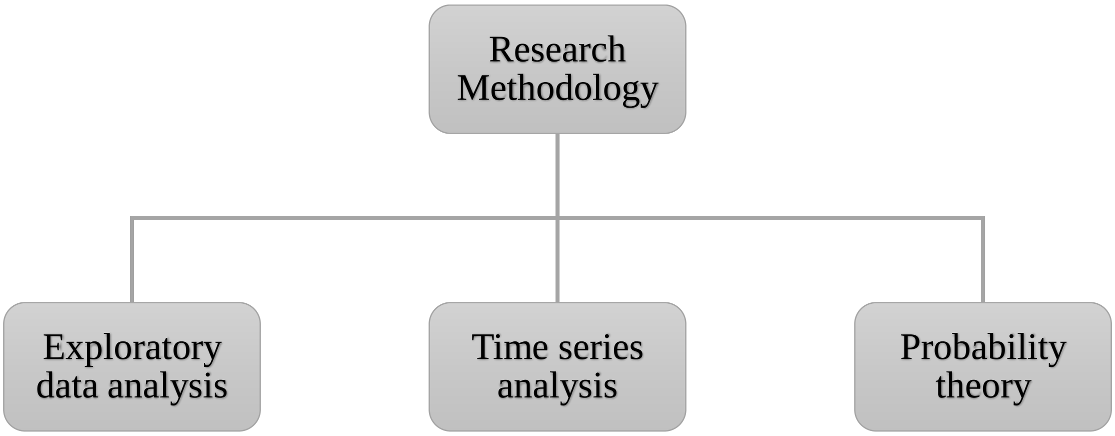
CHAPTER 5

RESEARCH METHODOLOGY

RESEARCH METHODOLOGY

As indicated in the title, this chapter includes a detailed description of the research methods employed in the study and the logic behind the use in context of the study carried out.



For the purpose of the research, the analysis is divided into three parts. The first part follows the exploratory data analysis; while, the second part adopts time series analysis. The final part of the study adheres to probability theory and statistics. The section containing the exploratory data analysis discusses the basic features of the data used in the study which provides simple summary statistics of the sample used. Thereafter, time series analysis is carried out which consist of models used for variety of reasons – predicting future outcomes, developing a pattern of past outcomes, making policy recommendations and much more. The present data is modelled with the help of Auto regressive conditional heteroscedasticity and General Auto regressive conditional heteroscedasticity models. These are the time-domain approach of time series modelling future values as a function of past as well as present values. This approach is based on time series regression of present value on its own past values and also past values of other variables. This kind of an approach is very popular in econometric time series analysis.

Further in the study, copula models are constructed which are high dimensional statistical applications that estimate the distribution of the random variables and help in determining the strength of dependence structure prevailing in the data. These are basically mathematical tools applied to areas of finance to identify market risk, credit risk, operational risk, option pricing, portfolio value-at-risk and economic capital adequacy. Since distribution in financial markets are always non-normal in nature such techniques are used to deal with skewed or asymmetric distributions.

5.1 DATA

The indices used in the study are widely accepted benchmark indices for two major Indian stock markets. These popular indices are the Bombay Stock Exchange's Sensex and the National Stock Exchange's Nifty50. The BSE Sensitivity Index (Sensex) was the company's first index, released in 1986. The BSE Sensex index of equity share prices was first introduced in 1978-79, with a base value of 100. It is made up of 30 firms and the companies were chosen to represent all of the key economic sectors based on their market capitalization, turnover, and the quality of their fundamentals. It has offered 21 indices in the last 15 years, including 12 sectoral indices. The Nifty50 of the National Stock Exchange is another prominent index (NSE). The National Stock Exchange (NSE) was founded in November 1992. The Nifty50 was first introduced on November 3, 1995, with a base value of 1000 rupees. It is made up of 50 equities that are ranked based on market capitalization and liquidity. The NSE has a total of 23 indicators.

In addition, sectoral indices of the Bombay stock exchange that are - S&P BSE Basic Materials, S&P BSE Energy, S&P BSE Fast Moving Consumer Goods, S&P BSE Finance, S&P BSE Healthcare, S&P BSE Industrials, S&P BSE Information Technology, S&P BSE Telecom, S&P BSE Utilities, S&P BSE AUTO, S&P BSE BANKEX, S&P BSE Capital goods, S&P BSE Consumer durables, S&P BSE Metal, S&P BSE OIL & GAS, S&P BSE Power, S&P BSE Realty, S&P BSE TECK are used to investigate the objectives. Further, the indices of National stock exchange that are included in the study have been selected from the sectoral as well as thematic indices. These are Nifty financial services index, Nifty media index, Nifty Pharma index, Nifty Private bank index, Nifty Public bank index from the sectoral index and Nifty commodities index, Nifty manufacturing index, Nifty India consumption index, Nifty infrastructure index and Nifty service sector index.

