Chapter 4
Methodology and Data base

As specified in Chapter 1, the present study examines the micro aspects of foreign exchange rate behaviour by using the high frequency data that are made available by the introduction of electronic broking system. With the help of literature survey, the objectives of the present study is formulated (see chapter 3). To examine these objectives, we have used different methodologies for each objective. In this chapter, we will discuss the methodology adopted for each objective separately. And the database for the study will be discussed briefly at the end of this chapter.

One major characteristic of any financial time series data is that the variance will change over time period. That is the variance of returns on the financial asset is not constant over a time. The studies done by Hsieh (1988) and Diebold (1988) on daily data and the study done by Zhou (1996) on high frequency data on foreign exchange rate showed that the distribution of daily returns is approximately symmetric and heavy tailed. These result shows that the data are independently distributed as a normal distribution whose mean and variance changes over time. The reason for
this hypothesis is that the information flow that causes change in prices is not constant over time. For studying the issue of volatility, there are array of time series models in the literature which are classified as Autoregressive Conditional Heteroscedastic models (ARCH) developed by Engle (1982) and generalized by Bollerslev (1986) (GARCH model); extended by Engle et.al (1987) (GARCH-in-Mean), Bollerslev & Ghysels (1996) (Periodic -ARCH) and Rabemananjara & Zakoian (1993) (Threshold ARCH). But in this study, we try to adopt the methodology developed by Zhou (1996) and Bollerslev & Ghysels (1996) for estimating the volatility of high frequency data on foreign exchange markets. This will be explained algebraically in the next section.

To test the impact of information (public) on the returns, we use the methodology developed by Goodhart et.al (1993) where the study estimates simple ordinary least squares method and further it applies GARCH-M model. The study presents the results of these two models with and without information and examines the effect of information on the rate changes.
Further, the study tries to examine the presence of private information and its impact on the exchange rate changes. Here the private information variable considered is order flows (this is based on the multi dealer model that is explained in Chapter 2). Since, order flows are of daily based, the study examines the impact of order flows on the daily end period exchange rate. Finally, to study the superiority of micro variables over macro fundamentals, we considered the interest rates of India and the US, as no other macro information is available at the daily frequency. For this purpose, we follow the methodology developed by Evans & Lyons (1999). All the three methodologies are briefly explained in the next section.

**Volatility Estimates**

*Periodic-GARCH model*

In the financial markets literature, to test the volatility of any time series data, the researchers generally used the conditional variance based models as the variance on the returns are not uniform over the period. From the literature survey that has been carried out in Chapter 3, we found that the most conventional methods in examining the volatility are Autoregressive Conditional Heteroscedasticity (ARCH) based models. In
this study also we try to test the volatility of exchange rate data that are at high frequency. Here we follow the Periodic-GARCH model that was developed by methodology developed by Bollerslev & Ghysels (1996). Before discussing this method, for our better understanding, we discuss the simple ARCH model and extend it to P-GARCH model.

The simple version of the ARCH model that was developed by Engle is

\[ Y_t = \beta'X_t + \varepsilon_t \quad (4.1) \]

\[ \varepsilon_t = \eta_t(\alpha_0 + \alpha_t \varepsilon_{t-1}^2)^{1/2} \]

where \( \eta_t \) is a standard normal variate. That is

\[ E(\varepsilon_t/\varepsilon_{t-1}) = 0 \]

Therefore

\[ E(\varepsilon_t) = 0 \text{ and } E(Y_t) = \beta'X_t \]

Also

\[ \text{Var}(\varepsilon_t/\varepsilon_{t-1}) = E(\varepsilon_t^2/\varepsilon_{t-1}^2) \]

\[ = E(\eta_t^2)(\alpha_0 + \alpha_t \varepsilon_{t-1}^2) \]

\[ = \alpha_0 + \alpha_t \varepsilon_{t-1}^2 \]

Here, \( \varepsilon_t \) is heteroscedastic since it is conditional upon \( \varepsilon_{t-1} \).

If \( \varepsilon_t \) is unconditional, the variance of \( \varepsilon_t \) will be equal to
\[ \text{Var}(\varepsilon_t) = \alpha_0 + \alpha_1 \text{Var}(\varepsilon_{t-1}). \]

It is clear from the above equation that unconditional variance is unchanging over time.

The log likelihood function for this model was provided by Engle, which is conditioned on the initial values of \( Y_0 \) and \( X_0 \). Therefore the log-likelihood function for \( N \) observations is

\[
\ln L = \left(-1/2\right)\sum_{i=1}^{N} \ln(\alpha_0 + \alpha \varepsilon_{i-1}^2) - \left(1/2\right)\sum_{i=1}^{N} \frac{\varepsilon_i^2}{\alpha \varepsilon_{i-1}^2},
\]

where

\[ \varepsilon_t = Y_t - \beta' X_t \]

Through maximisation of 4.2 we get the estimate of \( \beta \).

