Chapter 4: Data and Methodology

4.1: Data

The main data under examination in this study consists of national income and total public expenditure covering the period of 1970 to 2000 with all variables expressed in real terms. We will be using both annual data and quarterly data in this study. A simple OLS estimation will be run for annual data since the period of a sample size of 31 observations is not long enough for unbiased empirical tests to be conducted. Instead a more thorough investigation will be conducted with the quarterly data that consists a total of 124 observations. In this study, we choose the period of 1970 to 2000. It marked the beginning of a new public policy that encouraged a more active participation of public sector in the economy after the 1969 racial riots. In addition, this period represents an era of rapid industrialization, a period conforming to Wagner's idea of increasing state public expenditure.

We utilize Gross Domestic Product (GDP) to represent national income in our analysis. GDP is used instead of Gross National Product (GNP) because Wagner was more concerned with the economic activity as it was unfolding within the boundaries of Prussia and consequently corresponds to the former rather than to the latter concept (Afxentiou and Serletis, 1993). Nevertheless, the high correlation between the two concepts proved to be immaterial whichever concept is used in most empirical works. The GDP deflator has been used to obtain real values. Annual GDP figure and GDP deflator are extracted from International Financial Statistics. However, GDP data is not available on a quarterly basis and the derivation of this series from the Industrial Production Index is resorted to.[1]

The second important variable in the formulation of Wagner's Law is real public expenditure[2]. The definition of public sector in Malaysia follows the classification used in accordance to the Government Finance Statistics Manual (GSFM). Public sector in Malaysia is defined as comprising (non-financial) general government and Non-Financial Public Enterprises (NFPEs)[3]. General government in turn is comprised of federal government, state governments, local governments and statutory bodies. In this study, we will be using two different sets of public expenditure in examining the annual and quarterly data respectively. They are stated as follows:

i) annual consolidated public expenditure data extracted from Economic Report, various issues

ii) quarterly federal government expenditure extracted from Bank Negara Quarterly Bulletin, various issues

The annual consolidated public expenditure data proved to be the best approximation of public expenditure data available where the financial position of the

---

[1] Further explanation in the derivation of GDP in a quarterly basis can refer to E.C Tan (1995)

[2] Real public expenditure is derived using Consumer Price Index (CPI) as the deflator.

[3] NFPEs are public sector agencies undertaking the sale of industrial and commercial goods and services. It includes statutory bodies, Government-owned and/or Government controlled companies and agencies owned by statutory bodies. Ownership and control refer to Government or a public-sector agency controlling more than 50 percent of the equity.

federal, state and local governments is consolidated with that of approximately 50 of the larger Non Financial Public Enterprises (NFPEs). In other words, the consolidated public expenditure data best reflects the pattern of public expenditure in Malaysia. Unfortunately, the consolidated public expenditure data is not available in quarterly basis. As alternative, we use federal government expenditure as a proxy for public expenditure in Malaysia when examining quarterly series. This can be justified on the ground that the federal government's expenditure represents a big share in total public expenditure. In addition, the federal government's expenditure is usually used as the main fiscal tool in influencing the economy.

4.2 Methodology

First of all, we need to specify the model needed to estimate the Wagner's Law regression. In investigating the existence of a long-run relationship between public expenditure and national income, we consider three categories of public expenditure (i.e. total public expenditure, operating expenditure and developing expenditure). Total expenditure in turn can be categorized into operating and developing expenditure. Operating expenditure represents exhaustive expenditure and theoretically fits well into Wagner's idea. On the other hand, developing expenditure seems to adhere to Keynesian hypothesis, as it is profit augmenting and regularly used as an important fiscal variable to dampen excessive fluctuations in the economy. Thus, in addition to equation (1) we also examine equation (2) and (3) using both operating expenditure and developing expenditure.

