Chapter 2

Methodology and Model Specification

2.1 Methodology

According to the objectives that to estimate univariate and multivariate conditional volatility models, and volatility spillovers models for the returns on spot, forward and futures prices for the Brent, WTI, Dibai and Tapis, and to study the volatility spillovers between crude oil futures returns and oil company stock returns for the major oil companies, the methodology for modelling world crude oil price volatility and volatility spillovers are explained as follows.

First, the data used in this thesis are the daily synchronous closing price of spot, forward and futures crudes oil prices from four major crude oil markets, namely Brent, WTI, Dubai and Tapis, which are expressed in USD per barrel. Three of them are obtained from DataStream database service, while the price series for Tapis are collected from Reuters. The data 10 oil company stock prices, which are composed of the “supermajor” group of oil companies, namely Exxon Mobil (XOM, US), Royal Dutch Shell (RDS, The Netherlands), Chevron Corporation (CVX, US), ConocoPhillips (COP, US), BP (BP, UK) and Total S.A. (TOT, French), and other large oil and gas companies in the world, namely Petrobras (PBRA:Brasil), Lukoil (LKOH, Russia), Surgutneftegas (SNGS, Russia), and Eni S.p.A. (ENI, Italy), are also achieved from DataStream database services and expressed in local currencies.

Second, the synchronous price returns i for each market j are computed on a continuous compounding basis as the logarithm of closing price at the end of the period minus the logarithm of the closing price at the beginning of the period, which is defined as

$$r_{ij,t} = \log\left(P_{ij,t} / P_{ij,t-1}\right) \quad (1),$$

where $P_{ij,t}$ and $P_{ij,t-1}$ are the closing prices of crude oil price i in market j for days t and $t-1$, respectively.

Third, the plot of synchronous price returns and their descriptive statistics, namely mean, maximum, minimum, standard deviation, skewness, kurtosis and Jarque-Bera Lagrange multiplier statistics, are expressed in order to check whether the distributions of these returns are volatility clustering and has a normal distribution.

Four, test for a unit root in every return series. The Augmented Dickey-Fuller (ADF) statistic and Phillips-Perron statistic (PP) are applied in the test. They are a negative number, which are more negative, the stronger to rejection of the hypothesis that there is unit root at some level of confidence.

Augmented Dickey-Fuller test: (Dickey and Fuller (1979))

To test the unit root against the alternative of stationarity of the return series the ADF test is given as follows:

$$\Delta y_t = \alpha + \beta t + \theta y_{t-1} + \sum_{i=1}^{p} \phi_i \Delta y_{t-1} + \varepsilon_t \quad (2),$$

where α is constant, β is the coefficient on a time trend and p is the log order of the autoregressive process. The presence of the deterministic element α and βt determine the difference regression. Imposing the constraints $\alpha = 0$ and $\beta = 0$ corresponds to modelling a random walk and using constraint $\alpha = 0$ similar to modelling random walk with drift, and equation (2) resemble to modelling both drift and a linear time trend model. The parameter of interest in all the regression equations is θ, if $\theta = 0$ the return series contains a unit root. Therefore, under the null hypothesis is $\theta = 0$ against alternative hypothesis is $\theta < 0$, it is rejected when compare with MacKinnon critical value (MacKinnon (1991, 1996)), means that the return series is stationary.

Phillips-Perron statistic (PP): (Phillips and Perron (1988))

PP test allows fairly mild assumption that does not assume the specific type of serial correlation and heteroskedastity in the disturbances, and can have higher power than the ADF test under a wide range of circumstance. The PP test is based on the statistic:

$$t_{\phi} = t_{\phi}\left(\frac{\gamma_0}{f_0}\right)^{1/2} - \frac{T\left(f_0 - \gamma_0\right)\left(se\left(\hat{\phi}\right)\right)}{2 f_0^{1/2} s} \qquad (3),$$

where $\hat{\phi}$ is the estimate, and t_{ϕ} is the *t*-ratio of ϕ, $se\left(\hat{\phi}\right)$ is coefficient standard error, and s is the standard error of the test regression. In addition, γ_0 is a consistent estimate of the error variance in (2). The remaining f_0 is an estimator of the residual spectrum at frequency zero and T is number of observation. Under the null hypothesis is $\theta = 0$, which if it is rejected when compare with MacKinnon lower- tail critical and *p*-value, means that the return series is stationary.

