDP_{t-1} + \varepsilon_t \quad (3.19)$$

$$\Delta INF_t = h_1 + h_2 t + \upsilon SEC\_TR_{t-1} + h \sum_{i=1}^{m} \Delta\, SEC\_TR_{t-1} + \varepsilon_t \quad (3.20)$$

$lnEDU(pri)_t$= natural logarithmic form of gross primary school enrollment rate of Myanmar (%) at time t and t

$lnEDU(sec)_t$= natural logarithmic form of gross secondary school enrollment rate of Myanmar(%) at time t and t-1

$lnEDU\_EX_t$ = natural logarithmic form of education expenditure as the percentage of GNI at time t ans t-1

$LnGDPPC_t$ = natural logarithmic form of GDPPC in terms of $US

$lnPRI\_TR_t$ = natural logarithmic form of numbers of primary school teachers

$lnSEC\_TR_t$ = natural logarithmic form of secondary school teachers

$RGDP_t$ = real gross domestic product

INF_t = inflation rate (%)

t = the time trend 30 years

m = lag term

ε_t =error term

Hypothesis test;

$H_0 : \rho = 0$ (Indicating a stationary at level)

$H_1 : \rho < 0$ (Indicating a non-stationary at level)

3.3.3 ARDL Approach to Co-integration

The autoregressive distributed lag (ARDL) model was formulated by Pesaran and Shin. This approach is more appropriate for co-integration analysis. There are some advantages for the estimation results. First, the researcher can get a better estimation result although the sample size is small. Second, ARDL approach can be used for variables with order of integration both I(0) and I(1). Third, the unbiased and better estimation results can be received although some of the variables are endogenous. Fourth, ARDL approach is a suitable procedure if the researcher included the dummy variable into the equation.

The bound testing of the ARDL approach is suitable approach to test the long-run relationship between variables. The simple formula of the ARDL (p,q) model,

$$Y_t = c + \emptyset_t + \omega_0 Y_{t-1} + \cdots + \omega_p Y_{t-p} + \delta_0 X_{t-1} + \cdots \delta_q Y_{t-q} + \gamma D_1 + \varepsilon_t \quad (3.21)$$

Where c is the intercept term, t is the time trend, D refers to the dummy variable and ε is white noise error term and also the dependent and independent variables are the stationary variables. The formula of the ARDL approach involved the error-correction model developed by Pesaran and Shin (1997,1999) and Pesaran, Shin, and Smith (2001).

$$\Delta lnY_t = \alpha_{0Y} + \sum_{i=1}^{n} \beta_{iY} \Delta lnY_{t-i} + \sum_{i=1}^{n} \gamma_{iY} \Delta lnX_{t-i} + \delta_{1Y} lnY_{t-1} + \delta_{2Y} lnX_{t-1} + \mu_t \quad (3.22)$$

For equation (3.21) and (3.22), using F-test is suitable test to determine one or more long-run relationship between dependent and independent variables.

For equation 3.22, null hypothesis (H_0) is: $\delta_{1Y} = \delta_{2Y} = 0$, which means that there is no long-run co-integration between variables. Alternative hypothesis (H_1) is: $\delta_{1Y} \neq \delta_{2Y} \neq 0$, which means that there is long-run co-integration.

To test the hypothesis, we used the critical value of the F-statistics in Pesaran, Shin, and Smith (2001). There are two sets of critical values called the upper bound values and lower bound values. The economists argued that if the computed F-statistics is higher than the critical value in the upper bound test, the null hypothesis is rejected. This means that there is long-run co-integration between variables. If it is below the lower bound test, the null hypothesis is accepted that there is no long-run co-integration between variables. The other condition is if its value between the lower and upper bounds, the estimation result would not be defined conclusive result.

The long-run relationship between the variables means that there is Granger-causality in at least one direction. After testing the long-run relation between the variables, the next procedure is to test short run dynamic parameter in the vector error correction model. The short-run causal effect is represented by the F-statistic on the explanatory variables, and the t-statistics on the coefficient of the lagged error-correction term represents the long-run causal relationships. (Odhiamb,2009; Narayan and Smyth, 2006). The error-correction model is:

$$\Delta lnY_t = \alpha_{0Y} + \sum_{i=1}^{n} \beta_{iY} \Delta lnY_{t-i} + \sum_{i=1}^{n} \gamma_{iY} \Delta lnX_{t-i} + \emptyset ECT_{t-1} + \mu_t \quad (3.23)$$

Where, β_{iY} and γ_{iY} are the coefficient terms that indicate the short run dynamic model.ECT_{t-1} is the error correction term lagged by one period and $\emptyset$ is adjustment coefficient. And μ_t is the error term.

After testing the unit root test for all of the variables which are used in this paper, Autoregressive Distributed Lagged model (ARDL) was used in order to know long-run co-integration between the variables. After determining the long-run relationship, we estimate the short-run error correction model.

