CHAPTER 4

DATA AND METHODOLOGY

4.1 Introduction

There are two sections in this chapter. The first section describes the data and defines the variables used for this analysis. The second section explains the testing procedure whereby the employed model is used to examine the relationship between tourism and economic growth in Malaysia. Brief explanations of the cointegration and causality tests are also presented in this section.

The objective of this study is to assess whether growth in international tourism in Malaysia has significantly contributed to the country's long-term economic development. This is accomplished by establishing whether a long-run equilibrium relationship is present between tourism and economic growth in Malaysia, using the Johansen Cointegration Test (Balaguer and Cantavella-Jorda, 2002). If the long-run relationship is determined, the causal relation among the same variables is examined using the Granger Causality Test in a cointegrating framework, to examine whether tourism-led growth hypothesis is valid for Malaysia.

4.2 Data Description

Data used in this study are quarterly time series data on income, tourism receipts and the real effective exchange rate of Malaysia. The time period covered for this study is from first quarter of 1988 to fourth quarter of 2002.

The income variable is represented by the real gross domestic product (GDP) which is measured in millions of Malaysian ringgit. Data are drawn from the Department of Statistics Malaysia (DOSM) and the International Financial Statistics (IFS), published by the International Monetary Fund (IMF). DOSM publishes quarterly real GDP data from 1991 onwards. Prior to 1991, only annual data are available. Accordingly, quarterly real GDP data series prior to 1991 are extrapolated from annual data using the method described by Tseng and Corker (1993). Based on this method, the fourth quarter GDP is assumed to be 25% of annual GDP for that year. By using the quarterly Malaysian industrial production index (published by DOSM), the remaining three quarters of GDP can be obtained as follows:

(a) GDP of first quarter for year t is computed as follows:

$$GDPQ_{1t} = \left[\frac{IPIQ_{1t}}{IPIQ_{4(t-1)}}\right] * GDPQ_{4(t-1)}$$

(b) GDP of second and third quarters for year t are calculated as follows:

$$GDPQ_{2t} = \left[\frac{IPIQ_{2t}}{IPIQ_{1t}}\right] * GDPQ_{1t}$$

$$GDPQ_{3t} = \left[\frac{IPIQ_{3t}}{IPIQ_{2t}}\right] * GDPQ_{2t}$$

where subscripts denote quarters of year t. Data from DOSM, which are expressed on 1987 base year, is then converted to the base year 1990, in line with the IMF data.

Tourism receipts (TOUR) variable is defined as Malaysia's international tourism earnings in real terms, which is measured in millions of Malaysia ringgit. Annual tourism receipt series are drawn from the Ministry of Tourism (formerly known as the Ministry of Culture, Arts and Tourism (MOCAT)), while the quarterly data are based on the actual quarterly total travel receipts (credit) from the balance of payment of Malaysia, excluding expenditures by transit passengers and excursionists to Peninsular Malaysia, Sarawak and Labuan, which are obtained from DOSM.

The real effective exchange rate (REER) is a weighted index that combines the exchange rates between a currency in particular and the currencies of industrialized countries (major trading partner and/or competitor countries). It is adjusted for relative movements in labour unit costs and express on 1990 year base. Defined as units of domestic currency per unit of foreign currency, an appreciation of REER is reflected by a decrease of the index while a depreciation of REER is reflected by an increase of the index. Data on the Malaysian REER are drawn from the IFS, published by IMF.

4.3 Model Specification

A few widely used approaches are used to explain the output growth over long periods. One of the approaches is growth theory, which models the interactions among factor supplies, productivity growth, savings and investment in the process of growth. Another approach is growth accounting, which attempts to quantify contribution of different determinants of output growth. In this case, most empirical studies have focused on the export-led growth hypothesis for both developing and developed countries. Given the fact that tourism income from foreign tourists is an export for the host country, the literature of the export-led growth hypothesis and recent theoretical models which consider non-traded goods such as tourism would justify the inclusion of tourism in a growth model in order to analyze the role of tourism for the economic growth. As in the export-led growth hypothesis, the model includes the most fundamental variables, namely, gross domestic product and tourism revenue. In view of the fact that Malaysia is a small open economy, foreign exchange rate is another contributory factor to the economic growth in the long run. Accordingly, the real effective exchange rate is also included in the model.