The research is based on secondary data. Secondary data is usually published in some manner (printed, electronic CD, or the Internet). Some government or research body has already gathered these data. The study's stock price data was gathered from www.bseindia.com and www.nseindia.com.

The study uses data on daily closing prices of BSE and NSE benchmark indices as well as the selected sectoral and thematic indices of India, from January 2006 to August 2020. All the 19 sectors of the Bombay stock exchange are included in the study followed by the sectors of National stock exchange (excluding the ones already included from BSE sectoral indices) and five thematic indices selected on the basis of their importance and performance also considering the availability of the data.

5.2 METHODOLOGY

5.2.1 Exploratory data analysis

1. Rate of return

The following formula was used to compute the rate of return of the indices as well as the sectors:

$$r_t = \frac{p_t - p_{t-1}}{p_{t-1}} \dots (5.1)$$

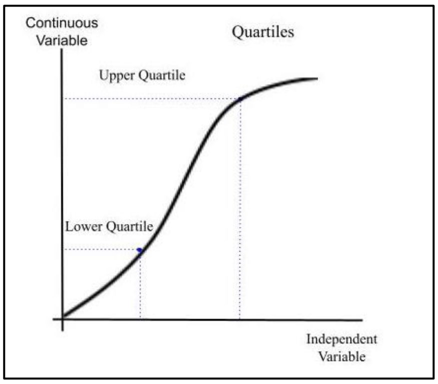
Where, r_t is the rate of return of the index/sector at time t, p_t price of the index/sector at time t, p_{t-1} price of the index/sector at the time t-1.

2. Descriptive statistics

The descriptive analysis includes the statistic values such as minimum, maximum, mean, median, standard deviation, skewness, kurtosis.

The minimum and the maximum values are selected from the data set as per the minimum and maximum values of prices as well as returns calculated. The first which is also known as the lower quartile basically separates the lowest 25% of the data from the highest 75%. The third quartile or the upper quartile distinguishes between the highest 25% of data from the lowest 75%. The median, which is the second quartile divides the number into two equal parts.

Figure 5.1: Quartiles



Source: Mathematical statistics by Gupta & Kapoor

When the data set is arranged in ascending order the quartiles are represented as,

1. First quartile Q_1 = [(n+1)/4]th term
2. Second quartile Q_2 = [(n+1)/2]th term
3. Third quartile Q_3 = [3(n+1)/4]th term

Mean, which is the average of the numbers is calculated as follows:

$$\mu = \frac{sum\ of\ the\ observations}{number\ of\ observation} \ldots (5.2)$$

here, μ is the mean value of the observations. Similarly, standard deviation is calculated using the below formula:

$$\sigma = \sqrt{\frac{\sum(x_i - \mu)^2}{N}} \ldots (5.3)$$

here, σ is the population standard deviation, N is the number of observations, x_i is each value from the population and μ if the population mean.

Further, to categorize the data as per its location, variability and shape the descriptive analysis includes skewness and kurtosis. Skewness measures lack of symmetry, while kurtosis measures whether the data are heavy-tailed or light-tailed as compared to the normal distribution. The histogram is an effective graphical technique to represents skewness and kurtosis of the data set.

In case of univariate data $X_1, X_2, \ldots, X_N$, the formula for skewness is:

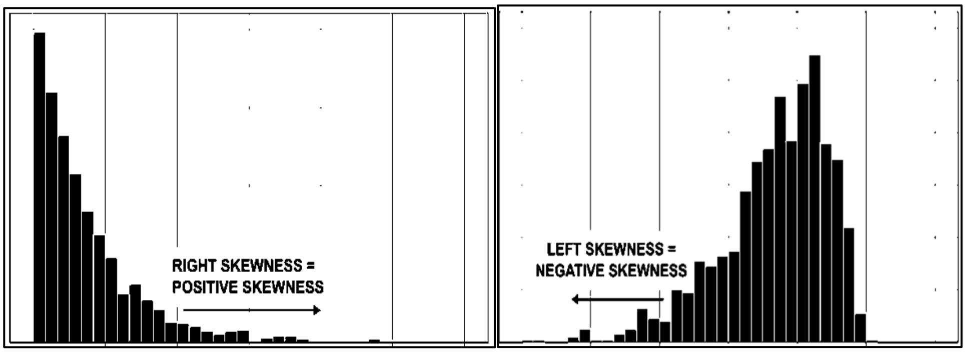
$$Skewness = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^3 / N}{s^3} \ldots (5.4)$$

Where, $\bar{X}$ is the mean, s is the standard deviation, X_i is each value from the population, and N is the number of observations. This formula is the Fischer-Pearson coefficient of skewness used by many software programs. Similarly, for univariate data $X_1, X_2, \ldots, X_N$, the formula for kurtosis is:

$$Kurtosis = \frac{\sum_{i=1}^{N} (X_i - \bar{X})^4 / N}{s^4} \ldots (5.5)$$

Where, $\bar{X}$ is the mean, s is the standard deviation, X_i is each value from the population, and N is the number of observations. A standard normal distribution has skewness of zero and kurtosis of three. The case of positive and negative skewness is shown in the graphs below:

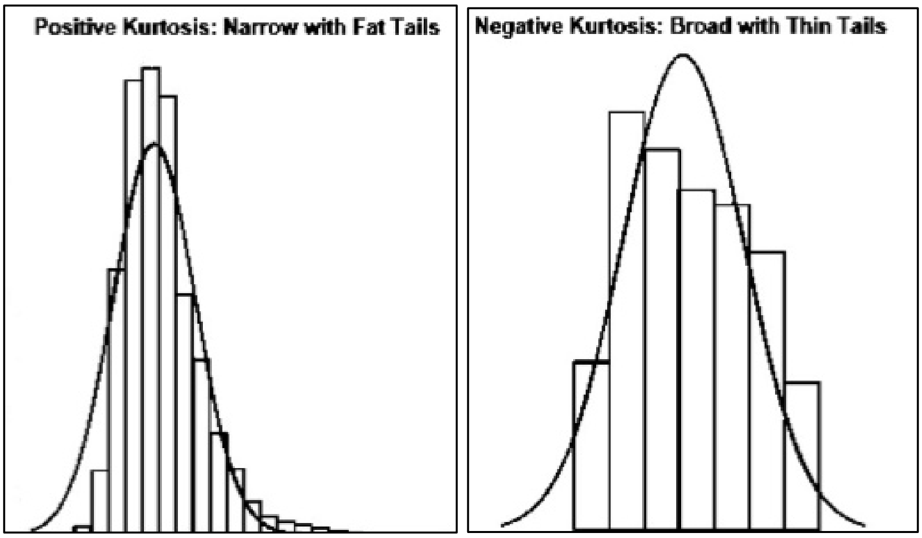
Figure 5.2: Different kinds of Skewness



Source: Mathematical statistics by Gupta & Kapoor

Histogram of positive and negative kurtosis looks something like this:

Figure 5.3: Different kinds of Kurtosis



Source: Mathematical statistics by Gupta & Kapoor

3. Beta values

The beta value measures the fluctuations in the stock prices to the overall changes in the stock market. It is a basic measure of volatility, a shot-term risk used in the CAPM model. In the thesis, the beta values of every sectoral indices as well as thematic indices are calculated with respect to the market. The formula used for the same is as follows:

$$Beta\ coefficient\ (\beta) = \frac{Covariance\ (R_e, R_m)}{Variance(R_m)} \ldots (5.6)$$

Where, R_e is the return on an individual stock (in this case, sectoral/thematic index). R_m is the return of the overall market. The beta values of sectoral indices of Bombay stock exchange with respect to the S&P BSE Sensex has been calculated. Similarly, beta values of sectoral and thematic indices are calculated with respect to Nifty 50 index. The beta values are very insightfully for investors to understand the movement of the sectoral/thematic index in comparison to the market. If the beta value is equal to zero, it indicated the sectoral/thematic index are strongly correlated with the market. If the value is greater than zero, it indicates that the sectoral/thematic index is more volatile than the market. Similarly, if the value is less than zero, it shows that the sectoral/thematic index is less volatile than the market.

4. Correlation

The correlation coefficient depicts relationship between movements of any two variables. The values range between -1 (negative correlation) to +1 (positive correlation). This statistic can be of great use to the investors in adjusting their portfolios. It can also be used for diversification of the portfolio. There are a number of ways to calculate correlation coefficient but the most common is the Pearson's coefficient (r)(Pearson,1895). The formula for calculating correlation coefficient is as follows:

$$\rho_{xy} = \frac{Cov(x, y)}{\sigma_x \sigma_y} \ldots (5.7)$$

Here, ρ_{xy} represents correlation between x and y variables, $\sigma_x \sigma_y$ are standard deviation of x and y. In the study, the correlation of all the sectoral and thematic indices with the respective market index is being calculated.

Accordingly, the formula employed in the study is as follows:

$$\rho_{sm} = \frac{Cov(sect\&thema\ index, market\ index)}{\sigma_{sect\ \&\ themaindex}\sigma_{marketindex}} \ldots (5.8)$$

Where, ρ_{sm} represents correlation between sectoral/thematic index and market index, $\sigma_x \sigma_y$ are standard deviation of sectoral/thematic index and market index. The correlation coefficient of both Bombay as well as National stock exchange is measured with respect to their sectoral/thematic index.

5. Jarque-Bera test of normality

Jarque-Bera Test, is a test for normality, which is a form of Lagrange multiplier test (Jarque & Bera, 1980). Many statistical analyses assume normality; the Jarque-Bera test is commonly used before one of these tests to validate normality. Because other normality tests are unreliable when n is big, it is typically employed for huge data sets. The test compares the skewness and kurtosis of data to check if they are similar to a normal distribution. A normal distribution has a skewness of zero (i.e. it's completely symmetrical around the mean) and a kurtosis of three, which indicates how much data is in the tails and how "peaked" the distribution is.

