Chapter 5

Methodology

5.1 Introduction

For the purpose of empirical studies, this chapter will look into the methodology and statistical tests used in estimating the wage equation. Section 5.2 outlines the scope of study, followed by the data of study and the sources of data in section 5.3 and section 5.4 respectively. In section 5.5, we use few econometric techniques to investigate long-run relationships as well as short-term dynamics in the Malaysian manufacturing wage formation process for the period 1975-1997. Section 5.6 will conclude this chapter.

5.2 Scope of Study

This study will be divided into two main parts. In part one, we investigate wage formation in the Malaysian manufacturing industry as a whole in the long-run using econometric analysis. We employed the Unit Root (to examine whether the data used in the analysis is stationary or nonstationary) and Cointegration test. The Engle and Granger procedure is then employed to parameterize the short-run dynamics of real wage determination. In the second part, we impose the same long-run structure found for the whole of the industry with the aggregate data to verify the influences of labour productivity on wages for 13 selected sub-sectors of Malaysian manufacturing. The analysis is using annually data from 1975 to 1997.

The definition of manufacturing[11] used in this study follows that of the Malaysian Industrial Classification 1972 (updated 1979). Manufacturing is defined as the machinery or chemical transformation of inorganic or organic substance into new products. The assembly of the component parts of manufactured product is considered manufacturing except in cases where the activities are appropriately classified under "construction". Establishments' primary engaged in repair work are included and classified according to the type of product required.

5.3 Data

Analysis of this study is based on the secondary data. The main data collected are salaries and wages paid, value added, employment, consumer price index, unemployment rate, and union density. All variables used in this study are expressed in log form. The definitions of the variables used are given in the following section.

5.3.1 Definition of Variables

The real wage (RW) is defined as:

$$RW = \frac{\text{Nominal wage}}{CPI/100}$$

Where CPI = consumer price index (1994 base year)

[11] As cited from Technical Notes, *Annual Survey of Manufacturing Industries, 1992*. Department of Statistics, Malaysia.

In this definition wages and salaries paid refer to the value of money payments, including bonuses, cash allowances, etc. made to all paid employees during the calendar year. The employees' contribution towards the provident fund is included but the employer's contribution is excluded. Allowances to working proprietor, working partners and unpaid family workers are not included.[12]

The labour productivity (LP) is defined as (see NPC, Annual Productivity Report):

$$LP = \frac{\text{Value added}}{\text{Number of employees}}$$

According to the Department of Statistics, value added is derived as the difference between the gross value of output and the cost of input. Employees refer to all persons engaged in the establishment during December or the last pay period of the reference year. It includes working proprietors, active partners and unpaid family workers, together with paid employees, both full-time and part-time.

The union density (UD) is defined as:

$$UD = \frac{\text{Number of union membership}}{\text{Number of employees}}$$

[12] Refer to definition of concepts and term used, *Annual Survey of Manufacturing Industries, 1992*. Department of Statistic, Malaysia.

The union density refers to the proportion of the employed labour force which is unionized and as such is a more potent indicator of union power than simple membership[13].

5.4 Sources of Data

The wage, employment and value added data used are obtainable from the Annual Survey of Manufacturing Industries (1975-1997). Number of unions and membership to measure union density are obtained from Annual Trade Unions Reports compiled by Ministry of Human Resource (1975-1997). Unemployment rate are obtained from Economic Reports (1975-1997).

5.5 Analysis of Method

The econometric analysis is done using computer package *Econometric Views (EVIEWS) 3.1.*

5.5.1 Unit Root Tests

A time series data is usually claimed as being generated by stochastic or random process. A stochastic process is said to be stationary if its mean value and variance do not vary systematically over time and the value of covariance between two time periods depends only on the distance or lag between two time periods and not on the actual time

[13] See Stephen W. Smith (1994). *Labour Economics*. Routledge. London. P.120.

at which the covariance is computed. In order to see if the time series is stationary or nonstationary, a unit root test is applied.

Two equations based on the unit root principle were formed to test all the variables used in the data to examine whether the data is stationary or non stationary (Gujarati, 1997). The first equation [as shown in equation (5.1)] is Dickey-Fuller (DF) test and the second equation [equation (5.2)] is Augmented Dickey-Fuller (ADF) test. The difference between the two equations is that ADF test incorporates lagged left-hand side variables as additional explanatory variables to approximate the possible autocorrelation in the error process u_t (MacKinnon and James, 1993). In this study, the ADF test is modified to allow for the presence of an intercept (β_1) to test for the order of integration of a series.

