# CHAPTER 3

# METHODOLOGY

## 3.1 Data Collection

Due to data limitation, data used for this study is secondary data for the period of 1961 to 2010 as most of the variables analyzed were available in the International Rice Research Institute (IRRI), World Food Programme (WFP), Food and Agriculture Organization of the United Nations (FAO), United States Department of Agriculture (USDA), World Trade Organization (WTO), Asian Development Bank (ADB), ASEAN food security information system (AFSIS), CEIC database, Cambodian Development and Research Institute (CDRI) and government reports. Dependent variable is rice output in Cambodia. This variable will be used to achieve three significant purpose of the study.

## 3.2 Conceptual Framework

The study aims to estimate the rice production by employing three models. Rice production output models based on input components such as fertilizer use, machinery (tractor), harvested area, paddy yield, and irrigated agricultural area. Rice production forecasting model uses to estimate the rice production in the near future. Rice export contribution to GDP model can analyze the contribution of rice production and export to stimulate economic growth.

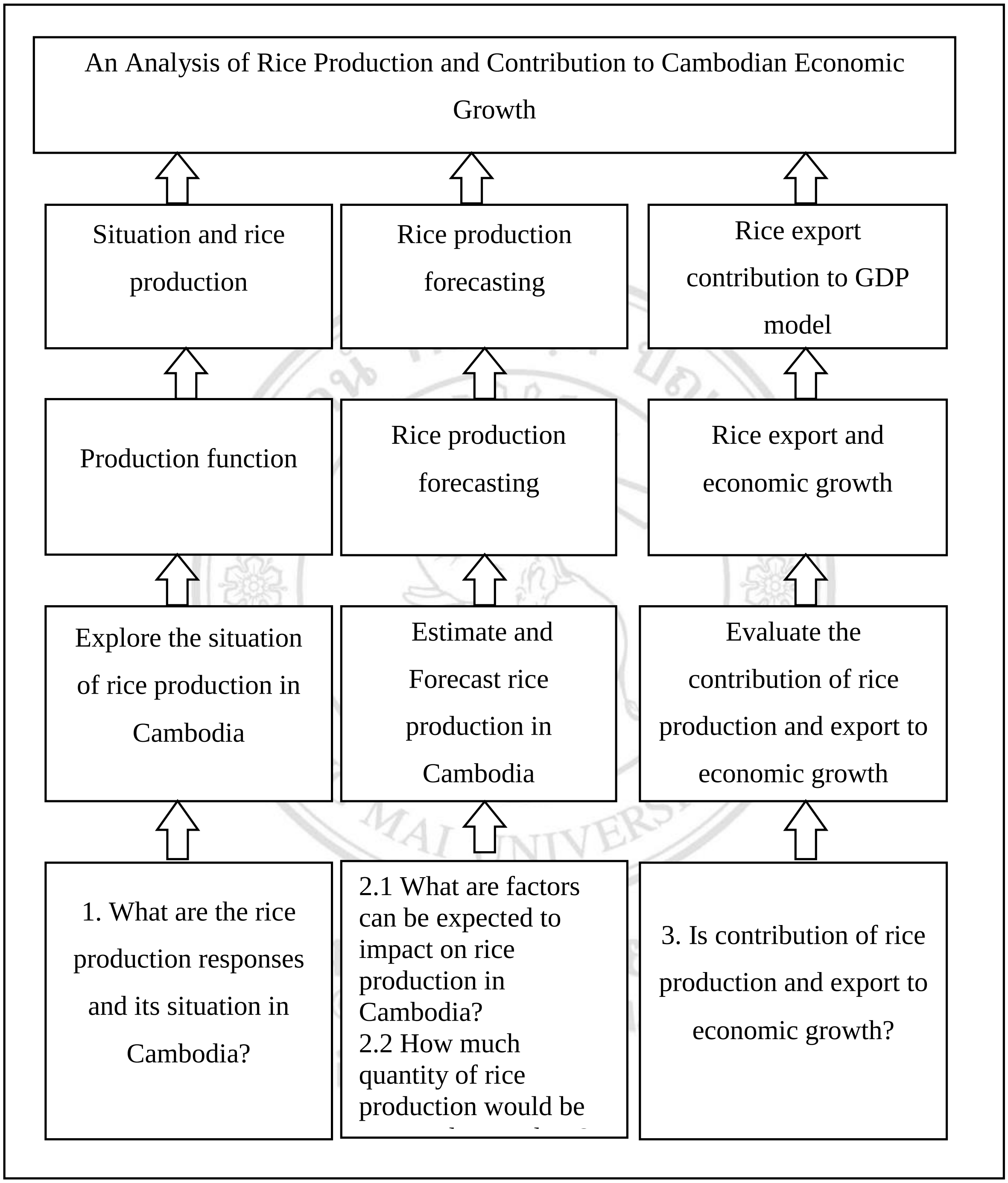


**Figure 3.1:** Conceptual Framework of Cambodian Rice Production

## 3.3 Research Methodologies

### 3.3.1 Porter's Diamond Model

Michael Porter is professor of the Harvard Business School, has introduced new theory to explain about national or industry competitive advantage; it is called diamond model. Porter focuses on competitiveness in which is a core driven force to success in international trade. He combines all possible factors that could influence one's industry or nation competitiveness (Smit, 2010; Charoen, 2012). There are four critical factors that are needed to consider to reveal competitiveness of the industry or nation. Factor conditions, demand conditions, firm strategy, structure and rivalry and related and supporting industries are four edges of diamond model. However, there are two other exogenous factors that indirectly influence on the model such as government and chance (Smit, 2010; Charoen, 2012).