The Jarque-Bera test statistic (also known as the JB test statistic) has the following formula:

$$JB = \frac{n}{6}\left(S^2 + \frac{1}{4}(K-3)^2\right) \ldots (5.9)$$

Where, n is sample size, S is the sample skewness coefficient, K is the kurtosis coefficient.

The test's null hypothesis is that the data is normally distributed; the alternative hypothesis is that the data is not normally distributed.

5.2.2 Granger causality

Granger causality is a statistical hypothesis testing which checks whether a time series can be used to forecast the other. It is very close to determining the cause effect relation although not in exact way. It basically tells us whether a time series is caused by the other. This relationship can be one way and two way. If time series A is caused by time series B, and time series B is also caused by time series A it is called a bi-direction relation. Whereas, if time series A is caused by time series B but time series B is not caused by time series A then it is called a uni-directional relationship. The null hypothesis of the test states that time series A does not cause time series B. If the probability value is less than five percent level of significance, than the hypothesis will be rejected.

The causality test of each sectoral and thematic indices is carried out with respect to the market index. It is also run amongst itself, the results of which is presented in the appendix. To run the granger causality, the following bivariate regression is carried out:

$$b_t = \gamma_0 + \gamma_1 b_{t-1} + \cdots + \gamma_1 b_{t-1} + \delta_1 a_{t-1} + \cdots + \ \delta_1 a_{-1} + \varepsilon_t \ldots (5.10)$$

$$a_t = \gamma_0 + \gamma_1 a_{t-1} + \cdots + \gamma_1 a_{t-1} + \delta_1 b_{t-1} + \cdots + \ \delta_1 b_{-1} + u_t \ldots (5.11)$$

$$(\delta_1 = \delta_2 = \cdots = \delta_t = 0)$$

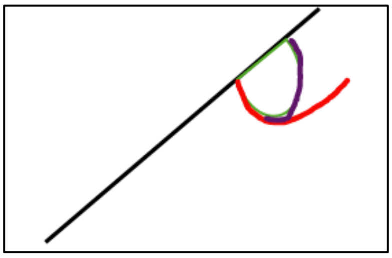
Here, a and b are time series variables, a represents the various sectoral and thematic indices and b represents the market index. The null hypothesis states that a does not granger-cause b in the equation 12 and that b does not granger-cause a in the equation 13.

5.2.3 ARCH and GARCH models

1. Unit root test

A unit root is a stochastic trend present in a time series, which is also called a “random walk with drift”. If a time series has a unit root, it depicts a systematic pattern which is unpredictable in nature.

Figure 5.4: Unit root



Source: Basic econometrics by Gujarati

The red line represents the fall in output and path of recovery if the timeseries has a unit root. Blue, on the other side, shows the recovery if there is no unit root and the series is stationary. The presence of unit root can cause the analysis to have a spurious regression and errant behavior in the distributions. So, unit root is a test for stationarity in a time series model. A time series is stationary only if a change in time has no significant impact in the shape of the distribution.

There are a number of tests to check for stationarity, but this study makes use of the two most significant tests – Dickey Fuller test and Phillips-Perron (PP) test.

Augmented Dickey Fuller Test- This test was devised by David Dickey and Wayne Fuller in 1979 (Dickey & Fuller, 1979). It is based on linear regression and handles bigger, more complex models. The ADF test, when applied to a return series rt, regresses the first difference of the series against the series lagged k times, as illustrated below:

$$r_t = \rho r_{t-1} + x_t{}'\delta + \epsilon_t \ldots (5.12)$$

Where, x_t is optional regressors which may consist of constant, or a constant and trend, ρ and δ are parameters to be estimated, while ϵ_t is the white noise.

Null hypothesis (H_0) = The series contains a unit root.
Alternative hypothesis (H_1) = The series is stationary.

Acceptance of the null hypothesis implies that the dataset is non-stationary. The acceptance is based on the values of the probability, i.e., if the value is lesser than 5% confidence interval than the null hypothesis is accepted. Alternatively, if the probability value is greater than 5% confidence interval than the null hypothesis is rejected and the alternative hypothesis is accepted stating that the series is stationary. The ADF test is commonly used in most of the timeseries model; however, it cannot be applied in case of serial correlation.