Five, a wide range of univariate and multivariate volatility models have been used to estimate and forecast volatility and volatility spillovers with symmetric and asymmetric effects. These models are presented in model specification part.

2.2 Model Specification

2.2.1 Univariate Conditional Volatility Models

Following Engle (1982), consider the time series $y_t = E_{t-1}(y_t) + \varepsilon_t$, where $E_{t-1}(y_t)$ is the conditional expectation of y_t at time $t-1$ and ε_t is the associated error. The generalized autoregressive conditional heteroskedastity (GARCH) model of Bollerslev (1986) is given as follows:

$$\varepsilon_t = \sqrt{h_t}\eta_t \quad , \quad \eta_t \sim N(0,1) \tag{4}$$

$$h_t = \omega + \sum_{j=1}^{p} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{q} \beta_j h_{t-j} \tag{5}$$

where $\omega > 0$, $\alpha_j \geq 0$ and $\beta_j \geq 0$ are sufficient conditions to ensure that the conditional variance $h_t > 0$. The parameter α_j represents the ARCH effect, or the short-run persistence of shocks to returns, and β_j represents the GARCH effect, where $\alpha_j + \beta_j$ measures the persistence of the contribution of shocks to return i to long run persistence.

Equation (5) assumes that the conditional variance is a function of the magnitudes of the lagged residuals and not their signs, such that a positive shock $(\varepsilon_t > 0)$ has the same impact on conditional variance as a negative shock $(\varepsilon_t < 0)$ of equal magnitude. In order to accommodate differential impacts on the conditional variance of positive and negative shocks, Glosten, et al. (1993) proposed the asymmetric GARCH, or GJR model, which is given by

$$h_t = \omega + \sum_{j=1}^{r} \left(\alpha_j + \gamma_j I\left(\varepsilon_{t-j} \right) \right) \varepsilon_{t-j}^2 + \sum_{j=1}^{s} \beta_j h_{t-j} \tag{6}$$

where

$$I_{it} = \begin{cases} 0, & \varepsilon_{it} \geq 0 \\ 1, & \varepsilon_{it} < 0 \end{cases}$$

is an indicator function to differentiate between positive and negative shocks. When $r = s = 1$, sufficient conditions to ensure the conditional variance, $h_t > 0$, are $\omega > 0$, $\alpha_1 \geq 0$, $\alpha_1 + \gamma_1 \geq 0$ and $\beta_1 \geq 0$. The short run persistence of positive and negative shocks are given by α_1 and $(\alpha_1 + \gamma_1)$, respectively. When the conditional shocks, η_t, follow a symmetric distribution, the short run persistence is $\alpha_1 + \gamma_1/2$, and the contribution of shocks to expected long-run persistence is $\alpha_1 + \gamma_1/2 + \beta_1$.

In order to estimate the parameters of model (4)-(6), maximum likelihood estimation is used with a joint normal distribution of η_t. However, when η_t does not follow a normal distribution or the conditional distribution is not known, quasi-MLE (QMLE) is used to maximize the likelihood function.

Bollerslev (1986) showed the necessary and sufficient condition for the second-order stationarity of GARCH is $\sum_{i=1}^{r}\alpha_i + \sum_{i=1}^{s}\beta_i < 1$. For the GARCH(1,1) model, Nelson (1991) obtained the log-moment condition for strict stationary and ergodicity as $E\left(\log\left(\alpha_1\eta_t^2\right)+\beta_1\right)<0$, which is important in deriving the statistical properties of the QMLE. For GJR(1,1), Ling and McAleer (2002a, 2002b) presented the necessary and sufficient condition for $E\left(\varepsilon_t^2\right)<\infty$ as $\alpha_1+\gamma_1/2+\beta_1<1$. McAleer, et al. (2007) established the log-moment condition for GJR(1,1) as $E\left(\log\left(\alpha_1+\gamma_1 I\left(\eta_t\right)\eta_t^2+\beta_1\right)\right)<0$, and showed that it is sufficient for consistency and asymptotic normality of the QMLE.