Factor conditions is closely related to Cobb-Douglas production function in which land, labor and capital are the key inputs. However, Porter has divided these factors into two categories such as basic factors and advanced factors. Basic factors are what already exist or need little investment as unskilled labor, natural resources, weather condition and raw materials (Smith, 2010). For example, infrastructure or advanced information and communications technology (ICT). Demand conditions refers to home-country high quality services or products demand cause intra-industry competition to be tighter to serve the sophisticated requirement of customer. It may help industry to find the way to innovate its services or products. Productivity is needed to be high with a competitive price to meet the high standard. The high competition in one's nation or industry will later on transfer home-country demand into international demand. Firm strategy, structure and rivalry happen as a firm is trying to compete each other. National business environment where firm strategy and structure is built determine country's competitiveness. Rivalry drives firm to invest its budget in research and development in order to produce high quality products or service with a competitive price to satisfy its customer also maintain its operation in the business (Smit, 2010).

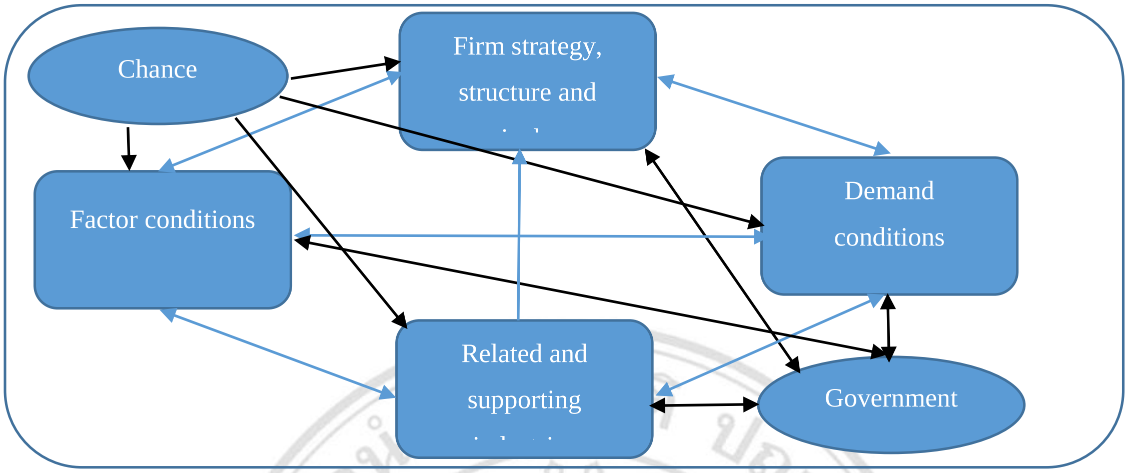


**Figure 3.2:** Port's Diamond model

Related and supporting industries are the last determinant of nation or industry competitive advantage (Smit, 2010). The strong network of specialized supply chain such as providers and institutions is the true competitive advantage since cluster of related and supporting industries would flourish innovation and learning environment (Smit, 2010). External factors as government and chance also influence country's competitiveness. Government could use its law or policies to influence supply conditions, demand conditions in the home market also competition between firms in the industry. Chance is the outside events that provide opportunities or threats to industry (Smit, 2010).