Phillips-Perron Test – This test is a modification of the Augmented dickey fuller test and corrects the errors for autocorrelation and heteroscedasticity (Phillips & Perron, 1988). The Phillips-Perron test is calculated on the basis of following formula:

$$\tau_\alpha = t_\alpha (\frac{\gamma_0}{f_0})^{1/2} - \frac{T(f_0 - \gamma_0)\big(s\epsilon(\hat{\alpha})\big)}{2f_0^{1/2}s} \ldots (5.13)$$

Where $\hat{\alpha}$ is the estimate, and τ_α the t-ratio of α. $\big(s\epsilon(\hat{\alpha})\big)$ is coefficient of standard error, s is the standard error of the test regression. γ_0 is the consistent estimate of the error variance and f_0 is an estimator of the residual spectrum at frequency zero.

2. Autoregressive conditional heteroscedasticity

A simple autoregressive (AR) model is denoted as:

$$r_t = \theta_0 + \theta_1 r_{t-1} + a_t \dots (5.14)$$

Here, time series r_t is the dependent variable and r_{t-1} is the explanatory variable, a_t is the white noise series with zero mean and variance. In the literature of the time series, equation 16 is represented as an autoregressive model of order 1 or AR (1) model.

Similarly, an AR(p) model can be denoted as:

$$r_t = \theta_0 + \theta_1 r_{t-1} + \cdots + \theta_p r_{t-p} + a_t \dots (5.15)$$

Here, p is assumed to be a nonnegative integer and a_t is the white noise series with zero means and variance. The model states that the past p variables r_{t-1} (i=1,…, p) jointly determine the conditional expectation of r_t. Autoregressive indicates that heteroscedasticity observed over different time periods may be autocorrelated. Conditional indicates that variance is based on past errors; while, heteroscedasticity means that the series has unequal variance. Thus, forming an Autoregressive conditional heteroscedasticity (ARCH) model. It simply indicates that the series has a time-varying variance that depends on lagged effects. Engle (1982) introduced the ARCH models. It assumes that the variance of the current error term is akin to the size of previous period error terms.

ARCH models the attitude of investors not only towards expected returns but also risk. It can be used to form an economic forecast and also to measure volatility. For instance, inflation in itself may not be bad but its variability is bad which makes financial planning difficult. Similarly, for importers, exporters, traders in the foreign exchange markets, face variability in the exchange rates resulting in either huge losses or profits. In case of stock markets, investors are interested in the volatility in stock prices. A high volatility could mean huge losses or gains. Hence, it becomes important to model these variabilities in order to understand the existing patterns and forecast the future.
Considering the following model:

$$Y_t = c + u_t \dots (5.16)$$
$$u_t \sim iid\ N(0, \sigma_t^2) \dots (5.17)$$

$$\sigma_t^2 = b_0 \ldots (5.19)$$

This model is a simple regression model containing a constant term and no explanatory variables. Equation 18 is the mean equation, where Y_t dependent variable is equal to the sum of constant and white noise error term, u_t. The error term is normally distributed with mean 0 and variance σ_t^2. The variance σ_t^2 is a constant i.e., b_0 which changes over time.

Further making three major adjustments in the model:

1. Allowing the error variance to be time-varying, which is, heteroscedastic h_t.

$$\sigma_t^2 = h_t \ldots (5.20)$$

2. Allowing the distribution of error term to be conditionally normal representing information available at time t-1.

$$u_t[I_{t-1} \sim N(0, \sigma_t^2) \ldots (5.21)$$

3. Allowing h_t to be a function of a constant term and lagged squared error, u_{t-1}^2.

$$h_t = b_0 + b_1 u_{t-1}^2 \ldots (5.21)$$

$$\sigma_t^2 = b_0 + b_1 u_{t-1}^2 \ldots (5.22)$$

$$b_0 > 0, 0 \leq b_1 < 1 \ldots (5.23)$$

Equation 19 is the variance equation of the ARCH (1) process.

For ARCH(q) model the variance equation would be as follows:

$$h_t = b_0 + \sum_{i=1}^{q} b_1 u_{t-i}^2 \ldots (5.24)$$

Therefore, in totality the ARCH(1) model would be as follows:

$$a_t = \varepsilon_t \sqrt{\alpha + \alpha_1 a_{t-1}^2} \dots (5.25)$$

Here, a_t is a time series,

ε_t is the white noise or error term,

α is a constant and

a_{t-1}^2 is the value of the time series yesterday.

$\sqrt{\alpha + \alpha_1 a_{t-1}^2}$ is the volatility of time series a_t.

Now, the problem with the ARCH model is that it is quite "bursty". This means it has burst of volatility rather than persistence of volatility. That is whey generalized ARCH model comes into the picture.

3. Generalized autoregressive conditional heteroscedasticity

Constant variance has been one of the most classical assumptions of conventional time series and econometric models. However, that changed after the introduction of ARCH (Autoregressive Conditional Heteroskedastic) process introduced by Engle (1982) which emphasized on the connection between conditional and the unconditional variance. The conditional variance was allowed to change over time as a function of past errors in the ARCH model. It was the first model that provided a systematic method to model volatility and was used widely by financial practitioners, academicians and policy makers.

