CHAPTER FOUR: METHODOLOGY

4.0 INTRODUCTION

The methodology used is adapted from **Baharumshah (2003)**. The empirical process comprises three parts: firstly, testing for unit root in each series; secondly, testing for the number of co-integrating vectors in the system; and thirdly, estimating and testing for causality in the framework of a multivariate Vector Error-Correction Model (VECM). If the variables for a particular country are found to be stationary in their level representation, then the Vector Auto-regression (VAR) model is appropriate in detecting the direction of causality between saving and its determinant.

4.1 DATA

The sample covered the period from the first quarter of 1970 through the fourth quarter of 2000 for the case of Malaysia. The data is based on the following series: gross national savings, gross national product, interest rate, dependency ratio and current account. The data sources derived from various issues of International Monetary Fund (IMF), World Saving Database[2] from World Bank and Asian Development Bank (ADB). All variables are in natural logarithmic (except for the interest rates) so the first difference represents the growth rates. The transformation of time series variables by taking natural logarithms has the advantage of stabilizing the variances of those series.

4.2 ECONOMETRIC MODEL

Based on the above discussion, the saving function for the ensuing empirical analysis can be specified as follows:

$$sav_t = \alpha_0 + \beta_1 gnp_t + \beta_2 i_t + \beta_3 dr_t + \beta_4 (ca_t)^2 + \varepsilon_t \qquad (18)$$

Where *sav* here is represents gross national savings and the independent variables are (with the signs expected for the regression coefficients are given in parentheses):

gnp (+ or -) = gross national product

i (+ or -) = interest rate

dr (+ or -) = dependency ratio

ca (-) = current account

The disturbance term ε is to represent the influence of those unobservable variables that are excluded in the model and is assumed to be N (0, σ^2). We used squared current account because these country affected by crisis recorded current account deficits in the 1990s.

4.3 ECONOMETRIC TESTS

In order to test the causal relationship between savings and its determinant, three kinds of tests are needed: the unit root tests, the co-integration test and the Granger-causality test.

4.3.1 UNIT ROOT TEST

To avoid spurious regression, the unit root test is performed on those times series variable to investigate the stationary properties. Testing for unit root is the first step in time series model building. We begin the estimation process by testing the time series properties of the data using **Augmented Dickey and Fuller (1981) and Phillips and Perron (1988)** unit root tests to test for the order of integration.

4.3.1.1 AUGMENTED DICKEY AND FULLER

Augmented Dickey and Fuller (ADF) test is used to determine the degree of integration of the relevant variables with assumption that correlated errors have constant variance. The ADF test consists of running a regression of the first difference of the series against the series at lag one, lagged difference term an optional, a drift and a time trend. With k lagged difference term, the regression is:

$$\Delta y_t = \mu + \beta_t + \delta y_{t-1} + \sum_{k=1}^{m} \phi_k \Delta y_{t-k} + \varepsilon_t \qquad (19)$$

Where Ä represents first difference operator and the error term ε_t is assumed to follow a white noise process. y_{t-1} is the one-period lagged value of the variable y_t was included to ensure that the error term becomes white noise. Then, μ represent the drift and β_t which implies time trend for equation (19). Let y_t refer to the variables of the series and $t = 1,2,...,n$, which n refer to number of observations in the sample period.

Besides that, lag length m is determined by choosing the value that minimizes the Akaike Information Criterion (AIC) or the Schwarz Information Criterion (SIC). **Hall (1994)** stated that, the asymptotic null distribution of the Dicker-Fuller statistic with augmentation lag order, which selected by SIC is the same as if the true order were know. As a result, SIC provides a useful guide to augmentation lag order in Dicker-Fuller regressions even it based on the parsimony principle.

The test for a unit root is a test on the coefficient of y_{t-1} in the regression. If the coefficient is significantly different from zero then the hypothesis that y_t contains a unit root i.e., H_0:δ = 0 is rejected, y_t is consider stationary and follow an I(0) process. Otherwise, if null hypothesis is not rejected, y_t contains at least one unit root and is non-stationary. Therefore, the high order of integration (1st difference) should be test for presence of unit root. If the H_0 of a unit root is rejected for the high order of integration, the variable is said follow an I(1) process.

Critical value for the ADF test is based on t-statistic, which calculated, by regression. However, the t-statistic cannot be referred to the critical value in the standard t table. Because, under the null hypothesis the left hand side variable is non-stationary. As a result, the critical value applied in this test is based on critical value that tabulated by **MacKinnon (1991)**.

4.3.1.2 PHILLIPS AND PERSON

The Phillips and Person (PP) t-test for unit root is also employed in this analysis. This PP test allows for weakly correlated and heterogeneously distributed errors. The test equation includes a constant and a deterministic time trend, the regression is:

$$\Delta y_t = \mu + \beta_t + \phi y_{t-1} + \varepsilon_t \qquad (20)$$

Where μ over here refer to constant term and others else symbol were same as discussed in equation (19). Non-parametric corrections on the t-statistic for testing null hypothesis $\phi = 0$ are used. If the null hypothesis is not rejected, y_t contains at least one unit root and is non-stationary. However, y_t is stationary if the null hypothesis is rejected. The critical values, which referred to, are also tabulated by **MacKinnon (1991)**. The rejection rule is also same as discussed in ADF test.