### 3.3.2 Rice production

In this study, a typical Cobb-Douglas functional form is used to estimate the rice production function in Cambodia. The utilized inputs include harvested area, fertilizer, irrigation, seed, and machinery. Output is rice production. The use of the Cobb-Douglas function offers two main advantages: Firstly, it allows outcomes to calculate returns to scale as constant, increasing or decreasing; and secondly, the estimated coefficient of an input from a linearized Cobb-Douglas function is the explanation of output with respect to the input quantity. This method is used in many empirical works. More specifically, the Cobb-Douglas production function is used to represent the rice production for Cambodian farmers can be written as:

$$Y = \alpha_0 HA^{\alpha_1} SE^{\alpha_2} MA^{\alpha_3} FE^{\alpha_4} IR^{\alpha_5} \tag{43}$$

Taking natural logarithm,

$$Ln\ Y_t = \alpha_0 + \alpha_1 lnHA_t + \alpha_2 lnSE_t + \alpha_3 lnMA_t + \alpha_4 FE_t + \alpha_5 IR_t + U_t$$
(44)

where,

| | |
|---|---|
| $LnY$ | is logarithm of quantity of rice production output in (ton) |
| $LnHA$ | is logarithm of harvested area in (ha) |
| $LnSE$ | is logarithm of seed use in (ton) |
| $LnMA$ | is logarithm of machinery use (tractors) in (tractor) |
| $LnFE$ | is logarithm of fertilizer use in (ton) |
| $LnIR$ | is logarithm of irrigated agricultural area (ha) |
| $\alpha_i$ | is parameter (i = 0, 1, 2,…,5) |

The expected signs are based on the previous studies. For instance, Yu and Fan (2009) indicated that the high price and output expansion in the Cambodian rice sector, by increasing their inputs such as fertilizers, irrigation, harvested area, seed use, and machinery use. Furthermore, According to Przsanna, Kumar and Singh (2009), rice production is based on input-use pattern in rice production like fertilizer, irrigation, seed use and harvested area to push quantity and quality of rice production. For these, all the input-use variables are expected to be positive signs.

**Table 3.1:** The expected sign of parameters of variables

| Parameters | Variables | Sign |
|---|---|---|
| $\alpha_1$ | $HA_t$ | + |
| $\alpha_2$ | $SE_t$ | + |
| $\alpha_3$ | $MA_t$ | + |
| $\alpha_4$ | $FE_t$ | + |
| $\alpha_5$ | $IR_t$ | + |

### 3.3.3 Co-integration and Vector Error Correction Model of Rice Production

Before we go through for testing co-integration, unit root test is previously employed to judge whether variables is stationary or non-stationary in which level. Augmented Dickey Fuller (ADF) test is used to test the stationarity of variables. For time series data, ADF test is a preferable test for unit root test. The ADF test can be conducted using the regression as below:

$$\Delta Y_t = \beta_1 + \beta_t + \alpha Y_{t-1} + \beta_i \sum_{i=1}^{m} \Delta Y_{t-1} + \varepsilon_t \qquad (45)$$

When added term ($\beta_i \sum_{i=1}^{m} \Delta Y_{t-1}$) in equation (45) then t-statistics value of α before $Y_{t-1}$ to be changed as well as all t-statistics value of them to be changed too. So ADF-Test corrects for higher order serial correlation by adding lagged differenced terms on the right-hand side. The hypothesis test for unit root in time series data by ADF-Test method as same as the DF-Test method and same conclusion about time series data is stationary or non-stationary (Chaiboonsri, 2008). Stationary or non-stationary can be concluded by defining null hypothesis for ADF as:

$H_0$: $\delta = 0$ (it has unit root)

$H_1$: $\delta < 0$ (it has no unit root)