This study employs the model developed by Balaguer and Cantavella-Jorda (2002) and in econometric terms, the equation is expressed as follows:

$$LY_t = \alpha_0 + \alpha_1 LTOUR_t + \alpha_2 LREER_t + u_t \tag{4.1}$$

where Y = real gross domestic product

$TOUR$ = international tourism earnings in real terms

$REER$ = real effective exchange rate (index)

u = the error term

t = 1988Q1....2002Q4

It has been suggested that for long-run analysis between growth and tourism, the number of variables to be included in the model should be kept to a minimum without excluding any relevant and important variables. The rationale is that the more variables are included in the model, the higher the likelihood to obtain more than one relationship in the long run. From the economic point of view, this situation may appear somewhat confusing (Balaguer and Cantavella-Jorda, 2002).

A priori, tourism driven growth is positively related to both tourism revenues and real effective exchange rate. As such, αs are the parameters of the model to be estimated. All variables in equation 4.1 are expressed in natural logarithms so that the estimated elasticities are obtained as the coefficients of equation 4.1. In this study, Eviews (1997) software package version 3 is used to perform the necessary estimation.

4.4 Cointegration Technique

The concept of cointegration was first introduced by Granger (1981) and was further extended by Engle and Granger (1987) and others. The basic idea behind the cointegration is that despite being individually nonstationary, there exists certain linear combination of two or more nonstationary time series that is stationary. In this case, the variables are said to be cointegrated. This concept implies that the relevant time series would have a long-term relationship despite apparent short-term divergences. Conversely, lack of cointegration between the variables would suggest that there exists no long-run relationship among the variables such that they can wander from each other randomly. Their relationship is thus spurious.

4.4.1 Unit Roots Tests

The phenomenon of spurious regression, which was first described by Granger and Newbold (1974), arises due to regression of a nonstationary time series on the other when the variables involved in the analysis are not cointegrated. In order to avoid spurious regression, appropriate tests should be conducted to verify the stationary properties for the time series data before testing for cointegration. Furthermore, cointegration test cannot be conducted if some of the variables are stationary on levels, while others are stationary only after first differencing.

Stationarity of a time series could be identified by finding out whether the time series contains a unit root. In this study, the unit root tests developed by Dickey and Fuller (1979, 1981) and Phillips and Perron (1988) are employed for this purpose. These methods are preferred as the Augmented Dickey-Fuller (ADF) test procedure has better small sample properties compared to its alternatives, while the Phillips-Perron (PP) test allows for mildly correlated and heteroscedastic errors. The ADF and PP tests are conducted respectively on individual time series on levels based on the following equations:

$$\Delta LY_t = \mu + \beta_t + \delta LY_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta LY_{t-i} + \epsilon_{1t} \qquad (4.2)$$

$$\Delta LTOUR_t = \mu + \beta_t + \delta LTOUR_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta LTOUR_{t-i} + \epsilon_{2t} \qquad (4.3)$$

$$\Delta LREER_t = \mu + \beta_t + \delta LREER_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta LREER_{t-i} + \epsilon_{3t} \qquad (4.4)$$

The test equations contain a constant and deterministic time-trend. Lagged terms are also included to ensure the error terms are not autocorrelated.

Both unit root tests are computed to test the null hypothesis of a unit root (Ho: $\delta=0$) against the alternative of stationarity (H_a: $\delta<0$). The critical values for rejection of the null hypothesis are tabulated by MacKinnon (1991). If the null hypothesis is rejected, the series does not contain a unit root and is stationary. Accordingly, this series is said to be integrated of order zero or I(0) as the series is stationary on level. If the null hypothesis is not rejected, the series contains at least a unit root and is nonstationary. Subsequently, we have to proceed to test for the presence of unit root in the first difference of the series on the basis of the following equations:

$$\Delta^2 LY_t = \mu + \beta_t + \delta \Delta LY_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta^2 LY_{t-i} + \varepsilon_{1t} \qquad (4.5)$$

$$\Delta^2 LTOUR_t = \mu + \beta_t + \delta\Delta LTOUR_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta^2 LTOUR_{t-i} + \varepsilon_{2t} \quad (4.6)$$

$$\Delta^2 LREER_t = \mu + \beta_t + \delta\Delta LREER_{t-1} + \sum_{i=1}^{m} \gamma_i \Delta^2 LREER_{t-i} + \varepsilon_{3t} \quad (4.7)$$

Rejection of the null hypothesis in the first difference implies that the series is stationary and thus is integrated of order one or I(1), as the series has to be differenced once before it becomes stationary.

