Chapter 4:

Data and Methodology

4.1 Introduction

In this paper, we shall examine the impact of oil price shocks and gold price shocks on the GDP in Malaysia and compared to the US. In doing so the following chapters will outline the date and methods used to carry out the examination.

4.2 The Data

The first step in developing a VAR model is to make a choice of the macroeconomic variables that are essential for the analysis. Three key macroeconomic variables; real annual gross domestic product or GDP, oil prices and gold prices are used in this research. The real annual national GDP data were obtained from the Monthly Statistical Bulletin, Department of Statistics Malaysia, various publications. In this study, since the GDP figures obtained has different constant years, the GDP constant years were modified into 1 constant year, being 1987. Real annual GDP was used instead of current annual GDP to take into account the effect of inflation.

Crude oil commodity prices is classified under world oil prices; being the average real oil prices obtained from the three main benchmark oil prices used in world trade, namely, West Texas International or WTI (from the US), Brent (from Europe) and Dubai (from Middle East). The data were obtained from the World Economic Outlook, various publications, published by the International Monetary Fund.

The gold prices, more specifically, the London PM Fix Gold Commodity prices, were obtained from Gold Council. All the data were collected from the year 1970-2002. Economic growth rates (GDP), gold prices (GP) and oil prices (OILP) are calculated from the difference of logs of real gross domestic product (GDP at 1987 constant prices), annual gold and oil prices.

4.3 Methodology

4.3.1 Initial Testing

For the method, a simple OLS regression is first created on GDP and world oil prices as well as gold prices. Economic growth rates (y) are calculated from the difference of logs of real gross domestic product (GDP at 1987 constant prices). Likewise, world oil prices (o) and gold prices (g) are calculated from the difference of logs of World Oil Prices (compiled by Energy Information Administration) and London Gold Prices (compiled by Gold Council) respectively for the period 1970-2002.

The purpose of the OLS regression for the annual data is to provide an overall picture of the relationship between the two commodities and the GDP in Malaysia and the US. The regression is stated as follows:

$$LNMGDP = c_0 + c_1\, LNMOILP + c_2\, LNMGP + u_t \qquad (4.1)$$

$$LNUSGDP = e_0 + e_1\, LNUSOILP + e_2\, LNUSGP + v_t \qquad (4.2)$$

MOILP = Annual Oil Prices (in RM)

USOILP = Annual Oil Prices (in US Dollar)

MGP = Annual Gold Prices (in RM)

USGP = Annual Gold Prices (in US Dollar)

MGDP = Malaysia's annual real gross domestic product

USGDP = U.S.'s annual real gross domestic product

Then the Augmented Dickey-Fuller (ADF) test is carried on the regression to determined whether the variables gold prices, oil prices and GDP are stationary. In a regression model, to avoid spurious regression situations, the variables must be stationary or cointegrated (i.e., a linear combination of the variables are integrated of order 0). If the time series variables have unit roots, then it will create a spurious relationship. Thus in order to avoid spurious relationship between the two commodity prices and the GDP, the series must satisfy stationary condition. If the series are non-stationary, cointegration test will be carried out to check whether the linear combination of the variables are stationary or otherwise. On the other hand, if the variables are cointegrated, we regress on the levels of the variables.

4.3.2 Vector Autoregression

Vector autoregression is also frequently used, although with considerable controversy, for analyzing the dynamic impact of different types of random disturbances on systems of variables. The VAR is a linear model used also for forecasting, impulse responses and variance decomposition. The VAR technique is appropriate because of its ability to characterize the dynamic impact of /structure of the model as well as its ability to avoid the imposition of excessive identifying restrictions associated with different economic theories. In other words, VAR does not require any explicit econometric theory to estimate the model.

A system of vector autoregression is written as:

$$A(L)\,Z(t) = \varepsilon(t) \qquad (4.3)$$

Where Z(t) is a nx1 vector of covariance stationary non-deterministic variables. A(L) is a nxn matrix polynomial in the lag operator, that is

$$A(L) = 1 - \vartheta_1 L - \ldots\ldots - \vartheta_n L^n$$

$\varepsilon(t)$ is a nx1 vector of random shocks or innovations with zero mean and covariance matrix Σ. The elements of Σ are assumed to have properties that cov (ε_{it}, ε_{it-s}) = 0 for i = 1,....,n and for t = s, cov(ε_{it}, ε_{js-t})= 0 for i=j and t=s, i=1,...,n.