Alternatively, we can also estimate the \( \beta \) with the help of Lagrange multiplier test for ARCH process against the null hypothesis that the disturbance in the series is conditionally homoscedastic. The statistic for Lagrange multiplies test is \( N \times R^2 \), where \( R^2 \) is estimated through the simple regression of \( \varepsilon_t^2 \) on \( \varepsilon_{t-1}^2 \).
In the recent times, there have been so many extensions on the ARCH model with minor modifications on the estimation process. Major extension of this model is its generalisation as suggested by Bollerslev. The model of generalised ARCH model is defined as follows.

Only difference from the ARCH models is the expression of conditional variance, which is given below:

$$\sigma_t^2 = \alpha_0 + \delta_1 \sigma_{t-1}^2 + \delta_2 \sigma_{t-2}^2 + ... + \delta_p \sigma_{t-p}^2 + \alpha_1 \epsilon_{t-1}^2 + \alpha_2 \epsilon_{t-2}^2 + ... + \alpha_q \epsilon_{t-q}^2$$  \hspace{1cm} (4.3)

where \(\sigma_t^2\) is the variance of \(\epsilon_t\), which is conditional upon the information set at time \('t'\). (The notations in this chapter are independent of the notations that are used in the previous chapters).

In equation 4.3, the conditional variance evolves over time in what might be a very complicated manner, depending on the parameter values and on \('p'\), the order of autoregressive part, and \('q'\). Further, this GARCH\((p, q)\) model is also extended. The major extension of this model is the Periodic-GARCH model developed by Bollerslev & Ghysels (1996). The logic behind this extension is to adjust for the seasonality in the high
frequency financial time series data, which is a well-known phenomena.

The P-GARCH model will be briefly explained below.

Let \( \varepsilon_t \) be a discrete-time real-valued stochastic process which satisfies, like in GARCH model,

\[
E[\varepsilon_t/\varphi_{t-1}] = 0
\]

and

\[
E[\varepsilon_t^2/\varphi_{t-1}] = \sigma_t^2 = \alpha_0 + \sum_{j=1}^{q} \alpha_j \varepsilon_{t-j}^2 + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2
\]

Where \( \varphi_{t-1} \) denotes the Borel \( \sigma \)-field filtration based on the realisation of the stochastic process upto time \( t-1 \).

To adjust for the periodicity in the model, we need to have time varying coefficient model than the fixed coefficient model like equation 4.5. To define such models, consider a modified Borel \( \sigma \)-field filtration in which the usual \( \varphi_{t-1} \) is augmented by a process adjusting the stage of periodic cycle at each point in time, say \( \varphi_{t-1}^t \). Then the P-GARCH model can be defined as

\[
E[\tilde{\varepsilon}_t/\varphi_{t-1}^t]
\]
And

$$E[\tilde{\varepsilon}^2 / \varphi_{i,t-1}] = \alpha_{0,t(i)} + \sum_{i=1}^{q} \alpha_{i,t(i)} \tilde{\varepsilon}_{i,t}^2 + \sum_{j=1}^{p} \beta_{j,t(i)} \tilde{\sigma}_{i-j}^2$$  \hspace{1cm} (4.7)$$

where \(s(t)\) refers to the stage of periodic cycle at time 't'. One should note from the above equations that \(\varphi_{i,s}^t\) appears in both conditional mean and variance equations. Therefore, \(\tilde{\varepsilon}\), may differ from the \(e\), that is defined in equation 4.4. Here, the lag lengths 'p' and 'q' will not change according to \(s(t)\) set as they are set to their optimal orders across all stages of the periodic cycle.

The estimation and testing procedures for this model is based on the log-likelihood estimation. Let \(\theta\) denote the vector of parameters for all 'S' seasons to be estimated. That is, from equation 4.7,

$$\theta = (\alpha_{0,s}, \alpha_{15}, ..., \alpha_{qS}, \beta_{15}, ..., \beta_{ps}) \quad \text{for} \quad s = 1, ..., S.$$ 

Then the log-likelihood estimation for a strong P-GARCH model is

$$L_N(\theta / \varphi_{0}^S) = \sum_{i=1}^{N} l_i(\theta_{s(i)})$$  \hspace{1cm} (4.8)$$

**Zhou model**

In this model volatility is estimated between a given interval \([a,b]\) from a series of observations within the interval.
Let \( \sigma^2 = \tau(b) - \tau(a) \quad (4.9) \)
be the volatility of the interval and \( S(t_i), i=0,1,\ldots,n; \) be series of observations within the interval. Further, it is assumed that
\[
S(t) = B(t) + \varepsilon_t \quad (4.10)
\]
Where \( B(.) \) is the standard Brownian motion.