2.2.2 Multivariate Conditional Volatility Model

The typical specification underlying the multivariate conditional mean and conditional variance in returns are given as follows:

$$y_t = E\left(y_t \middle| F_{t-1}\right)+\varepsilon_t \qquad (7)$$

$$\varepsilon_t = D_t\eta_t$$

$$\operatorname{var}\left(\varepsilon_t \mid F_{t-1}\right) = D_t\Gamma D_t$$

where $y_t = (y_{1t},...,y_{mt})'$, $\eta_t = (\eta_{1t},...,\eta_{mt})'$ is a sequence of independently and identically distributed (i.i.d.) random vectors, F_t is the past information available up to time *t*, $D_t = \text{diag}\left(h_1^{1/2},...,h_m^{1/2}\right)$, m is the number of returns, and $t = 1,...,n$, (see Chan, et al. (2003), and Bauwens, et al. (2006). The constant conditional correlation (CCC) model of Bollerslev (1990) assumes that the conditional variance for each return, h_{it}, $i = 1,..,m$, follows a univariate GARCH process, that is

$$h_{it} = \omega_i + \sum_{j=1}^{r} \alpha_{ij} \varepsilon_{i,t-j}^2 + \sum_{j=1}^{s} \beta_{ij} h_{i,t-j} \tag{8}$$

where α_{ij} represents the ARCH effect, or short run persistence of shocks to return *i*, and β_{ij} represents the GARCH effect, or the contribution of shocks to return *i* to long run persistence, namely $\sum_{j=1}^{r} \alpha_{ij} + \sum_{j=1}^{s} \beta_{ij}$.

The conditional correlation matrix of CCC is $\Gamma = E\left(\eta_t \eta_t' \middle| F_{t-1}\right) = E\left(\eta_t \eta'\right)$, where $\Gamma = \{\rho_{it}\}$ for $i, j = 1,...,m$. From (7), $\varepsilon_t \varepsilon_t' = D_t \eta_t \eta' D_t$, $D_t = \left(\text{diag } Q_t\right)^{1/2}$, and $E\left(\varepsilon_t \varepsilon_t' \middle| F_{t-1}\right) = Q_t = D_t \Gamma D_t$, where Q_t is the conditional covariance matrix. The conditional correlation matrix is defined as $\Gamma = D_t^{-1} Q_t D_t^{-1}$, and each conditional correlation coefficient is estimated from the standardized residuals in (7) and (8). Therefore, there is no multivariate estimation involved for CCC, which involves *m* univariate GARCH models, except in the calculation of the conditional correlations.

Although the CCC specification in (8) is a computationally straightforward "multivariate" GARCH model, it assumes independence of the conditional variances across returns and does not accommodate asymmetric behavior. In order to incorporate interdependencies, Ling and McAleer (2003) proposed a vector autoregressive moving average (VARMA) specification of the conditional mean in (7), and the following specification for the conditional variance:

$$H_t = W + \sum_{i=1}^{r} A_i \vec{\varepsilon}_{t-i} + \sum_{j=1}^{s} B_j H_{t-j} \tag{9}$$

where $H_t = \left(h_{1t},...,h_{mt}\right)'$, $\vec{\varepsilon} = \left(\varepsilon_{1t}^2,...\varepsilon_{mt}^2\right)'$, and W, A_i for $i=1,..,r$ and B_j for $j=1,..,s$ are $m \times m$ matrices. As in the univariate GARCH model, VARMA-GARCH assumes that negative and positive shocks have identical impacts on the conditional variance. In order to separate the asymmetric impacts of the positive and negative shocks, McAleer, Hoti and Chan (2009) proposed the VARMA-AGARCH specification for the conditional variance, namely

$$H_t = W + \sum_{i=1}^{r} A_i \vec{\varepsilon}_{t-i} + \sum_{i=1}^{r} C_i I_{t-i} \vec{\varepsilon}_{t-i} + \sum_{j=1}^{s} B_j H_{t-j} \tag{10}$$

where C_i are $m \times m$ matrices for $i=1,..,r$, and $I_t = \mathrm{diag}\left(I_{1t},...,I_{mt}\right)$, where

$$I_{it} = \begin{cases} 0, & \varepsilon_{it} > 0 \\ 1, & \varepsilon_{it} \le 0 \end{cases}.$$

If $m = 1$, (9) collapses to the asymmetric GARCH, or GJR model. Moreover, VARMA-AGARCH reduces to VARMA-GARCH when $C_i = 0$ for all *i*. If $C_i = 0$ and A_i and B_j are diagonal matrices for all *i* and *j*, then VARMA-AGARCH reduces to the CCC model. The parameters of model (7)-(10) are obtained by maximum likelihood estimation (MLE) using a joint normal density. When η_t does not follow a joint multivariate normal distribution, the appropriate estimator is QMLE.

