Chapter 3: Methodology

3.1. Data and Definition of Variables

Oil prices are measured using the Producer Price Index for fuels, for the Malaysian domestic economy. Economic activity, in other words, industrial output is measured using the proxy, Index for Industrial Production. Both figures are published by the Department of Statistics. Interest rates are measured using the 3-month Bank Negara Treasury Bill. Stock market activity is represented by the general stock market index, the Kuala Lumpur Composite Index (KLCI). Real stock return is denoted by *rsr*, which is calculated by taking the difference between continuously compounded return on the KLCI and the domestic inflation rate, computed based on the Consumer Price Index (CPI). All financial data are obtained from publications by the Central Bank, Bank Negara Malaysia.

Data are logarithmically transformed and the variables of concern are:

lnppi = natural log of Producer Price Index

lnipi = natural log of Industrial Production Index

lntbill = natural log of the 3-month Treasury Bill

rsr = real stock return

Monthly data are obtained for the period 1990:1 – 2000:12.

3.2. Test of Unit Root, Order of Integration and Cointegration

Before any regression can be performed on a time series, it is necessary to identify the order of integration of each variable because if the time series are processes containing unit roots, the application of standard classical regression approach may yield spurious results.

A time series process is called *stationary* if the mean and the variance are constant over time and the covariance between the values of the process between two points, say *t* and *s*, depends only on the distance between these two points and not on the time period itself. Unit root test was first introduced in the studies of Fuller (1976) and Dickey and Fuller (1979) but it does not take into account the possibility of autocorrelation in the error term, ε_t. However, the modified Augmented Dickey-Fuller (ADF) test, which involves running a higher order autoregressive regression, would take this into account. It can also be used together with a time trend component to allow for deterministic trend. The ADF test regression used in this study takes the form of:

$$\Delta X_t = \mu + \beta t\,(\text{ time }) + \delta X_{t-1} + \sum_{i=1}^{m} \delta_i\, \Delta X_{t-i} + \varepsilon_t \qquad (3.1)$$

where Δ is the first difference operator and ε_t is a stationary random error. t is the trend term and m is the number of lags of ΔX_t included, which is iid$(0,\sigma^2)$. The objective is to test if δ is not statistically different from zero. The lags of

ΔX_t are included to account for high-order serial correlation in the series as this makes a parametric correction by assuming that the X_t series follows an autoregressive process. The Dickey-Fuller t_α statistics is:

$$t_\alpha = \hat{\delta} / s.e.(\hat{\delta}) \qquad (3.2)$$

The distribution of t_α under the H_0 does not follow a t-distribution but the empirical distribution tabulated by MacKinnon (1991). The hypothesis tested is:

H_o: $\delta = 0$ against

H_1: $\delta < 0$

If $\delta = 0$, the time series X is stationary, or integrated of order zero I(0). If the null hypothesis cannot be rejected, in the level form, we try again by running the equation for the next higher-order of differencing for presence of unit root in the first differences of the variables. To reject the null hypothesis, the coefficient δ must be statistically significant and larger in absolute terms than the critical values proposed by Mackinnon (1991). Rejection of the null hypothesis in the case of testing for unit roots in the first difference implies that the series X_t is integrated of order one I(1).

Oil price, industrial production, interest rates and stock market (*lnppi, lnipi, lntbill, rsr*) are subjected to ADF test at levels and then, at first differences (*Δlnppi, Δlnipi, Δlntbill, Δrsr*) for lags, m = 0, 1, 3, 6, and 12.

If the variables of interest themselves are not stationary, it is then important to examine if any linear combination of the variables is stationary. To do this, the test of cointegration is used. Cointegration is a technique used to avoid spurious regression and the same time being able to identify long-run relations. Cointegration can be defined as: "*cointegration exists between two or more time series if one or more linear combinations of different nonstationary time series produce stationary time series.*" Norgaard et. al. (1999).