Assuming the \( \varepsilon_i, i=1,2,\ldots,n \) are independent, the unbiased estimator of the volatility can be estimated as:
\[
\hat{\sigma}^2_{t_i} = \sum_{i=1}^{n} (X_i^2 + X_iX_{i-1} + X_{i+1}X_i) \quad (4.11)
\]
where \( X_i = S(t_i) - S(t_{i-1}) \)

Therefore, the variance of the estimator (4.11) can be estimated as follows:
Let \( \text{Var}(X_i) = \sigma^2/n \) and \( \text{Var}(\varepsilon_i) = \eta^2 \). Then the variance will be
\[
\text{Var}(\hat{\sigma}^2_{t_i}) = \sigma^4 \left( \frac{6}{n} + \frac{8 \eta^2}{\sigma^2} + \frac{8 n \eta^4}{\sigma^4} - 2 \frac{\sigma^4}{n^2} - 4 \eta^4 \right) \quad (4.12)
\]
(For the proof of the estimator, see Zhou(1996)).

This procedure has little advantage over the ARCH type models. It is simple and easy to update when the new information comes into the market. But the disadvantage with this model is that it cannot take the full advantage of high-frequency data with optimal observation frequency.
Estimation of Information Effects on the exchange rates

To study the impact of information arrivals on the exchange rate, we followed the methodology developed by Goodhart, et.al (1993) where they try to examine the macroeconomic news with the help of GARCH-in-Mean model. The study also uses the simple ordinary least square technique by incorporating the information in the form of dummy variable. The estimated ordinary least square model is

\[ \Delta X_t = \alpha_0 + \alpha_t X_{t-1} + \sum \beta_i \Delta X_{t-i} + \gamma N \]  

(4.13)

where 'N' is the news variable.

And the model with GARCH-M term is

\[ \Delta X_t = \alpha_0 + \alpha_t X_{t-1} + \sum \beta_i \Delta X_{t-1} + \delta \sigma_t^2 \]  

(4.14)

Both the equations were estimated with and without the presence of news variable in the system.

To examine the relative importance of micro and macro variables, we estimated a multiple regression model by taking different combinations in including the variables.

All these models are estimated with the help of RATS program and are presented in a tabular form in Tables section. The results were
thoroughly discussed in Chapter 6 and the conclusions were drawn accordingly.

**Data Base**

The present study uses the high frequency data (tick-by-tick data) in the case of Indian rupee/US dollar (INR/USD), US dollar/Euro (USD/EURO) and Japanese Yen/US dollar (YEN/USD) in examining the objectives that are formulated in Chapter 1. The period for which the data obtained is from the whole month of August 1999 in the case of INR/USD and the number of observations being 4981 points; from August 5th to 6th, 1999 in the case of YEN/USD and the observations being 8158; and in the case of USD/EURO we have taken for one day, i.e., on August 2nd 1999 with 13318 observations. Though we have the data for whole month of August 1999 in the case of YEN/USD and USD/EURO, we could not take, as the software could not deal with the huge data sets of these currencies. Hence, we restricted the period for two days in the case of YEN/USD and for one day in the case of USD/EURO. The choice of days in the case of YEN/USD is basically to examine the effect of employment report of US which will be released on every Friday at 8.30 EST (Friday falls on August 6th). And in
the case of USD/EURO, the choice of August 2nd is being first trading day of the month and also we could not open the data set for more than two days. One more reason for choosing this day is to examine the impact of macro economic report that will be released on every Monday at 8.30 EST. These data are collected from the Reuters screen which exhibits the bid, ask (both high and low) and the timing of the trading. These data are collected by the Olsen & Associates and are tabulated. Olsen & Associates has provided these data sets to the present researcher.

In the case of interest rates, the daily Indian call money rates were taken from the various issues of Reserve Bank of India Monthly Bulletin. For US, we have taken federal fund rate from the Federal Reserve's website (ftp.federalreserve.org). Order flows in the case of INR/USD, we have calculated from the RBI's sales and purchases as the proxy for effective demand in the market. We have also used daily total number of transactions occurred on the Reuters screen as a proxy for order flow (we have taken number of transactions as the proxy for the order flow based on the view of Demsetz (1968) and Stoll (1999) that number of transactions in the market reflects the market activism). In the case of YEN/USD, we
could not get the information on the sale and purchases of the central bank, but could use the daily total transactions as the only proxy variable.

In the succeeding chapters, we will study the nature of high frequency data in foreign exchange markets and discuss the results that are estimated, which are tabulated at the end of the thesis.