We follow the most famous formulation of Wagner's Law, known as the Peacock-Wiseman version, in charting the relationship between public expenditure and national income. This regression is stated as follows:

$$LNTPE = c_0 + c_1\ LNGDP + u_t \qquad ....(1)$$

$$LNOPE = c_0 + c_1\ LNGDP + v_t \qquad ....(2)$$

$$LNDPE = c_0 + c_1\ LNGDP + e_t \qquad ....(3)$$

where TPE = Real Total public expenditure of consolidated public sector/federal government

OPE = Real Operating public expenditure of consolidated public sector/ federal government

DPE = Real Net Development expenditure of consolidated public sector/ federal government

GDP = Real Gross Domestic Product

u_t, v_t and e_t is serially uncorrelated random disturbance terms and LN denotes natural logarithm.

We only estimate simple OLS regressions for the annual data. As mentioned, it is not be feasible to conduct thorough tests since the sample size for the annual data is too small. According to Charemza & Deadman (1992), "..very little is known about power of cointegration tests for small samples". OLS estimation may produce substantial bias in the results of small samples. Therefore, the aim to run the simple

OLS regression for the annual data is just to provide an overall picture of the behaviour of various categories of public expenditure in Malaysia. The results obtained may not be necessarily reliable but it will be interesting to explore it since we will be using the best approximation of public expenditure data available (i.e. the consolidated public expenditure data).

The main focus of this study is on the quarterly data where thorough tests will be conducted to investigate the true behaviour of public expenditure in Malaysia where the second set of data (i.e. the federal government expenditure data) will be utilized. Earlier studies of the growth of public expenditure did not examine the stationary properties of the variables involved. The data were regarded as stationary. However, recent developments in time series analysis show that most macroeconomic time series variables have unit roots that will create spurious relationship. Thus, in order to avoid spurious relationship between the public expenditure and national income, the series must satisfy stationarity condition. In order to develop a meaningful relationship, we must first examine the stationary properties of the data using Augmented Dickey-Fuller (ADF).

If the series are found to be non-stationary (i.e. they have unit roots), co integration test needs to be run in order to check whether the linear combination of the variables is stationary or not. A test for co integration can be thought of as a pre-test to avoid spurious regression situations. If the variables can be co integrated, we regression on the levels of the variables is meaningful. We utilise two co integration tests in confirming whether co integration exists between the variables.

They are the Engle-Granger co integration test and the Johansen co integration test method. Engle-Granger provides at most one co integrating vector whereas the Johansen full Information Maximum Likelihood (ML) method provides possibility of multiple co integration vectors. (However, this does not matter in this study since we have limited variables in our co integrating regressions.) An important condition for co integration is for the variables to be integrated to the same order. A co integration test can determine the existence of a long-run relationship between the variables.

The Engle-Granger method is conducted with estimating a static long-run regression by OLS to investigate whether the residuals are stationary or not. The co integration regression's residuals is tested using the DF/ADF test and can be written as follows:

$$\Delta e_t = \phi\, e_{t-1} + \sum_{j=1}^{p} \phi_j\, \Delta e_{t-j} + v_t \qquad (4)$$

and test H_0: $\phi = 0$ versus H_1: $\phi < 0$ using appropriate critical values.

It is important to emphasize that the above equation has no intercept and time trend since residuals are expected to have zero mean and do not follow a deterministic trend. In other words, we are just testing whether the residuals of each co integrating regressions are stationary or not. Residuals, which are stationary, indicate that the variables are co integrated and vice versa.

The second method to reconfirm the results of Engle-Granger co integration test is the Johansen procedure. The Johansen test is often regarded as superior to the Engle-Granger single equation method. The statistical properties of the Johansen procedure are generally better and the power of co integration test is higher. The Johansen model is deduced by VAR model like below:

$$y_t = A_1\, y_{t-1} + \ldots + A_p\, y_{t-p} + B\, x_t + \varepsilon_t \qquad (5)$$

where y_t is a m-vector of non-stationary I(0) variables, x_t is a d vector of deterministic variables and ε_t is a vector of innovations.