In this paper, cointegration test is conducted on the variables to check for long-term relationship between the variables of the same order of integration. The cointegration test follows the "Johansen's maximum likelihood procedure" (Johansen and Juselius, 1990; Johansen 1995) which allows estimation of multiple cointegrating vectors in a multivariate framework. Let $Y_t' = (Y_{1t}, Y_{2t}, \ldots\ldots, Y_{mt})$. Say, each of the component variable is I(1). Therefore, a VAR(p) (vector autoregression of order p) model for Y_t is:

$$Y_t = A_1Y_{t-1} + A_2Y_{t-2} + \ldots\ldots\ldots + A_pY_{t-p} \qquad (3.3)$$

The VAR in (3.3) can also be written as;

$$\Delta Y_t = \Pi\, Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i\, \Delta Y_{t-i} + \varepsilon_t \qquad (3.4)$$

The cointegration test is premised on finding the existence of long-run relationships, and how many if exist. Let r be the number of long-run cointegrating relationships. The Johansen's maximum likelihood procedure involves the following hypothesis:

H_0: r = 0 (i.e. for no cointegrating relationship) against

H_1: r > 0 (i.e. at least one cointegrating relationship)

If H_0 cannot be rejected, i.e r = 0, we can conclude that the model is;

$$\Delta Y_t = \Pi\, Y_{t-1} + \sum_{i=1}^{p-1} \Gamma_i\, \Delta Y_{t-i} + \varepsilon_t \qquad (3.5)$$

That is, it is the VAR model and no cointegration is found among the elements of Y_t. If we reject the null hypothesis of r = 0, the test is continued by testing for more than one cointegrating relationship as follows:

H_0: r = 1 (i.e. at least 1 cointegrating relationship) against

H_1: r > 1 (i.e. more than 1 cointegrating relationship)

The Johansen procedure requires that the process be repeated until a non-rejection is obtained. The statistic used for the cointegrating test is the likelihood ratio trace test, whereby the test statistic is given by:

$$Q_r = -N \sum_{i=r+1}^{m} [\log(1-\lambda_i)] \qquad (3.6)$$

where r is the hypothesized number of cointegrating vector under H_0 and λ_i is the i-th largest eigenvalue. The critical values for the trace test is given by Osterwald-Lenum (1992). If Q_r is greater than the critical value, then the null hypothesis is rejected.

For any value of $0 < r < m$, at least one cointegrating relationship exists, the model that is used should incorporate this long-run relationship. Such models is known as the Error Correction Model (ECM). For this study, $Y_t' = [lnppi, lnipi, lntbill]$.

3.3. Vector Autoregression (VAR)

If no cointegrating relationships are found in section 3.2., a VAR model may be recommended. Traditionally, regression analysis defines one dependent (endogenous) variable and several other independent (exogenous) variables. However, in the real world, the variables are quite often dependent on each other thus, they are, in fact, endogenous. This is an important implication for the statistical treatment of the time series. It requires estimation of a system of equations or otherwise, we risk loosing information, encountering biasedness, obtaining inefficient estimates and incurring invalid inferences. Hence, a dynamic statistical model such as VAR(p) is used. Let $Y_t = [Y_{1t}, Y_{2t}, \ldots\ldots\ldots\ldots, Y_{mt}]$, where all the component variables are I(0). Define the VAR(p) to be:

$$Y_t = A_0 + A_1 Y_{t-1} + A_2 Y_{t-2} + \ldots\ldots\ldots\ldots\ldots + A_p Y_{t-p} + \varepsilon_t \qquad (3.7)$$

A_0 is a m x 1 vector of constants and A_j is a m x m matrix of parameters, j = 1, 2, 3,........, p. The model can be written as;

$$Y_{1t} = a_{10} + \sum_{j=1}^{p} a_{11,j} Y_{1,t-j} + \ldots\ldots + \sum_{j=1}^{p} a_{1m,j} Y_{m,t-j} + \varepsilon_{1t}$$

$$Y_{2t} = a_{20} + \sum_{j=1}^{p} a_{21,j} Y_{1,t-j} + \ldots\ldots + \sum_{j=1}^{p} a_{2m,j} Y_{m,t-j} + \varepsilon_{2t}$$

$$\vdots$$

$$Y_{mt} = a_{m0} + \sum_{j=1}^{p} a_{m1,j} Y_{1,t-j} + \ldots\ldots + \sum_{j=1}^{p} a_{mm,j} Y_{m,t-j} + \varepsilon_{mt}$$

The VAR(p) is in its unrestricted form, treating all variables as endogenous and imposing no constraints based on supposed *a priori* knowledge. In the case of this study, $Y_t' = [\Delta lnppi, \Delta lnipi, \Delta lntbill, rsr]$.