In order to forecast 1-day ahead conditional correlation, we use rolling windows technique and examine the time-varying nature of the conditional correlations using VARMA-GARCH and VARMA-AGARCH. Rolling windows are a recursive estimation procedure whereby the model is estimated for a restricted sample, then re-estimated by adding one observation at the end of the sample and deleting one observation from the beginning of the sample. The process is repeated until the end of the sample. In order to strike a balance between efficiency in estimation and a viable number of rolling regressions, the rolling window size is set at 2008 for all data sets.

However, many empirical studies have presented that the conditional correlations are not constant over time. For example, Solnik, et al. (1996) and Hunter and Simon (2005) showed that the US and other major bond market returns correlations are not constant, but are influenced by fundamentals and market conditions. De Santis and Gerard (1998), Longin and Solnik (2001), found that equity correlations increase (decrease) during bear (rally) markets. Cappiello, et al. (2006) presented evidence that conditional correlation between equity and bond returns decline when stock markets suffer from financial turmoil.

Since the correlations can change over time, it is unavoidable to model the dynamic conditional correlations across assets. In case of volatility spillovers, the dynamic conditional correlations are also important in constructing multivariate models that incorporate mean and volatility spillovers. Two dynamic conditional correlation models are the DCC model of Engle (2002) and the GARCC model of McAleer, et al. (2008).

Unless η_t is a sequence of iid random vectors, or alternatively a martingale difference process, the assumption that the conditional correlations are constant may seen unrealistic. In order to make the conditional correlation matrix time dependent, Engle (2002) proposed a dynamic conditional correlation (DCC) model, which is defined as

$$y_t \mid \Im_{t-1} \sim (0, Q_t) \quad , \quad t = 1, 2, \ldots, n \tag{11}$$

$$Q_t = D_t \Gamma_t D_t , \tag{12}$$

where $D_t = \mathrm{diag}\left(h_1^{1/2}, \ldots, h_m^{1/2}\right)$ is a diagonal matrix of conditional variances, and $\Im_t$ is the information set available to time *t*. The conditional variance, h_{it}, can be defined as a univariate GARCH model as follows:

$$h_{it} = \omega_i + \sum_{k=1}^{p} \alpha_{ik} \varepsilon_{i,t-k} + \sum_{l=1}^{q} \beta_{il} h_{i,t-l} \ . \tag{13}$$

If η_t is a vector of i.i.d. random variables, with zero mean and unit variance, Q_t in (12) is the conditional covariance matrix (after standardization, $\eta_{it} = y_{it} \big/ \sqrt{h_{it}}$). The η_{it} are used to estimate the dynamic conditional correlations, as follows:

$$\Gamma_t = \left\{ (diag(Q_t)^{-1/2} \right\} Q_t \left\{ (diag(Q_t)^{-1/2} \right\} \tag{14}$$

where the $k \times k$ symmetric positive definite matrix Q_t is given by

$$Q_t = (1 - \theta_1 - \theta_2)\bar{Q} + \theta_1 \eta_{t-1} \eta'_{t-1} + \theta_2 Q_{t-1} \tag{15}$$

in which θ_1 and θ_2 are scalar parameters to capture the effects of previous shocks and previous dynamic conditional correlations on the current dynamic conditional correlation, and θ_1 and θ_2 are non-negative scalar parameters. When $\theta_1 = \theta_2 = 0$, $\bar{Q}$ in (15) is equivalent to the CCC model. As Q_t is a conditional on the vector of standardized residuals, (15) is a conditional covariance matrix, and $\bar{Q}$ is the $k \times k$ unconditional variance matrix of η_t. DCC are not linear but can often be estimated very simply with two step method based on the likelihood function the first is a series of univariate GARCH estimates and the second the correlation estimate. For further details, and critique of the DCC model, see Caporin and McAleer (2009).