The macroeconomic variables namely the gross domestic product per capita(GDPPC), education expenditure as the percentage of GDP, primary school teachers, secondary school teachers, inflation rate and the growth rate of the real gross domestic product influencing on primary school enrollment rate, secondary school enrollment covering three ARDL economics models for the two lag length structure as following:

$$lnEDU(pri)_t = \beta_0 + \alpha_1 LnEDU(pri)_{t-1} + \alpha_2 lnEDU(pri)_{t-2} + \rho_0 lnGDPPC_t + \rho_1 lnGDPPC_{t-1} + \rho_2 lnGDPPC_{t-2} + \emptyset_1 lnEDUEX_t + \emptyset_1 lnEDUEX_{t-1} + \emptyset_2 lnEDUEX_{t-2} + \delta_1 lnPRITR_t + \delta_2 lnPRITR_{t-1} + \delta_3 lnPRITR_{t-2} + \theta_1 lnINF_t + \theta_2 lnINF_{t-1} + \theta_3 lnINF_{t-2} + \mu_0 lnRGDP_t + \mu_1 lnRGDP_{t-1} + \rho_2 lnRGDP_{t-2} + D_0 + \varepsilon_t \quad (3.24)$$

$$lnEDU(sec)_t = \beta_1 + \theta_1 LnEDU(sec)_{t-1} + \theta_2 lnEDU(sec)_{t-2} + \rho_0 lnGDPPC_t + \rho_1 lnGDPPC_{t-1} + \rho_2 lnGDPPC_{t-2} + \emptyset_1 lnEDUEX_t + \emptyset_1 lnEDUEX_{t-1} + \emptyset_2 lnEDUEX_{t-2} + \upsilon_1 lnSECTR_t + \upsilon_2 lnSECTR_{t-1} + \upsilon_3 lnSECTR_{t-2} + \theta_1 INF_t + \theta_2 INF_{t-1} + \theta_3 INF_{t-2} + \mu_0 RGDP_t + \mu_1 RGDP_{t-1} + \rho_2 RGDP_{t-2} + D_1 + \varepsilon_t \quad (3.25)$$

Where,

D denotes the dummy variables for the structural break

$\beta_0$denotes the constant term of the primary school enrollment rate

β_1 denotes the constant term of the secondary school enrollment rate

$\alpha_i$denotes coefficient of primary school enrollment rate

θ_i denotes secondary school enrollment rate.

$\rho_i$denotes coefficient of gross domestic product per capita ($US)

$\emptyset_i$denotes coefficient education expenditure as the percentage of GDP(%)

δ_i denotes coefficient of the primary school teachers.

υ_i denotes coefficient of the secondary school teachers.

θ_i denotes coefficient of the inflation rate

λ_i denotes coefficient of the real GDP growth rate

μ_t denotes error terms.

Table 0.2: ARDL bound test for long-run co-integration

Variable	F-statistic	5% critical value		Number of m
		Lower bound	Upper Bound	
F(EDUpri|,GDPPCt,EDUEXt,PRI_TRt,INFt,RGDPt,Dt).	H0,H1	H0,H1	H0,H1	M
F(EDUsec|,GDPPCt,EDUEXt,SEC_TRt,INFt,RGDPt,Dt).	H0,H1	H0,H1	H0,H1	M

Note: * Statically significant at 1% level, **statically significant at 5% level, ***statically significant at 10% level.

If the computed F-statistics is higher than the critical value in the upper bound test, the null hypothesis is rejected. This means that there is long-run co-integration between variables. If it is below the lower bound test, the null hypothesis is accepted that there is no long-run co-integration between variables. The other condition is if its value between the lower and upper bounds, the estimation result would not be defined as a conclusive result.

3.3.4 Short-run Error correction model

$$\Delta lnEDU(pri) = \beta_0 + \beta_1 \sum_{i=1}^{a} \Delta lnEDU(pri)_{t-i} + \beta_2 \sum_{i=1}^{a} \Delta lnGDPPC_{t-i} + \beta_3 \sum_{i=1}^{a} \Delta lnEDUEX_{t-i} + \beta_4 \sum_{i=1}^{a} \Delta lnPRI\_TR_{t-i} + \beta_5 \sum_{i=1}^{a} \Delta lnINF_{t-i} + \beta_6 \sum_{i=1}^{a} \Delta lnRGDP_{t-i} + \beta_6 t + \beta_7 D_0 + \theta ECT_{t-1} + \varepsilon_t \quad (3.26)$$

$$\Delta lnEDU(sec) \ = \beta_0 + \beta_1 \sum_{i=1}^{a} \Delta lnEDU(sec)_{t-i} + \beta_2 \sum_{i=1}^{a} \Delta lnGDPPC_{t-i} + \beta_3 \sum_{i=1}^{a} \Delta lnEDUEX_{t-i} + \beta_4 \sum_{i=1}^{a} \Delta lnSEC\_TR_{t-i} + \ \beta_5 \sum_{i=1}^{a} \Delta lnINF_{t-i} + \beta_6 \sum_{i=1}^{a} \Delta lnRGDP_{t-i} + \beta_6 t + \beta_7 D_1 + \theta ECT_{t-1} + \varepsilon_t \qquad (3.27)$$

Where,

EDU (pri) denotes gross primary school enrollment rate (%)

EDU (sec) denotes gross secondary school enrollment rate (%)

GDP_PC denotes GDP percapita in Myanmar

EDUEX denotes the education expenditure (as the percentage of GDP)

PRI_TR denotes number of the primary school teachers

SEC_TR denotes number of the secondary school teachers

INF denotes the inflation rate.