To determine the optimal lag length, six versions are estimated where, m = 1, 2, 3, 4, 5, and 6-lag versions. Then, the optimal lag length is given by the model with the minimum value of the Akaike Information Critierion (AIC) or, Bayesian Schwarz Criterion (BIC) defined as;

$$AIC = \ln |\hat{\Sigma}| + 2\,k/n \tag{3.8}$$

$$BIC = \ln |\hat{\Sigma}| + k\,[(\ln n)/n] \tag{3.9}$$

where k is the total number of parameters (including the constant) in the system and $\hat{\Sigma}$ is the variance-covariance matrix of the system residuals.

3.3.1 Impulse Response Function(IRF)

The impulse response function (IRF) traces the path that a variable, in a VAR system, follows if it is kicked by a single unit of shock, ε_t. That is, $\varepsilon_{t-j} = 0$, $\varepsilon_t = 1$, $\varepsilon_{t+j} = 0$ for $j \neq 0$. So, the impulse response function for say, a VAR(1) system:

$$Y_t = AY_{t-1} + C\varepsilon_t, \tag{3.10}$$

Then, the response function would be:

$$C, AC, A^2C, \ldots\ldots, A^kC, \ldots\ldots \tag{3.11}$$

for 0, 1, 2, ... period ahead after the shock has taken place. A and C are matrices so the above formula can capture the complicated dynamics of any infinite order of VAR. One would expect the impulse response to decay in sine waves after the initial shock. The purpose of this investigation is to find out how much each of the variables namely, *lnipi, lntbill* and *rsr* respond to shocks in oil price changes, *Δlnppi*.

3.3.2. Variance Decomposition(VDC)

The forecast-error of the variance decomposition (VDC) analysis reveals information about the proportion of the movements of a variable in sequence of time due to its 'own' shocks (say, $\varepsilon_{\Delta lnppi,t}$) and due to shocks of other variables (e.g., $\varepsilon_{\Delta lnipi,}$) included in the model. If the shocks of other variables do not explain the forecast error variance of one of the variables say, $Y_{1t,}$ then, Y_{1t} is an exogenous variable. On the other hand, if the shocks can explain in all the forecast error variance of $Y_{1t,}$ at all forecast periods, then Y_{1t} is an entirely endogenous variable. Let the VAR be:

$$Y_t = A_0 + A(\iota)Y_{t-1} + \varepsilon_t \tag{3.12}$$

where, $\varepsilon_t \sim (0, \Sigma)$. Stationarity ensures that the equation (3.12) is invertible so that we can compute the moving-average representation:

$$Y_t = (I - (A(\iota))^{-1}A_0 + (I - A(\iota))^{-1}\varepsilon_t \tag{3.13}$$

Since the covariance matrix Σ of the VAR disturbances, ε_t is non-diagonal, it is impossible to decompose movements in the component of Y_t into innovations. However, this can be rectified as for any positive semi-definite nonsingular Σ there always exist a decomposition Σ = VV′, where V is the lower triangular orthogonal matrix, so that the equation (3.13) can be transformed to:

$$Y_t = (I - (A(\iota))^{-1}A_0 + (I - A(\iota))^{-1}V\mu_t$$

$$= \mu_0 + \sum_{s=0}^{\infty} C_s\mu_{t-s} \qquad (3.14)$$

where $\mu_t \sim (0,I)$. In equation (3.14), μ_t's are contemporaneously uncorrelated and it becomes possible then to examine the responses to innovation in each variable in the system.

In summary, the VDC will list the proportion of the s-steps ahead forecast error of Y_t explained by $(\varepsilon_{1t,}.........\varepsilon_{pt})$ for different values of s. If the VDC shows a variable due to shocks in other variables is large, it means that the shocks are an important determinant of this variable.

3.3.3 Variance-Covariance and Correlation Matrices of Residuals

The ordering of the variables in the analysis of IRF and VDC does alter the results for both the IRF and VDC, but the choice of ordering is usually made theoretically. Sadorsky (1999) says that although the effect of ordering is statistically insignificant in his case, he recommends the order of: interest rate, oil price (or, the oil price volatility), industrial output and stock returns. This ordering assumes that the monetary policy shocks are independent of contemporaneous disturbances due to other variables.

In this study, the variance-covariance and correlation matrices of the residuals are computed to examine if the relationships between different series are significant. If the results show no or, weak relationship among the residuals, then the ordering does not matter.