The VAR model specified here focuses on three variables: real GDP (y), real oil prices (o) and real gold prices (g). A general VAR formation is as follows:

$$y_t = c_0 + \sum_{j=0}^{p} b_{yy,j}\, y_{t-1-j} + \sum_{j=0}^{p} b_{yo,j}\, o_{t-1-j} + \sum_{j=0}^{p} b_{yg,j}\, g_{t-1-j} + \varepsilon_{yt} \qquad (4.4)$$

$$o_t = c_0 + \sum_{j=0}^{p} b_{oy,j}\, y_{t-1-j} + \sum_{j=0}^{p} b_{oo,j}\, o_{t-1-j} + \sum_{j=0}^{p} b_{og,j}\, g_{t-1-j} + \varepsilon_{yt} \qquad (4.5)$$

$$g_t = c_0 + \sum_{j=0}^{p} b_{gy,j}\, y_{t-1-j} + \sum_{j=0}^{p} b_{go,j}\, o_{t-1-j} + \sum_{j=0}^{p} b_{gg,j}\, g_{t-1-j} + \varepsilon_{yt} \qquad (4.6)$$

The optimal lag order is chosen based on AIC. From the highest possible lag order, we perform sequential testing downward to find minimum AIC values. AIC is given by:

$$AIC = \log\left(\Sigma e_i^2 / N\right) + 2\,k/N \qquad (4.7)$$

where e_i^2 = sum of squared residuals k = number od parameters (including constant) in the system

The optimal lag chosen is subjected to the residual test to ensure the nonexistence of serial auto correlation. 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 will also used in the Granger Causality test.

4.3.3 Impulse Response Function (IRF) and Variance Decomposition (VD)

To interpret the estimated coefficients of a VAR, we look at impulse response functions (IFR) and variance decompositions (VD). IFR allows us to analyze dynamic behaviour while VD shows us the relative importance of each shock. The impulse response functions give the dynamic response of each endogenous variable to a shock to the system, that is by generating a moving average representation of the system. The VAR equation of (4.3) has a moving average representation:

$$
\begin{aligned}
Z(t) &= [\,A(L)]^{-1}\,\varepsilon(t) \\
&= B(L)\,\varepsilon(t) \\
&= \sum_{s=0}^{\infty} B_s\,\varepsilon\,(t-s) \qquad (4.8)
\end{aligned}
$$

where the normalization of A(L), B is an identity matrix.

Rewriting the moving average representation of equation (4.8) in term of othogonalized innovations yield the following equation:

$$Z(t) = \sum_{s=0}^{\infty} H_s v(t-s) \qquad (4.9)$$

The i^{th} equation of system (9) is :

$$Z_{it} = \sum_{s=0}^{k} h_{ij}(s)\, v_{ij}(t-s) \qquad (4.10)$$

The term $\sum_{s=0}^{k} h_{ij}(s)$ represents the impulse response function of Z_i with respect to an innovation in Z_j.

An impulse response function traces the response of an endogenous variable to a change in one of the innovations. An impulse response function describes the response of an endogenous variable to one of the innovations. Simulations for each of the aggregates are solved in response to a 1 percent innovation of the respective aggregate. In other words, the impulse response function is able to trace out the dynamic-effect adjustments for the purposes of comparative stability of the prices of oil, gold and the GDP.

Besides that, the impulse response function is also useful in providing the means to analyze the dynamic behavior of the target variables due to unanticipated shocks in the policy variables. If the innovations are not correlated with each other, interpretation is straightforward. For a series with a unit root, the IRF never dies out; however, for a trend-stationary series the IRF does die out. In any event, whether an individual time series is trend stationary or has a unit root, the relative magnitude of the IRF across different time horizons indicates the extent of the persistence of shocks to the individual series.

The variance decomposition of a VAR gives information about the relative importance of the random innovations. E-Views calculates separate variance decomposition for each endogenous variable. The first column is the forecast error of the variable for different forecast horizons. The source of this forecast error is variation in the current and future values of the innovations. The remaining columns give the percentage of the variance due to specific innovations. One period ahead, all of the variation in a variable comes from its own innovation, so the first number is always 100 percent. Again, this decomposition of variance depends critically on the ordering of equations.

4.1.3 Granger Causality Model

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

"The Granger approach to the question whether X causes Y is to see 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 Xs are statistically significant."

In other words, the variable X does not 'Granger' cause Y if and only if the past values of X do not explain Y (Granger, 1969). In terms of equation, 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 conclude that X Granger causes Y. The same applies if Y Granger causes X.

Granger causality tests requires the null hypothesis of no causality being tested on a joint test that the coefficients of the lagged causal variable are significantly different from zero. The null hypothesis is that X does not Granger causes Y in the first regression and that Y does not Granger causes X in the second regression. There are four possible causal relationship:

1) Independence is suggested when the set of X and Y coefficients are not statistically significant in both regressions.
2) Unidirectional causality from X to Y exist if the estimated coefficients on the lagged Y in (8) are statistically different from zero as a group (i.e. $\Sigma\ \alpha_i = 0$) and the set of estimated coefficients on the lagged X in equation (9) is not statistically different from zero (i.e. $\Sigma\ \delta_j = 0$).
3) Unidirectional causality from Y to X is indicated if the set of the lagged X in (8) are statistically different from zero as a group (i.e. $\Sigma\ \alpha_i = 0$) and the set of estimated coefficients on the lagged Y in equation (9) is not statistically different from zero (i.e. $\Sigma\ \delta_j = 0$).
4) Bilateral causality is suggested when the set of X and Y coefficients are statistically significant, different from zero in both regressions.