4.3.2 CO-INTEGRATION TEST

After testing unit roots for each variable, a co-integration test should be performing to ensure that the multiple regression models are not spurious before turning to the test of causality. The interpretation for co-integration is that 2 or more series are linked to form an equilibrium relationship over the long run. Even through these series themselves are no stationary, they will move together over time and their difference will be stationary.

Generally, a set of n difference-stationary variables are said to be co-integrated if at least one linear combination between them exists such that this linear combination is of a lower order of integration than the order for the series themselves. For instance, both the time series variables x_t and y_t are I(1), a linear combination exists as following:

$$y_t = \beta' x_t + u_t \tag{21}$$

$$y_t - \beta' x_t = u_t ; u_t \sim I(0) \tag{22}$$

Where x represents the vector of fundamental determinants, β is the co-integrating vector of coefficients, and u is the stationary residual. If above (21) and (22) equation are satisfies, x_t and y_t are said to be co-integrated. Since $u_t \sim I(0)$ implies constant variance, therefore, the regression is not spurious because x_t and y_t are co-integrated. It is because, the unknown trends in x_t and y_t are cancel out i.e., x_t and y_t are sharing a common trend.

4.3.2.1 JOHANSEN METHODOLOGY

The test of co-integration between saving and its determinant are based on a VAR approach introduced by **Johanson (1988)**. Let $y_t = (y_{1t} + y_{2t} + \cdots + y_{5t})$, a VAR model for y_t is:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \ldots + \alpha_5 y_{t-5} + \varepsilon_t \qquad (23)$$

This can be rewritten as an error correction model, which takes the following form:

$$\Delta y_t = \Pi y_{t-1} + \sum_{i=1}^{q-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \qquad (24)$$

$$\text{where } \Pi = \begin{bmatrix} \Pi_{11} & \Pi_{12} & \cdots & \Pi_{1r} \\ \Pi_{21} & \Pi_{22} & \cdots & \Pi_{2r} \\ \vdots & \vdots & \ddots & \vdots \\ \Pi_{51} & \Pi_{52} & \cdots & \Pi_{5r} \end{bmatrix} \qquad \text{or} \qquad \underset{5\times 5}{\Pi} = \underset{5\times r}{\alpha}\, \underset{r\times 5}{\beta'}$$

$$\Gamma_i = \begin{bmatrix} \Gamma_{11.j} & \Gamma_{12.j} & \cdots & \Gamma_{15.j} \\ \Gamma_{21.j} & \Gamma_{22.j} & \cdots & \Gamma_{23.j} \\ \vdots & \vdots & \ddots & \vdots \\ \Gamma_{51.j} & \Gamma_{52.j} & \cdots & \Gamma_{55.j} \end{bmatrix}$$

$$\varepsilon_t = [\varepsilon_{1t}, \varepsilon_{2t}, \ldots \varepsilon_{5t}]$$

Based on Granger's representation theorem, if Π has reduced rank $r < 5$, it can be expressed then as $\alpha\beta'$ $(\Pi = \alpha\beta')$, where α and β are $5 \times r$ matrices and r is the number of co-integrating relations. Then, each column of β yields an estimate of the co-integrating vector. Meanwhile, elements of α are representing the adjustment parameters in the error correction model. However, if $r = 0$ and $\Pi = 0$ (zero rank) then the model is:

$$\Delta y_t = \sum_{i=1}^{q-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \qquad (25)$$

Equation (25) is an example of VAR model. This implies that, no co-integration is found among the elements of y_t. It also indicates that there is no long run relationship between these variables. Furthermore, if $r = 5$ and Π has full rank, which would imply that the variables are follow by a stationary process. The model had become equation (24), which can be reduced to equation (23), which refers to a VAR model of y_t in levels form. This is a trivial case of co-integration, which implies that the co-integration is against with the long-term trend. Neither Π is full rank nor zero rank, will it be appropriate to estimate the model respectively in either levels or first differences. Consequently, if $0 < r < 5$ (reduced rank), the model is:

$$\Delta y_t = \alpha\beta' y_{t-1} + \sum_{i=1}^{q-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \qquad (26)$$

Equation (26) implies that y_t is co-integrated with r long-term equilibrium relationship. The number of lags used in the VAR is chosen based on the evidence provided by Akaike's Information Criterion (AIC). In this case, we use an error correction model to perform a causality test which running from dependent variable to independent variables since co-integration implies the existence of an error correction model.