It can be rewritten as a general model of the Johansen test:

$$\Delta y_t = \Pi\, y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i\, \Delta y_{t-i} + B\, x_t + \varepsilon_t \qquad (6)$$

where $\Pi = \sum_{i=1}^{p} A_i - I$, $\quad \Gamma_i = -\sum_{j=i+1}^{p} A_j$

Granger's representation theorem states that if the coefficient matrix Π has reduced rank r<m, then there exists mXr matrices α and β each with rank r such that $\Pi = \alpha\beta'$ and $\beta' y_t$ is stationary. r is the number of co integrating relations (the co integrating rank) and each column of β is the co integrating vector. The elements of α represent the adjustment parameters in the vector error correction (VEC) model. Johansen test is to estimate the Π matrix in an unrestricted form and then test whether we can reject the restrictions implied by the reduced rank of Π . E-Views provides tests for the following five model specifications considered by Johansen (1995):

a) The data have no deterministic trend and there is no intercept of trend in the co integrating regression.

b) The data have no deterministic trend and there is an intercept but no trend in the co integrating regression.

c) The data have a linear deterministic trend and there is an intercept but no trend in the co integrating regression.

d) The data have a linear deterministic trend and there is an intercept and trend in the co integrating regression; or

e) The data have a quadratic deterministic trend and there is an intercept and trend in the co integrating regression.

We consider public expenditure and national income to have a deterministic trend. As we cannot safely determine whether the data contain a time trend, both options (b) and (c) will be used to run the Johansen co integration test to see if the results differ.

Before we employ the Johansen test, we need to determine the optimal lag by running Vector Autoregressive (VAR) test. VAR is developed by Sims[4] and all the variables are considered to be endogenous. Therefore there is no a priori distinction between endogenous and exogenous variables.

---

[4] Further readings are referred to C.A. Sims, Macroeconomic and Reality, Econometrica, vol. 48, 1980, pg 1-48.

A general VAR formation is written as follows:

$$Y_t = c_0 + \sum_{i=1}^{m} \alpha_i \, Y_{t-i} + \sum_{j=1}^{m} \beta_j \, X_{t-j} \; + e_t \qquad (7)$$

$$X_t = c_0 + \sum_{i=1}^{n} \lambda_i \, X_{t-i} + \sum_{j=1}^{n} \delta_j \, Y_{t-j} \; + v_t \qquad (8)$$

Note: In our VAR test, a purely exogenous variable for seasonal dummy will be included in since we observed a jump at every fourth quarter.

The optimal lag order is chosen based on AIC. From the highest possible lag order, we perform sequential testing downward to find the minimum AIC values (here we choose lags 6 as the maximum lags). AIC is given by:

$$\text{AIC} = \log \left(\sum e_i^2 / \text{N} \right) + 2 \text{ k/N} \qquad (9)$$

where e_i^2 = sum of squared residuals

k = number of parameters (including constant) in the system

The optimal lag chosen is also subjected to residual test to ensure the nonexistence of serial autocorrelation. Number of lags should be long enough to capture the dynamics of the system but not too long in order to save degrees of freedom. The optimal lag order is to be used in the co integration, vector-error correction model and Granger causality test.

If the series can be co integrated, the short run relationship between public expenditure and GDP can be explained from an Error Correction Model (ECM). We derive ECM using the error term from the estimated co integrating regression. A standard error-correction model can be written as follows;

$$\Delta X_t = \gamma_0 + \sum_{i=1}^{m} \gamma_{1i} \, \Delta Y_{t-1} + \sum_{j=1}^{n} \gamma_{2j} \; X_{t-j} + \lambda u_{t-1} + v_t \qquad (10)$$

where Δ is the first difference operator and λ is the error-correction coefficient. The error-correction term, u_{t-1} in this model is obtained from lagging u_t in the cointegration equation by one period. λ is the speed of adjustment coefficient that measures how fast changes in Y_t adjust to return to long run equilibrium. Then, we can proceed on to employ Granger causality test to examine the directions of causality.