After an extensive use of the model, people found out a few weaknesses in the model such as long lag length, large number of parameters and it is not easy to control the existence of negative variance. Thus, in order to overcome these drawbacks, Bollerslev (1986) proposed the generalized ARCG, GARCH (Generalized Autoregressve Conditional Heteroscedasticity) model.

The GARCH(p,q) specification may be written as follows:

$$\sigma_t^2 = \beta_0 + \sum_{i=1}^{q} \beta_1 \varepsilon_{t-i}^2 + \sum_{j=1}^{p} \beta_2 \sigma_{t-j}^2 \ldots (5.26)$$

Where, $\beta_0 \beta_1$>0 and (β_1+β_0) < 1, σ_t^2 represents the conditional variance, ε_{t-i}^2 is the white noise. The conditional variance is a linear mixture of the conditional return's equation's q lags of the squared residuals and the conditional variance's p lags.

Considering one lag of σ_t^2 and ε_t^2 i.e., $\sigma_{t-1}^2, \varepsilon_{t-1}^2$, then the estimation of the conditional variance would be:

$$\sigma_t^2 = \beta_0 + \beta_1 \varepsilon_{t-1}^2 + \gamma_1 \sigma_{t-1}^2 \ldots (5.27)$$

Here, σ_t^2 is volatility at time t, σ_{t-1}^2 is volatility of previous period and ε_{t-1}^2 is the error of previous period. β_0 is a constant, β_1 and $\gamma_1$are coefficients.

A range of GARCH(p,q) models with different values of p and q along with different distributions of the error term will be estimated on the basis of the information criterion that will be discussed.

a. Exponential GARCH

The Exponential GARCH or EGARCH model was proposed by Nelson in 1991, based on log transformation of conditional variance. In this model the conditional variance always remains positive because of natural logarithm of the dependent variable. This model enables to capture the asymmetries present in the error term caused as a result of negative and positive news. The ARCH term is divided into two different independent variables. The first variable indicates the sign effect of shocks on index volatility and the second variable indicates the size or magnitude effect of shocks on volatility in the financial time series. The specification of EGARCH model used to study the conditional variance as in Brooks (2014) is as below:

$$\ln(\sigma_t^2) = \omega + \beta \ln(\sigma_{t-1}^2) + \gamma \frac{u_{t-1}}{\sqrt{\sigma_{t-1}^2}} + \alpha \left[\frac{|u_{t-1}|}{\sqrt{\sigma_{t-1}^2}} - \sqrt{\frac{2}{\pi}} \ldots (5.28)\right.$$

The model has several advantages over the model presented in equation 23. Firstly, the $\ln(\sigma_t^2)$ is modeled, so that none of the parameters will be negative. Thus, there is no need to impose non-negative constraints in the model and since asymmetries are allowed, if the relation between volatility and returns are negative, γ will be negative. The second term in the model represents the impact of GARCH term on the future conditional variance and the third term indicates the sign effect of ARCH on conditional volatility. The fourth term signals the size effect of ARCH on conditional volatility. The EGARCH will be tested with skewed distribution.

b. GJR GARCH

The GJR-GARCH is named after Glosten, Jagannathan and Runkle (1993) which models the positive and negative shocks on the conditional variance asymmetrically with the help of an indicator function I. The specification of the GJR GARCH model as in Brooks (2014) is as follows:

$$\sigma_t^2 = (\omega + \sum_{j=1}^{m} \zeta_j v_{jt}) + \sum_{j=1}^{q} (\alpha_j \varepsilon_{t-j}^2 + \gamma_j I_{t-j} \varepsilon_{t-j}^2) + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^2 \ldots (5.29)$$

Here, γ_j represents the leverage term, the value of indicator function is 1 for $\varepsilon \leq 0$ and 0 alternatively. In presence of the indicator function, persistence of the model relies on the asymmetry of conditional distribution and is represented as:

$$\hat{P} = \sum_{j=1}^{q} \alpha_j + \sum_{j=1}^{p} \beta_j + \sum_{j=1}^{q} \gamma_j \, k \ldots (5.30)$$

Here, k is the expected value of standardized residuals z_t.

$$k = E[I_{t-j} z_{t-j}^2] \ldots (5.31)$$

If the distributions are symetric, the value of k is 0.5. The ARCH(q) parameters are α_j , leverage (q) parameters are γ_j and GARCH (p) parameters are β_j. Finally, the variance intercept parameter is omega.

c. ASYMMETRIC – POWER GARCH

Ding, Granger and Engle (1993) while investigating the 'long memory' property of stock returns found that absolute returns are not the only ones having a higher correlation compared to normal returns, but the power transformation of the absolute returns also result in higher autocorrelation for long lags. The ARCH models are based on squared returns, however Ding et. al (1993) found that the autocorrelation function for a fixed lag has a unique maximal point which makes linear relationship among absolute returns neither a necessary or efficient property of ARCH specification. This gave rise to Asymmetric – Power GARH put forward by Ding et. al (1993). The specification of the APARCH model is as follows:

$$\sigma_t^\delta = (\omega + \sum_{j=1}^{m} \zeta_j v_{jt}) + \sum_{j=1}^{q} \alpha_j (|\varepsilon_{t-j} - \gamma_j \varepsilon_{t-j}|)^\delta + \sum_{j=1}^{p} \beta_j \sigma_{t-j}^\delta \ldots (5.32)$$

The σ_t and γ_j are the leverage term.