4.4.2 *Johansen Procedure and Error Correction Model*

Having established the order of integration of each series, the cointegration test is then applied to test whether the three series are cointegrated. For the time series to be cointegrated, all variables must be integrated of the same order. If the series are I(0), there is no need to proceed with cointegration tests as the standard time-series analysis would be applicable.

There are several methods available for testing cointegration. The two most widely used approaches are, namely, (i) tests based on the residuals estimated from a cointegrating regression suggested by Engle-Granger (1987) and (ii) tests based on the system of equations using vector autoregressive (VAR) models suggested by Johansen (1988, 1991) and Johansen and Juselius (1990, 1992). In this study, the Johansen methodology is preferred to the Engle-Granger approach due to the following reasons (Enders, 1995):

(i) The Johansen approach considers all variables in the system as endogenous, thus avoiding the arbitrary normalization inherent in the Engle and Granger method; and

(ii) The Johansen technique can identify all cointegrating vectors in a multivariate system, while the Engle and Granger methodology can detect at most one cointegrating vector.

According to the Granger representation theorem, if two variables are cointegrated, the dynamic relations between the variables could be examined within the framework of an error correction model [2] (ECM) instead of a VAR model. Within this framework, Johansen and Juselius use full maximum likelihood procedure to test for the number of cointegrating relationships and estimate the parameters of those cointegrating relationships in nonstationary time series (Johansen, 1988, 1991; Johansen and Juselius, 1990, 1992). The ECM can be represented as follows:

$$\Delta y_t = \mu + \prod_{i=1}^{p} y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i \Delta y_{t-i} + \varepsilon_t \quad ; \quad \varepsilon_t \sim \text{i.i.d}\,(0, \Omega) \qquad (4.8)$$

where y_t = (mx1) vector of (y_{1t}, y_{2t},….., y_{mt},)

Π_i = (mxm) matrices of coefficients

μ = intercept term

[2] Error correction mechanism was first used by Sargan (1964) and further enhanced by Engle and Granger (1987). The idea is that a proportion of the disequilibrium from one period is corrected in the next period (Engle and Granger, 1987, p254).

$$\Pi = \left(\sum_{i=1}^{p} \Pi_i - I \right) \quad \text{and} \quad \Gamma_i = \left(\sum_{j=1}^{i} \Pi_j - I \right)$$

The rank of the π matrix provides the number of cointegrating vectors (r). Johansen-Juselius method is based on estimating the π matrix in an unrestricted form, and then test if the restrictions implied by the reduced rank of π can be rejected. Johansen proposes two types of tests for determining r. The first is the maximum eigenvalue test where the null hypothesis is at most r cointegrating vectors against the alternative of exactly r+1 cointegrating vectors. Another test is the trace test where the null hypothesis is at most r cointegrating vectors against the alternative of more than r cointegrating relations.

In this study, the trace test statistic is used to test the null hypothesis, H_o^1: r=0 against H_a^1: r>0. The test statistic is as follows:

$$Q = -n \sum_{i=r+1}^{n} \ln\left(1 - \hat{\lambda}_i\right) \tag{4.9}$$

The cointegration test is based on the assumption that the series of y_t contain linear trends and the cointegrating equations have only intercepts. The critical values for the test can be obtained from Osterwald-Lenum (1992). A rejection of the null hypothesis implies that there is at least one cointegration. Then we proceed to test for higher order of cointegrating relation, where H_o^2: $r \leq 1$ against H_a^2: $r>1$. If the null hypothesis is not rejected, thus there is only one cointegrating relation. Otherwise, this implies that there is more than one cointegrating relations. Thus, further tests for higher order of cointegrating relation would be conducted until the null hypothesis is not rejected. Nevertheless, if a series has m components, there can be at most m-1 linearly independent cointegrating vectors.

As the Johansen cointegration test results are sensitive to the number of lag length of the VAR in the procedure, information criteria such as Akaike's (1969) and Schwarz's (1978) for a system of equations are applied to unrestricted VAR models to determine the appropriate order of the VAR.

4.5 Granger Causality Approach

The cointegration test detects the presence of a long-run relationship, but it does not indicate the direction of this relationship. According to Engle and Granger (1987), if variables are cointegrated, the Granger causality would run in, at least, one direction. In other words, cointegration implies causal effects. Thus, the Granger causality tests are applied to analyze the causality structure of the variables considered.