3.4. Oil Price Volatility

In the preceding discussions, the VAR for oil price changes does not isolate the volatility of oil price shocks from the model. It may, therefore, be worthwhile to extend the study by using a low-order Generalised Conditional Heteroskdastic (GARCH) model to examine the growth rate of oil prices, as suggested by Sadorsky (1999). Thus, a GARCH model is used to construct the conditional variation in oil prices, which in turn is used to compute normalized unexpected movements in oil prices.

3.4.1. GARCH models

In this study, a low order GARCH (1,1) for oil prices is fitted. The model is:

$$\Delta lnppi_t = \mu + \sum \beta_i \Delta lnppi_{t-i} + \varepsilon_t \qquad (3.15)$$

where;

ε_t follows that the distribution of $N(0, h_t)$,

$h_t = \alpha_0 + \alpha_1 \varepsilon^2_{t-1} + \alpha_2 h_{t-1},$

I_{t-1} is the information set available at time t-1.

The residuals of the random variable is denoted by $\hat{\varepsilon}_t$ ($\hat{\varepsilon}_t = \Delta lnppi_t - E$ $(\Delta lnppi_t \mid I_{t-1})$ where E is the expectation operator. This way, we can isolate the unexpected oil price shock or, volatility. Both the magnitude and the volatility is reflected in the forecast error, ε_t:

$$\upsilon_t = \varepsilon_t / h_t^{1/2} \qquad (3.16)$$

The GARCH equation (3.15) is fitted for different lags, m = 1, 2, 3, 4, 5, and 6. The coefficients that are significant are selected. The selected model is then tested to ensure that it is sufficient to capture the Autoregressive Conditional Heteroskedasticity(ARCH) effect.

3.4.2. Testing for ARCH Effects

Like most economic time series, oil prices tend to be volatile. We have to ensure that the volatility variable, v_t has been satisfactorily isolated in the GARCH model and that when the VAR model is estimated it is not plagued with heteroskedasticity.

This is done by using what is known as the Lagrange Multiplier Autoregressive Conditional on Heteroskedasticity test or, simply LM ARCH test. This test is discussed for a general regression model:

$$Y_t = \beta_1 + \beta_2 X_{2t} + \dots\dots\dots + \beta_k X_{kt} + \mu_t \qquad (3.17)$$

The ARCH model proposed by Engle (1982) makes the assumption that conditional on the information available at time t-1, the disturbance term is distributed as:

$$\mu_t \sim N\left[\, 0, (\alpha_0 + \alpha_1 \mu^2_{t-1})\right] \qquad (3.18)$$

Normality of distribution remains but now, the variance μ_t is dependent on the squared disturbance of time period t −1, thus giving the appearance of serial correlation. In general, the **ARCH (p)** process is written as;

$$\text{var}(\mu_t) = \sigma^2_t = \alpha_0 + \alpha_1\mu^2_{t-1} + \alpha_2\mu^2_{t-2} + \ldots\ldots\ldots.+ \alpha_p\mu^2_{t-p} \qquad (3.19)$$

If there is no autocorrelation in the error variance then, the hypothesis of H_0 : $\alpha_1 = \alpha_2 = \alpha_3 = \ldots\ldots.= \alpha_p = 0$ holds true. Acceptance of the null hypothesis suggests that $\text{var}(\mu_t) = \alpha_0$, which is a constant. The error variance is then, homoskedastic. The preceding null hypothesis can easily be tested by running the following regression:

$$\mu_t = \alpha_0 + \alpha_1\mu^2_{t-1} + \alpha_2\mu^2_{t-2} + \ldots\ldots\ldots.+ \alpha_p\mu^2_{t-p} \qquad (3.20)$$

where μ_t denote the residuals estimated from the original regression model. The test statistic is given by nR^2 where R^2 is the coefficient of determination for equation (3.20), and it is distributed by χ^2 under H_0.

In this study, the equation (3.20) is fitted for p = 1, 2, ……, 6 to check for presence of ARCH.

3.4.3. Responses to Oil Price Volatility

This research paper also examines the responses of industrial output, interest rates and real stock returns to oil price volatility. Once the price volatility has been isolated in Section 3.4.1., the same methodology for responses of these variables to oil price changes is applied. For this purpose, the VAR model of (3.7) is now run on $Y_t' = [v_t, \Delta lnipi, \Delta lntbill, rsr]$.