RGDP denotes the real GDP growth rate.

D denotes the dummy variables for the structural break

α_0 denotes constant

β_i denotes coefficient of primary school enrollment rate and secondary school enrollment rate.

δ_i denotes coefficient of gross domestic product per capita ($US)

$\emptyset_i$ denotes coefficient education expenditure as the percentage of GDP(%)

γ_i denotes coefficient of the secondary school teachers.

θ_i denotes coefficient of the inflation rate

λ_i denotes coefficient of the real GDP growth rate

t denotes time trend

μ_t denotes error ter

$\beta_1, \dots, \beta_7$ denotes coefficient of the short run dynamics of the model

θ denotes coefficient of short-run

ECT_{t-1} denotes error correction term lagged by one period.

3.3.5 Tests for the model instability

1) Chow test (1960) for the structural break point

Chow test is used for the one brake date in the time series analysis. There is two sub-periods for the only one structural break. Suppose that the date is known when the break happened, so the Chow break point test is used. In analyzing of the Chow break point test, there are two group of the model. The first group of the model considered that the parameters of the model are over the total time period, and there is no break point date which is null hypothesis. The second group of the model is that there is a structural break which is alternative hypothesis.

The first group is

$$y_t = \alpha + \alpha_1 x_t + \varepsilon_t \quad : \quad t=T \tag{3.28}$$

The second group where there is the structural break, there are two models

$$y_t = \delta + \delta_1 x_t + \varepsilon_{1t} \tag{3.29}$$

$$y_t = \delta + \delta_2 x_t + \varepsilon_{2t} \tag{3.30}$$

In the second group, the first model is before the break, and the second one is after the break. The parameters of the second group are the same, and

H0 : $\alpha_1 = \delta_1$; parameters are stability

H1 : $\alpha_1 \neq \delta_1$; parameters are not stability

The F-statistics is

$$F = \frac{RSS-(RSS_1+RSS_2)}{(RSS_1+RSS_2)} \times \frac{T-2k}{k} \tag{3.31}$$

If the F-statistic exceeds the critical F-statistics, the hull hypothesis is rejected, and the parameters are not stability. The calculated F-statistics does not exceed the critical value, the null hypothesis is accepted, and the parameters are stability.

2) Correlogram Q-statistics and Lagrange Multiplier test for serial correlation

One assumption of the linear regression model is

$$cov(\varepsilon_t, \varepsilon_s) = 0 \quad s \neq t \tag{3.32}$$

In this model, we assume that the error terms of the used variables are uncorrelated. For most of the macroeconomics variables, the values of the current period depend on the previous period, so the error term of the variables are mostly correlated. The error correlation of variables is called auto-correlation or serial correlation. The correlogram Q statistics and Lagrange Multiplier test are used whether the error terms are correlated or not. The Q statistics take the sample serial correlation of the residuals, and the LM test is the formal auto-correlation test which also considered the no serial correlation up to the lag structure (k). Suppose that the parameters of the serial correlation of the k-th order:

$$\delta_k = \frac{cov(y_t, y_{t-1})}{var(y_t)} \tag{3.33}$$

The correlation of the variable y is determined the value of the δ_k which are k periods of time and the hypothesis is

H0 : $\delta_k = 0$

H1 : $\delta_k \neq 0$

The test statistics is

$Z = \sqrt{T}\,\delta_k$ which is the normal distribution N(0,1)

The null hypothesis is there is no serial correlation, and the alternative is serial correlation. The p values of the Q statistics is insignificant and large values which mean that auto-correlation and the partial correlation at all lags are zero, so the null hypothesis is failed to reject that are conclude that there is no serial correlation.

The LM test consider the lag length of the variables, so the residual regression model is

$$e_t = \alpha X_t + \delta_1 e_{t-1} + \delta_2 e_{t-1} + \cdots + \delta_3 e_{t-1} + \mu_t \qquad (3.34)$$

$$H_0 = \delta_1 = 0 \wedge \delta_2 = 0 \wedge \ldots \wedge \delta_k = 0$$

$$H_0 \neq \delta_1 \neq 0 \wedge \delta_2 \neq 0 \wedge \ldots \wedge \delta_k \neq 0$$

The p value is also used whether there is serial correlation or not. If the p values are significant (0.0000) at the significant level, the null hypothesis of the LM test is accepted, and concludes that there is auto-correlation or serial correlation between the variables.