In this study, Autoregressive Distributed Lag (ARDL) is used to find the appropriate lag latent among selected variables. ARDL was proposed by Pesaran (Pesaran and shin, 1995), is introduced to examine rice production and export to economic growth in Cambodia. Many current studies have demonstrated that the ARDL approach to co-integration is mostly used to estimate long-run and also short run relationship. Since all variables are confirmed to be stationary, ARDL is used to estimate the long-run and short run relationship. The ARDL model can be written as:

$$\Delta X_{t,j} = a_0 + \sum_{i=1}^{n_1} a_{1i,j}\,\Delta X_{t-i,j} + \sum_{i=0}^{n_2} a_{2i,j}\,\Delta HV_{t-i,j} + \sum_{i=0}^{n_3} a_{3i,j}\,\Delta SD_{t-i,j} + \sum_{i=0}^{n_4} a_{4i,j}\,\Delta MA_{t-i,j} + \sum_{i=0}^{n_5} a_{5i,j}\,\Delta FE_{t-i,j} + \sum_{i=0}^{n_6} a_{6i,j}\,\Delta IR_{t-i,j} + \sum_{i=0}^{n_7} a_{7i,j}\,\Delta Y_{t-i,j} + a_6 HVX_{t-1,j} + a_7 SDX_{t-1,j} + a_7 MA_{t-1,j} + a_8 FE_{t-1,j} + a_9 IR_{t-1,j} + a_{10} Y_{t-1,j} + v_t \qquad (46)$$

where,

Time t = 1961-2010

i = number of lag

j = rice as a commodity

In order to analyze the outcomes, F-test is used to test the existence of long-run relationship for all variables between dependent variable and independent variables. To test whether there is any existence of long-run relationship, two step tests need to be calculated. The first step consist of estimating the equation in level form I(0), while the second step consists of testing the stationarity of the residual I(1). The hypothesis can be written as:

$H_0$: has not the long run relationship or co-integration among the variables

$H_1$ : has the long-run relationship or co-integration among the variables

In order to determine the results whether there is reject or accept null hypothesis, there conclusion can be derived from F-test results. If the calculated F-statistics is greater than the upper bound critical value, mean null hypothesis is rejected. There is a long-run relationship among the variables. If the calculated F-statistics is less than the lower bound critical value, null hypothesis is accepted. The result can be concluded that there is not long-run relationship among the variables. However, the result is inclusive when the value of computed F-test falls between the lower and upper bound critical values.

In case the variables accept the existence of co-integration, then the conventional Vector Error Correction Model (VECM) is applied using OLS, in order to confirm the short run relationship effect speed of adjustment between dependent variable and independent variables within bounds testing approach. Therefore the VECM method can be written as:

$$\Delta X_{t,j} = \beta_0 + \sum_{i=1}^{k_1} \beta_{1i,j}\, \Delta X_{t-i,j} + \sum_{i=0}^{k_2} \beta_{2i,j}\, \Delta HV_{t-i,j} + \sum_{i=0}^{k_3} \beta_{3i,j}\, \Delta SD_{t-i,j} \sum_{i=0}^{k_4} \beta_{4i,j}\, \Delta MA_{t-i,j} + \sum_{i=0}^{k_5} \beta_{5i,j}\, \Delta FE_{t-i,j} + \sum_{i=0}^{k_6} \beta_{6i,j}\, \Delta IR_{t-i,j} + \sum_{i=0}^{k_7} \beta_{7i,j}\, \Delta Y_{t-i,j} + \lambda EC_{t-1,j} + \mu_t \tag{47}$$

The speed of adjustment is determined by estimating the value of λ. It is known as coefficient of adjustment and usually statistically negative sign. It can explain the economic adjustment to be recovered in the short run period if any shock would be occurred. It also provides vital information about economic recovery to long-run as equilibrium and the steady state.

### 3.3.4 Rice production forecasting model

Rice production forecasting employs Autoregressive Integrated Moving Average (ARIMA) (p, d, q) method to forecast annually rice production in Cambodia during the time periods 1960-2013. The Box-Jenkins (1976) can be used for a non-seasonal series in order to forecast rice production based on time series in Cambodia. According to (Box and Jenkins, 1970), the Box-Jenkins approaches consist of three steps, identification, estimation, and diagnostic checking. These steps can be explained in detail as the following:

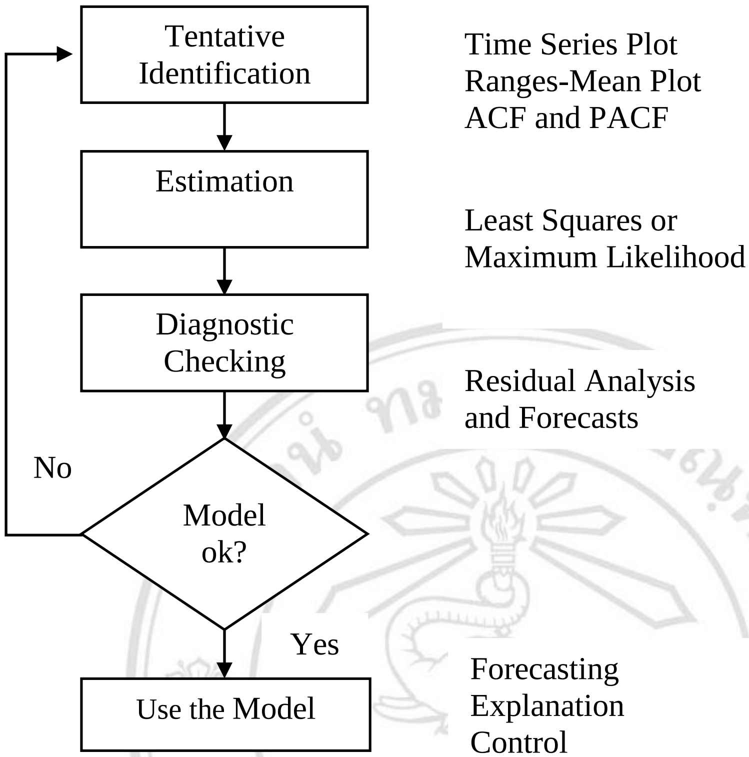


**Figure 3.3:** Box-Jenkins Methodology

**1) Identification**

Identification of the model for ARIMA (p,d,q) is based on the concept of using the techniques for determining the value of p, d and q. These values are defined by using autocorrelation function (ACF) and partial autocorrelation function (PACF) (Biswas and Bhattacharyya, 2013). In order to determine whether the series is stationary or not we could be considered the graph of ACF. If a graph of ACF cuts of falling quickly, then the time series value could be considered stationary. Here p represents the order of the autoregressive part, d represents the amount average of difference and q represents the order of the moving average part. If the original series is stationary, d=0 and the ARIMA models become to the ARMA models (Badmus and Ariyo, 2011).

The difference linear operator ($\Delta$), can be written as:

$$\Delta Y_t = Y_t - Y_{t-1} = Y_t - BY_t = (1 - B)Y_t \tag{48}$$

The stationary series:

$$W_t = \Delta^d Y_t = (1-B)^d Y_t = \mu + \theta_q(B)\varepsilon_t \tag{49}$$

$$\text{Or } \phi_p(B)W_t = \mu + \theta_q(B)\varepsilon_t \tag{50}$$

**2) Estimation**

Estimation of the appropriate values of p, d and q in the model and their statistical significance can be determined by t-distribution. A model has minimum values of Root Mean squared error (RMSE), Minimum Absolute Percentage Error (MAPE), Akaike's information criterion (AIC), Schwarz's Bayesian Information criterion (BIC), Mean Absolute error (MAE), Mean squared error (MSE) Q-statistics and high R-square, can be considered as the best model forecasting (Biswas and Bhattacharyya, 2013). The test statistics Q is given in equation:

$$Q_M = n(n+2)\sum_{k-1}^{n} \frac{r_k^2(e)}{n-k} \chi^2_{m-r} \tag{51}$$

where, $r_k(e)$ = the residual autocorrelation at lag K.

n = the number of residuals

m= the number of time lags includes in the test.

If the p-value associated with the Q statistics is small (p-value<α), the model is considered adequate. The analysts should consider a new or modified model and continue the analysis until a satisfactory model has been determined (Badmus and Ariyo, 2011).

**3) Diagnostic checking**

This step is to check whether the selected model appropriates the data based on ARIMA (p,d,q) model in order to find out if they are white noise. If the residuals are white noise the ACF of residuals and the Ljung and Box (1978) statistic will be applied. These may be determined by Ljung-Box statistic under null hypothesis that autocorrelation co-efficient is equivalent to zero. The best appropriate model for selecting when two or more competing models passing the diagnostic checking can be considered as the following criteria multiple $R^2$, Adjusted $R^2$, Root mean squared error (RMSE), Akaike's Information Criterion (AIC), Schwarz's Bayesian Information (BIC), Mean absolute error (MAE) and Mean absolute proportion percent error (MAPPE) (Biswas and Bhattacharyya, 2013).