The persistence model is given by:

$$\hat{P} = \sum_{j=1}^{p} \beta_j + \sum_{j=1}^{q} \alpha_j \, k_j \ldots (5.33)$$

Here, k_j is the standardized residual's (z_t) expected value.

$$k_j = E\ (|z| - \gamma_j z)^\delta \ldots (5.34)$$

The ARCH parameters are the α_j, leverage parameters are γ_j, power parameter is δ. The GARCH parameters are β_j and variance is omega.

d. Univariate fractional integrated volatility model

The long memory attribute can be explained using the properties of autocorrelation function, which are expressed as $\rho_k = \frac{cov(x_t - x_{t-1})}{var(x_t)}$ for integer lag k. Autocorrelation such that $\lim_{k\to\infty} \rho_k = 0$ is expected to be present in a covariance stationary time series process. For ARMA (p,q) process, the autocorrelation decay at relatively fast exponential rate so that $\rho_k \approx |m|^k$, $|m| < 1$ in case of stationary and invertible time series processes. While, in case of long memory processes, the autocorrelation decay at an hyperbolic rate, $\rho_k \approx\ c_1 k^{2d-1}$ as k increases without limit, where $c_1$is a constant and d is the long memory parameter.

In GARCH (1,1) process the estimated sum of parameters α_1 and β_1, if close to unity exhibits strong persistence. However, if the sum of the parameters is less than 1 then the ε_t process is second order stationary and a shock to conditional variance h_t has a decaying impact on h_{t+h}, when h is increasing and is asymptotically negligible. The influence on h_{t+h} does not die out asymptotically when the sum of parameters is higher than or equal to 1.

Engle and Bollerslev (1986) developed the IGARCH model under the restriction of α_1 + β_1 = 1 stating that current information is of importance while forecasting the volatility for all horizons.

$$\emptyset(L)(1-L)\varepsilon_t^2 =\ \omega + [1-\beta(L)]\upsilon_t \ldots (5.35)$$

Where, $\upsilon_t = \varepsilon_t^2 - \sigma_t^2$ is the innovation in the conditional variance mechanism, has mean 0 and absence of serial correlation. However, volatility changes over time, and the consequences of a shock might take a long time to decay as demonstrated by Ding et. al. (1993). As a result, it appears that the distinction between I(0) and I(1) processes is overly rigid. Moreover, shocks die out at an exponential pace in an I(0) process (capturing only short memory), while shock persistence is infinity in an I(1) process with no mean reversion, whilst 0<d<1 shocks die out at a slow hyperbolic pace.

Therefore, to determine the long memory effect in volatility, Baillie et al. (1996) presented the fractionally integrated GARCH model (FIGARCH), which permits a hyperbolic decay of the coefficient, β_j, that is positive, summable and fulfils the unit root requirement. Chung(1999) identified and argued that Baillie's method of parametrization of FIGARCH model may have been specification problem resulting in difficult interpretation of the estimated parameters. The fractional differencing operator applies to constant term in the mean equation and not in variance equation. Bollerslev and Mikkelsen (1996) expand the FIGARCH process to FIEGARCH in order to handle asymmetries between positive and negative shocks, known as the leverage effect. This corresponds to Nelson's (1991) exponential GARCH model, which allows for asymmetry.

The FIEGARCH (p,d,q) model is given as,

$$\ln(h_t) = \omega + \emptyset(L)^{-1}(1-L)^{-d}[1+\alpha(L)]g(z_{t-1}) \ldots (5.36)$$
$$g(z_{t-1}) = \theta Z_t + \gamma[|z_t| - E|z_t| \ldots (5.37)$$

θZ_t is the sign effect and $\gamma[|z_t| - E|z_t|$ is the magnitude effect. The roots of $\emptyset(L)$ and $\alpha(L)$

In the lag operator L, the roots are auto regressive polynomial and moving average polynomial. In case of d=0, the FIEGARCH model reduces to EGARCH proposed by Nelsen (1991), while, when d=1 the process becomes an integrated EGARCH process (IEGARCH). Bollerslev and Mikkelsen (1996) provided evidence for the effectiveness of QMLE when used to estimate the FIEGARCH process parameters.