The tests involve testing for presence of a unit root as indicated in the following hypothesis.

$$H_0: \rho = 1$$

$$H_1: \rho < 1$$

in the equation

$$\text{(DF)} \qquad \Delta Y_t = \rho Y_{t-1} + u_t \qquad (5.1)$$

$$\text{(ADF)} \qquad \Delta Y_t = \beta_1 + \rho Y_{t-1} + \sum_{i=1} \alpha_i \Delta Y_{t-i} + u_t \qquad (5.2)$$

where Δ is first differences, ρ is a parameter and u_t is assumed to be white noise error term.

The null hypothesis is that Y_t contains a unit root is tested against the alternative that it is stationary. If H_0 is not rejected, this shows that the series is nonstationary and contain one unit root. In this case, the series will proceed to test for presence of unit root with the next higher order of differencing for the equation

$$\text{(ADF)} \qquad \Delta^2 Y_t = \beta_1 + \rho \Delta Y_{t-1} + \sum \alpha_1 \Delta^2 Y_{t-1} + u_t \qquad (5.3)$$

If H_0 is rejected, ΔY_t~I(0) or Y_t~I(1). If H_0 is not rejected, this process is to be repeated with the next higher order of differencing until a rejection is found.

A time series Y_t that is stationary is said to be integrated of order zero, or denoted as Y_t~I(0). If a time series is differenced once before it comes to stationary, the original series is integrated of order 1, or Y_t~I(1). In a general form, if a time series is differentiated for d times, it is integrated of order d, or I(d).

5.5.2 Testing for Cointegration

Cointegration theory determines whether a linear combination of more than one nonstationary time series have a stationary relationship over the long run (Gujarati, 1995). If we regress one nonstationary time series on another nonstationary time series spurious regression would exist and cause the t and F test to be invalid. This problem arises because if both the time series involved exhibit strong trends (sustained upward or downward movements), the high R^2 observed is due to the presence of the trend and not to the relationship between the two.

Any equilibrium relationship among a set of nonstationary variables implies that their stochastic trends must be linked. This linkage among the stochastic trends necessitates that the variables be cointegrated (Enders, 1995). This happen because the underlying trends in the variables cancelled out and they share a common trend. After all, it means that the variables cannot move independently of each other but moving together over time. The concept of cointegration as developed by Engle and Granger (1987) provides the following definition of cointegration.

The components of the vector $y_t = (y_{1t}, y_{2t}, \ldots, y_{mt})'$ are cointegrated of order d,b, denoted by $y_t \sim CI(d,b)$ if all components of y_t are integrated of order d. There exists a vector $\beta = (\beta_1, \beta_2, \ldots, \beta_n)$ such that linear combination $\beta x_t = \beta_1 x_{1t} + \beta_2 x_{2t} + .. + \beta_n x_{nt}$ is integrated of order (d,b), where $b > 0$. The vector β is called the cointegrating vector (see Enders, 1995).

In order to establish whether there is a long-run equilibrium relationship among the variables in the model, the concept of cointegration introduced by Engle and Granger is used. Suppose that two variables – say, y_t and z_t are believed to be integrated of order 1 and we want to determine whether there exists an equilibrium relationship between the two. Engle and Granger two-step procedure (1987) as it name implies, involve two steps to propose a straightforward test whether two I (1) variables are cointegrated of order CI (1,1). The first step in the analysis is to pretest each variable to determine its order of intergration. If the results indicate that both y_t and z_t are I (1), the next step is to estimate the long-run equilibrium relationship using Ordinary Least Square (OLS) procedure in the form:

$$y_t = \beta_0 + \beta_1 z_t + e_t \tag{5.4}$$

In order to determine if the variables are actually cointegrated, the estimated residuals ($\hat{e}_t$) from the cointegrating regression are then tested for unit roots using DF or ADF test based on the following equation.

$$\text{(DF)} \qquad \Delta \hat{e}_t = a_1 \hat{e}_{t-1} + \varepsilon_t \tag{5.5}$$

$$\text{(ADF)} \qquad \Delta \hat{e}_t = a_1 \hat{e}_{t-1} + \sum_{i=1}^{p} \alpha_i \Delta \hat{e}_{t-i} + \varepsilon_t \tag{5.6}$$

where Δ = first difference operator.