### 3.3.5 Rice export contributions to GDP model

In order to achieve research objectives, this work obtained inspiration from the model used by Muhammad ZahirFaridi (2010). He establishes an econometric model based on a generalized Cobb Douglas production function.

$$Y_t = f(L_t, K_t) \tag{52}$$

The model included non-agricultural export as one of the independent variable determined using the principle component approach. Even though this study would use his model as a basis for the specification of our own model, we would avoid from being too generic i.e. looking at the entire contribution of agricultural exports to economic growth in Cambodia. This is due to the broadness of content which make it difficult for policy manipulations.

The study develops the same theoretical model based on the contribution of Agricultural export to economic growth in Cambodia with the case in point being rice export.

$$Y_t = \mathrm{f}\,(X_t^{EXP}, X_t^{LAB}, X_t^{GFC}) \tag{53}$$

The Cobb-Douglas form of neo-classical production function can be written as;

$$Y_t = \mathrm{A}_t(EXP_t^{\alpha} LAB_t^{\beta} GFC_t^{\lambda}) \tag{54}$$

This is essentially based on the production function framework, assuming a generalized Cobb Douglas production function and extending this Neo-classical growth model to include some selected agricultural exports indicators as additional inputs of the production function, alongside gross domestic fixed capital, and labor force as control variable written as;

$$RGDP_t = f(EXPX_t, LABX_t, GFCX_t) \tag{55}$$

where;

$RGDP_t$ is the annual real Gross Domestic Product

$EXP_t$ is rice export

$LAB_t$ is labor force for agricultural sector

$GFC_t$ gross domestic fixed capital

t is time trend.

Real Gross Domestic Product (RGDP) is our dependent variable due to we are looking at the relationship between the real GDP and rice export in Cambodia. It is defined as the sum of gross value added by all resident producers in the economy plus any product taxes and minus any subsidies not included in the value of the products. It is calculated without making deductions for depreciation of fabricated assets or for degradation of natural resources. These data are in current U.S. dollars (World Development Indicators, 2013).

In this study, total labor force variable capture the effect of labor force on economic growth since the development on the agricultural sector improves the productivity of labor. Labor force consist of people aged 15 and older, who meet the International Labor Organization definition of the economically active population. It includes both the employed and the unemployed. While national practices vary in the treatment of such group as the armed forces and seasonal or part time workers, in general the labor force incudes the armed forces, unemployed, and first time job seekers, but excludes home-makers and other unpaid caregivers and worker in the informal sector (World development indicator, World Bank 2013).

Gross Domestic Fixed Capital Formation (formerly gross domestic fixed investment) includes land improvements (fences, ditches, drains, and so on); plant, machinery, and equipment purchases; and the construction of road, railways, and the like, including schools, offices, hospitals, private residential dwellings, commercial and industrial buildings. Dara are in the current U.S. dollars. (World Development Indicators, World Bank 2013).

The rice export is measured in unit known as tons (FAOSTAT, 2013). They have been chosen because of their largest subsector contributions to Cambodian economic growth.

Finally, we estimate the following equation form our generalized model in equation (55), to empirically examine the effect of rice export on economic growth in Cambodia from 1996 to 2011.

By taking the natural logs ($ln$) on both sides of the equation (54) in order to rule out the differences in the units of measurements for variables;

$$LRGDP_t = lnA_t + \alpha lnEXP_t + ln\beta LAB_t + \lambda lnGFC_t + \varepsilon_t \tag{56}$$

where α, β and λ are parameters to be estimated.

To estimate the effect of rice export on economic growth in Cambodia, the study specifies the following model which is just a slight modification of equation (56).

$$LRGDP_t = \beta_0 + \beta_1 LEXPt + \beta_2 LAB_t + \beta_3 GFC_t + \varepsilon_t \tag{57}$$

Where; L is the natural logarithm of the variables, e.g. $LRGDP_t$= natural logarithm of real gross domestic product, $\varepsilon_t$is the stochastic error term $\beta_0$ is the constant term while $\beta_1$, $\beta_1$ and $\beta_3$ are parameters of the independent variables. $\beta_1$>0.

The expected signs are based on the previous studies. For instance, Gilbert, Linyong and Divine (2013) indicated that the agricultural exports have a positive and significant relationship with economic growth. Moreover, according to Anthony (2010) revealed that agricultural variables have positive and significant impact on economic growth and their contribution to export growth has been encouraging. For these, all the input-use variables are expected to be positive signs.