4.3.2.2 TRACE STATISTIC TEST

Besides that, for testing co-integration, we also used Likelihood Ratio (LR) test procedure established by **Johansen and Juselius (1990)** that based on trace statistics to testing the null hypothesis at least r co-integrating relations exist against an alternative hypothesis that states that more than r co-integrating relations exist. The trace test is given in equation below:

$$Q_r = -n \sum_{j=r+1}^{m} \log(1 - \lambda_i) \qquad (27)$$

Where λ is the eigen-value, m here are refers to number of both dependent and independent variables and n denotes to number of observations. The critical values for this test are adopted from **Qsterwald-Lenum (1992)**. If Q_r is greater than critical value, the null hypothesis $r = 0$ is rejected and showed that at least one co-integrating relation exists. Then, we need to proceed to test the null hypothesis $r = 1$ against the alternative hypothesis $r > 1$. In this case, if the null hypothesis is not rejected, we can conclude that $r = 1$ which means that there is one co-integrating relation. Otherwise, if null hypothesis is rejected, the high order of r should be testing in null hypothesis, i.e., H_0: $r = 2,..m-1, m$. Once r is determined, we can know the dimension of β given by $m \times r$ matrix of eigenvectors corresponding to the largest r eigen-value. With this β estimation, equation (27) can be estimated based on Ordinary Least Square (OLS).

4.3.3 GRANGER-CAUSALITY TEST

After that, if the result showed evidence of co-integration between these variables, the causality test should be performed using the Vector Error-Correction Model (VECM). The VECM incorporates both short run dis-equilibrium and long run equilibrium dynamics of these variables. Based on previous studies, as the co-integration test showed evidence of a long run relationship for these variables, a VECM based causality tests are appropriate **(A. Z. Baharumshah, 2003)**. The following equations were estimated to establish causality relations based on **Engle and Granger (1987)** methodology:

$$y_t - \hat{\beta}_0 - \hat{\beta}_1 z_t = \hat{\varepsilon}_t \qquad (28)$$

$$\Delta y_t = \delta_1 + \alpha_y \hat{\varepsilon}_{t-1} + \sum_{i=1}^{p} \theta_{11,i} \Delta y_{t-i} + \sum_{i=1}^{p} \beta_{12,i} \Delta z_{t-i} + \mu_{yt} \qquad (29)$$

$$\Delta z_t = \delta_2 + \alpha_z \hat{\varepsilon}_{t-1} + \sum_{i=1}^{p} \theta_{21,i} \Delta y_{t-i} + \sum_{i=1}^{p} \beta_{22,i} \Delta z_{t-i} + \mu_{zt} \qquad (30)$$

The α coefficient vector reveals the speed of adjustment to the error correction term, which measures the deviation from the empirical long run relationship. Then, μ_{yt} and μ_{zt} are disturbances which are uncorrelated. In the case of two variables, said z and y, the Granger causality approach is measures precedence and information provided by z in explaining current value of y. This means that, y is said to be granger caused by z if z helps in the prediction of y or equivalently lagged values of z are statistically significant and vice-versa.

Then, let say if Δy_t not granger cause Δz_t, the null hypothesis i.e., H₀: $\alpha_z = 0$ and $\theta_{21,i} = 0 \;\; \forall j$ should be rejected. Otherwise, the reverse is true if $\alpha_z \neq 0$ and $\theta_{21,i} \neq 0 \;\; \forall j$, which conclude that Δy_t was granger caused by Δz_t. On the other hand, the null hypothesis i.e., H₀: $\alpha_y = 0$ and $\beta_{12,i} = 0 \;\; \forall j$ should be rejected if Δz_t does not granger caused Δy_t. Similarly, the reverse is true, Δz_t was granger caused by Δy_t if $\alpha_y \neq 0$ and $\beta_{12,i} \neq 0 \;\; \forall j$. The significance of those parameters can be tested using the F-statistics based on Wald test to investigate any short-run causality between these variables.

Furthermore, the error correction term that measures the deviations of the series from the long run equilibrium relation can be tested using a significant F-statistic. The VECM exists when y_t and z_t was co-integrated which implies that at least one of α_y and α_z must be significant. For instance, if y_t does not granger cause z_t in the long run, the null hypothesis i.e., $H_0: \alpha_z = 0$ should be rejected. In contrast, the reverse is true if $\alpha_z \neq 0$, which conclude that Δy_t was granger caused by Δz_t in the long run.

As a result, the causal effect of one variable on another can be appears in two ways, neither through the effect of lagged changes in the independent variable that implies the short-run causality nor through the error-correction term, which refers to long-run causality. Thus, the VECM representation allows us to differentiate between the short and long run dynamic relationships.

4.4 CONCLUSION

To end up, the first step is to investigate the order of integration by using two popular unit root tests such as **Augmented Dicker and Fuller (1981)** and **Phillips and Person (1988)** statistics test. Furthermore, the second step is to establish the co-integration relationship among the five series. For this purpose, we employed **Johanson (1988)** methodology based on a VAR model to detect the co-integrating relation. Lastly, when co-integration is established, an error-correction model is estimated in order to observe the causal linkages among the variables. The investigation of causal relationship is based on **Engle and Granger (1987)** methodology through the VECM approach.