If we cannot reject the null hypothesis $|a_1| = 0$, we can conclude that the residual series contains a unit root. Hence, the variables in the model are not cointegrated. Instead, the rejection of the null hypothesis implies that the residual sequence is stationary. Given that both y_t and z_t were found to be I (1) and the residuals are stationary, we can conclude that the series are cointegrated of order (1,1).

The estimated $\hat{e}_t$ is based on the estimated cointegrating parameter β_i, that minimize the sum of squared residuals ($\sum \hat{e}_t^2$). Since the residual variance is made as small as possible, the procedure is prejudiced toward finding a stationary error process; therefore the DF and ADF critical significance values are not appropriate. Engle and Granger (1987) have calculated these values and known as Engle-Granger (EG) test and Augmented Engle-Granger (AEG) test. We obtain the estimated residuals from the regression and subject it to the ADF unit root test. The Engle-Granger 5% critical values taking one lag and four lags are respectively, -3.82 and −3.75[14]. If the computed t statistic is more than any of these values (omit the negative sign), the hypothesis of a unit root is rejected. This indicates that the residuals are stationary or I (0) and consequently the variables in the model are cointegrated, that is, there is a long-run equilibrium relationship between the variables.

5.5.3 Error Correction Model (ECM)

After establishing long-run relationship between the variables, the model is examined by Error Correction Model (ECM) to evaluate the short-run dynamics between the variables. Cointegration between the variables is a necessary condition for ECM to

[14] Source: The ADF(1) critical values were generated by PC-NAIVE using 10,000 replications. The ADF(4) critical values have been taken from Engle and Yoo (1987).

hold. The ECM is a means of reconciling the short-run behaviour of an economic variable with its long-run behaviour (Gujarati, 1995). It showed that if a set of variables is cointegrated, then there exists an error-correction representation which allows flexibility in the short-run dynamic process. In an ECM, the short-run dynamics of the variables in the system are influenced by the extent of deviation from long-run equilibrium. If the variables are cointegrated, the residuals from the equilibrium regression can be used to estimate the ECM. The value of the residuals estimates the deviation from long-run equilibrium in period (t-1). Using the saved residuals from the estimation of the long- run equilibrium relationship, the ECM can be written as,

$$\Delta Y_t = \alpha_1 + \alpha_y (Y_{t-1} - \beta_1 Z_{t-1}) + \sum_{i=1} \alpha_{11,i} \Delta Y_{t-i} + \sum_{i=1} \alpha_{12,i} \Delta Z_{t-i} + \varepsilon_{yt} \qquad (5.7)$$

$$\Delta Z_t = \alpha_2 + \alpha_z (Y_{t-1} - \beta_1 Z_{t-1}) + \sum_{i=1} \alpha_{21,i} \Delta Y_{t-i} + \sum_{i=1} \alpha_{22,i} \Delta Z_{t-i} + \varepsilon_{zt} \qquad (5.8)$$

where $\alpha_2(Y_{t-1} - \beta Z_{t-1})$ denotes error correction term, α_1, α_2, α_y, α_z, $\alpha_{11,i}$, $\alpha_{12,i}$, $\alpha_{21,i}$ and $\alpha_{22,i}$ are all parameters and ε_{yt} , ε_{zt} are random disturbances. The dependent variable Y_t and independent variable Z_t are assumed to be cointegrated of order CI (1), which implies that the ECT $\alpha_2(Y_{t-1} - \beta Z_{t-1})$ is I (0).

The speed of adjustment coefficients, α_2 have important implications for the dynamics of the system. It shows that for any given value of ECT term, a large value of α_2 is associated with a large value of ΔY_t to the previous period's deviation from long-run equilibrium. If α_2 is zero, the change in Z_t does not at all respond to the deviation from long-run equilibrium in (t-1). If α_2 and α_3 is zero, then it can be said that ΔZ_t does not Granger cause ΔY_t. One or both of these coefficients should be significantly different from zero if the variables are cointegrated.

5.6 Conclusion

The data used is basically time series data. Therefore, the methodology used in this study is basically a method to estimate the regression models as well as various statistical tests to avoid any distortions in the estimation results. The unit root test is applied to test if the time series data is stationary to avoid any spurious regression. After establishing long-run relationship between the variables using OLS, the model is examined by Error Correction Mechanism (ECM) to evaluate the short-run dynamics between the variables. This chapter also discusses the important variables, sources of data and scope of study for the purpose of studying within the specified scope.