**Table 3.2:** The expected sign of parameters of variables

| Parameters | Variables | Sign |
|---|---|---|
| $\beta_1$ | $EXP_t$ | + |
| $\beta_2$ | $LAB_t$ | + |
| $\beta_3$ | $GFC_t$ | + |

### 3.3.6 Co-integration and Vector Error Correction Model of Rice Export on Economic growth

Before we go through for testing co-integration, unit root test is previously employed to judge whether variables are stationary or non-stationary in which level. Augmented Dickey Fuller (ADF) test is used to test the stationarity of variables. For time series data, ADF test is a preferable test for unit root test. The ADF test can be conducted using the regression as below:

$$\Delta Y_t = \beta_1 + \beta_t + \alpha Y_{t-1} + \beta_i \sum_{i=1}^{m} \Delta Y_{t-1} + \varepsilon_t \tag{58}$$

When added term ($\beta_i \sum_{i=1}^{m} \Delta Y_{t-1}$) in equation (58) then t-statistics value of α before $Y_{t-1}$ to be changed as well as all t-statistics value of them to be changed too. So ADF-Test corrects for higher order serial correlation by adding lagged differenced terms on the right-hand side. The hypothesis test for unit root in time series data by ADF-Test method as same as the DF-Test method and same conclusion about time series data is stationary or non-stationary (Chaiboonsri, 2008). Stationary or non-stationary can be concluded by defining null hypothesis for ADF as:

$H_0: \delta = 0$ (it has unit root)

$H_1: \delta < 0$ (it has no unit root)

In this study, Autoregressive Distributed Lag (ARDL) was proposed by Pesaran (Pesaran and shin, 2001), is introduced to examine rice production and export to economic growth in Cambodia. Many current studies have demonstrated that the ARDL approach to co-integration is mostly used to estimate long-run and also short run relationship. Since all variables are confirmed to be stationary, ARDL is used to estimate the long-run and short run relationship. The ARDL model can be written as:

$$\Delta X_{t,j} = a_0 + \sum_{i=1}^{n_1} a_{1i,j} \Delta X_{t-i,j} + \sum_{i=0}^{n_2} a_{2i,j} \Delta EXP_{t-i,j} + \sum_{i=0}^{n_3} a_{3i,j} \Delta LAB_{t-i,j} + \sum_{i=0}^{n_4} a_{4i,j} \Delta GFC_{t-i,j} + a_5 EXP_{t-1,j} + a_6 LAB_{t-1,j} + a_7 GFC_{t-1,j} + v_t \tag{59}$$

where,

Time t = 1996-2011

i = number of lag

j = rice as a commodity

In order to analyze the outcomes, F-test is used to test the existence of long-run relationship for all variables between dependent variable and independent variables. To test whether there is any existence of long-run relationship, two step tests need to be calculated. The first step consist of estimating the equation in level form I(0), while the second step consists of testing the stationarity of the residual I(1). The hypothesis can be written as:

$H_0$: has not the long run relationship or co-integration among the variables

$H_1$ : has the long-run relationship or co-integration among the variables

In order to determine the results whether there is reject or accept null hypothesis, there conclusion can be derived from F-test results. If the calculated F-statistics is greater than the upper bound critical value, mean null hypothesis is rejected. There is a long-run relationship among the variables. If the calculated F-statistics is less than the lower bound critical value, null hypothesis is accepted. The result can be concluded that there is not long-run relationship among the variables. However, the result is inclusive when the value of computed F-test falls between the lower and upper bound critical values.



Since the variables accept the existence of co-integration, then the conventional Vector Error Correction Model (VECM) is applied using OLS, in order to confirm the short run relationship effect speed of adjustment between dependent variable and independent variables within bounds testing approach. Therefore the VECM method can be written as:

$$\Delta X_{t,j} = \beta_0 + \sum_{i=1}^{k_1} \beta_{1i,j}\, \Delta X_{t-i,j} + \sum_{i=0}^{k_2} \beta_{2i,j}\, \Delta EXP_{t-i,j} + \sum_{i=0}^{k_3} \beta_{3i,j}\, \Delta LAB_{t-i,j} \sum_{i=0}^{k_4} \beta_{4i,j}\, \Delta GFC_{t-i,j} + \lambda EC_{t-1,j} + \mu_t \quad (60)$$

The speed of adjustment is determined by estimating the value of λ. It is known as coefficient of adjustment and usually statistically negative sign. It can explain the economic adjustment to be recovered in the short run period if any shock would be occurred. It also provides vital information about economic recovery to long-run as equilibrium and the steady state.