The concept of causality was first introduced by Granger (1969) and further expanded by Sims and others. According to Granger (1969), a time series X_t is said to Granger-cause another time series Y_t, if series Y_t can be predicted better by using all information on past values of X_t together with the past values of Y_t, rather than by using only the past history of Y_t, ceteris paribus. In other words, variable X_t fails to Granger-cause Y_t if

$$Pr\ (Y_{t+m} \mid \Psi_t) = Pr\ (Y_{t+m} \mid \Omega_t)$$

where Pr(•) denotes conditional probability, Ψ_t is the set of information available at time t and Ω_t is the information set obtained by excluding all information on X_t from Ψ_t.

While there are a number of different techniques used for causality test, the Granger Causality Test is employed in this study as this test is recognized to perform better than others, particularly in small samples (Guikley and Salemi, 1982; Nelson and Schwert, 1982, Geweke, Meese and Dent, 1983). Nevertheless, in accordance to the Granger representation theorem (Engle and Granger, 1987), with cointegrated variables, the standard Granger-causality tests are misspecified and an ECM should be used, that is by inserting the error correction term (ECT) as additional explanatory variable, in order to capture the short-run deviations of series from their long-run equilibrium path.

Assuming cointegration exists, the Granger Causality Test for tourism-led growth hypothesis in Malaysia is performed based on the following system of equations with the corresponding null hypotheses of noncausality:

$$\Delta LY_t = \beta_1 - \gamma_1 e_{t-1} + \sum_{i=1}^{p} \beta_{11,i} \Delta LY_{t-i} + \sum_{i=1}^{p} \beta_{12,i} \Delta LTOUR_{t-i} + \sum_{i=1}^{p} \beta_{13,i} \Delta LREER_{t-i} + \varepsilon_{1t} \quad (4.10)$$

(a) H$_o$: $\gamma_1 = 0$ & $\beta_{12,i} = 0$, ∀ i ($\Delta LTOUR_t$ does not Granger cause ΔLY_t)

(b) H$_o$: $\gamma_1 = 0$ & $\beta_{13,i} = 0$, ∀ i ($\Delta LREER_t$ does not Granger cause ΔLY_t)

$$\Delta LTOUR_t = \beta_2 - \gamma_2 e_{t-1} + \sum_{i=1}^{p} \beta_{21,i} \Delta LY_{t-i} + \sum_{i=1}^{p} \beta_{22,i} \Delta LTOUR_{t-i} + \sum_{i=1}^{p} \beta_{23,i} LREER_{t-i} + \varepsilon_{2t} \quad (4.11)$$

(c) H$_o$: $\gamma_2 = 0$ & $\beta_{21,i} = 0$, ∀ i (ΔLY_t does not Granger cause $\Delta LTOUR_t$)

(d) H$_o$: $\gamma_2 = 0$ & $\beta_{23,i} = 0$, ∀ i ($\Delta LREER_t$ does not Granger cause $\Delta LTOUR_t$)

$$\Delta LREER_t = \beta_3 - \gamma_3 e_{t-1} + \sum_{i=1}^{p} \beta_{31,i} \Delta LY_{t-i} + \sum_{i=1}^{p} \beta_{32,i} \Delta LTOUR_{t-i} + \sum_{i=1}^{p} \beta_{33,i} \Delta LREER_{t-i} + \varepsilon_{3t} \quad (4.12)$$

(e) H$_o$: $\gamma_3 = 0$ & $\beta_{31,i} = 0$, ∀ i *(*ΔLY_t does not Granger cause $\Delta LREER_t$)

(d) H$_o$: $\gamma_3 = 0$ & $\beta_{32,i} = 0$, ∀ i ($\Delta LTOUR_t$ does not Granger cause $\Delta LREER_t$)

where e_{t-1} is the ECT obtained from cointegrating regression. The coefficient of the ECT (γ) indicates the rate of adjustment to deviations from the long-run equilibrium, which is also known as the speed of adjustment.

The Wald-coefficient restrictions test, which is then used to test the null hypotheses of noncausality, would yield the F-statistics with associated p-values for each hypothesis. For any F-statistic, the null hypothesis is rejected when the p-value is significant (less than 0.05 or 5% level of significance). A rejection of the null hypothesis would imply that the first series Granger-causes the second series.

4.6 Conclusion

This chapter discusses the procedure in analyzing the relationships between tourism and economic growth in Malaysia. The investigation of both the long-run and causal relationships begins with an examination of the integration properties of the data, followed by the cointegration analysis and the Granger causality tests based on a vector error-correction model. A summary of the testing procedures is illustrated in the form of flow chart is shown in Appendix II.