There is also possibility that some of the variables tested are not co integrated. No co integration leads us to say to that a long-run equilibrium relationship does not exist. In the absence of a long-run relationship between the variables, the short-run linkages between them may still remains of interest. However, without evidence of co integration, an error-correction procedure cannot be used to model short-run relationship between national income and public expenditure (Ansari et al. 1997). The short run behaviour of the relationship between national income and public expenditure can still be modeled by Granger causality test because it may still be possible that the variables are causally related in the short-run. Without co integration, the Granger causality test will be conducted using first difference series (i.e., the stationary values).

4.22 Granger Causality Model

As mentioned, the original formulation of Wagner's Law implies that in the industrialization process of an economy, government expenditure increases in size accompanying the growth in national income. This implies that the causality in Wagner's Law runs from national income to public expenditure. In other words, support for Wagner's Law requires unidirectional causality from GDP to public expenditure. The causality test is a very crucial test in this study. We cannot conclude whether our findings support Wagner's Law without conducting the Granger-Causality testing procedure.

The general definition of Granger causality is defined as follows in E-Views:

"x causes y and how much of the current y can be explained by past values of y and then to see whether adding lagged values of x can improve the explanation. y is said to be Granger-caused by x if x helps in the prediction of y, or equivalently if the coefficients on the lagged x's are statistically significant."

Since the future cannot predict the past, if variable X (Granger) causes variable Y, then changes in X should precede changes in Y. Therefore, in a regression of Y on other variables (including its own past values) if we include past or lagged values of X and it significantly improves the prediction of Y, then we can say that X (Granger) causes Y. The same applies if Y (Granger) causes X.

Eviews runs bivariate regressions in the following VAR form:

$$Y_t = c_0 + \sum_{i=1}^{m} \alpha_i \, Y_{t-i} + \sum_{j=1}^{m} \beta_j \, X_{t-j} + e_t \qquad (11)$$

$$X_t = c_0 + \sum_{i=1}^{n} \lambda_i \, X_{t-i} + \sum_{j=1}^{n} \delta_j \, Y_{t-j} + v_t \qquad (12)$$

where it is assumed that e_t and v_t are two uncorrelated white-noise series and m and n are the maximum number of lags.

Granger-causality test requires the null hypothesis of no causality is tested on a joint test that the coefficients of the lagged causal variable are significantly different from zero. The null hypothesis is therefore that x does not Granger-cause y in the first regression and that y does not Granger-cause x in the second regression. There are four possible causal relationships.

a) Independence is suggested when the sets of X and Y coefficients are not statistically significant in both regressions
b) Unidirectional causality from X to Y is indicated if the estimated coefficients on the lagged Y in (11) are statically different from zero as a group (i.e., $\sum \alpha_i \neq 0$) and the set of estimated coefficients on the lagged Y in eq is not statistically different from zero (i.e., $\sum \delta_j = 0$)

c) Unidirectional causality from Y to X exists if the set of lagged X coefficients in equation (11) is not statistically different from zero (i.e., $\sum \alpha_i = 0$) and the set of the lagged Y coefficients in equation (12) is statistically different from zero (i.e., $\sum \delta_j \neq 0$)

d) Bilateral causality is suggested when the sets of X and Y coefficients are statistically significant different from zero in both regressions.

Granger causality tests are sensitive to lag lengths. Employing arbitrarily chosen lag lengths will lead to the 'omission of relevant variable bias' if a shorter than optimal lag order is applied while it may lead to 'inclusion of irrelevant variable bias' if a longer than optimal lag order is selected. In our study, the optimal lag is chosen after running the VAR test as explained above. Results of these tests would be presented in the next section.