5.2.4 Copula models

Copula, a joint distribution function, is made up by two marginal distributions, say, f(x) and f(y) integrated into a copula function C(f(x),f(y)|¥), where ¥ is a set of data that is available at t-1point in time.
F(x) denotes the marginal distribution of a variable x, while F(y) denotes the marginal distribution of a variable y. If the joint distribution of x and y is G(x,y), then the copula function C for every (x,y) is as follows:

$$C\big(F(x), F(y)\big) = G(x.y) \ldots (5.38)$$

In financial time series data, however, changes in a variable is always dependent on changes in others. Therefore, a conditional parameter is used to include the conditionalities, where $F_{1|\Omega}$ and $F_{2|\Omega}$ signifies the marginal distribution of x and y, dependent on Ω. As a result, the copula function is stated as follows:

$$F_{1,2|\Omega}(x, y|\Omega) = C\big(\, F_{1|\Omega}(x|\Omega), F_{2|\Omega}(y|\Omega)\big|\Omega\big) \ldots (5.39)$$

This bivariate distribution does not have to be normal; it can be represented by several copulas from the copula family. Gaussian copula and Student t distributions, both of which are classified as elliptical copulas, are two of the most well-known copulas. In addition, the research uses three prominent Archimedean copulas that are Gumbel copula. Clayton copula, and Frank copula. All of these copulas represent various sorts of tail dependence, allowing for the simulation of a wide range of dependent structures.

1. ELLIPTICAL - GAUSSIAN COPULA

The Gaussian copula, whose distribution is symmetric, is the simplest fundamental copula and is represented as follows:

$$\emptyset_p\left(\emptyset^{-1}(x), \emptyset^{-1}(y)\right) = \int_{-\infty}^{\emptyset^{-1}(x)} \int_{-\infty}^{\emptyset^{-1}(y)} \frac{1}{2\pi(1-\rho^2)} e \int \{-\frac{s^2 - 2\rho st + t^2}{2(1-\rho)}\} dsdt \ldots (5.40)$$

Here, $\emptyset^{-1}$ = inversion of the standard normal distribution's cumulative distribution function (cdf)

$\emptyset_p$ is the cumulative distribution function of a two variate normal distribution

ρ is the Pearson correlation.

Tail dependence (left/right) does not exist in Gaussian copulas.

2. ELLIPTICAL - STUDENT T COPULA

The student t-copula is represented as below:

$$C_d^T(u; v; \Sigma) = \int_{-\infty}^{t_v^{-1}(u_1)} \cdots \int_{-\infty}^{t_v^{-1}(u_d)} \frac{\tau\left(\frac{v+d}{2}\right)}{\tau\left(\frac{v}{2}\right)\sqrt{(\pi v)^d|\Sigma|}} \left(1 + \frac{x'\Sigma^{-1}x'}{v}\right)^{-\frac{v+d}{2}} dx \ldots (5.41)$$

Here, t_v^{-1} is the inverted cumulative distributed function of a standard Student's t-distribution with v degrees of freedom, $\Sigma(\rho_{ij})_{i,j=1,\ldots,d}$ is a dispersion matrix. The Student t distribution is symmetric and a bell shaped quite similar to gaussian; however, it has heavier tails and is prone to producing values that fall far from its mean.

3. ARCHIMEDEAN - CLAYTON COPULA

Archimedean copulas allow modelling dependence structures of distributions in arbitrarily high dimensions with the help of only one parameter. These copulas can be defined as:

$$C(u_1, \ldots \ldots, u_d; \theta) = F(F^{-1}(u_1; \theta) + \ldots + F^{-1}(u_d; \theta); \emptyset) \ldots (5.42)$$

Here, F is the generator function of the copula. It is a continuous function but strictly decreasing on [0, $F^{-1}(0)$]. F^{-1} is the pseudo inverse of the function F.

A table on the generator function and its inverse along with the parameters is represented below:

Table 5.1: Generator functions and parameter range of Archimedean Copulas

COPULA	**GENERATOR FUNCTION (∅)**	**GENERATOR FUNCTION INVERSE ($\emptyset^{-1}$)**	**PARAMETER RANGE**
CLAYTON	$(1+\theta t)^{-1/\theta}$	$\frac{1}{\theta}(t^{-\theta}-1)$	$\theta \in [-1,\infty)\{0\}$
GUMBEL	$\frac{1-\theta}{\exp(t)-\theta}$	$\log\left(\frac{1-\theta(1-t)}{t}\right)$	$\theta \in [-1,1)$
FRANK	$\frac{1}{\theta}\log\left(1+\exp(-t)\left(\exp(-\theta)-1\right)\right)$	$-\log\left(\frac{exp(-\theta t)-1}{exp(-\theta)-1}\right)$	$\theta \in R\{0\}$

Source: Copula methods in finance by Cherubini, Luciano & Vecchiato

Embrechts et al. (2005) suggested the use of Clayton copula which incorporates the lower tail dependence and Gumbel copula that incorporates the upper tail dependencies.

The clayton copula is represented as below with dependence parameter $\theta \in (0,\infty)$,

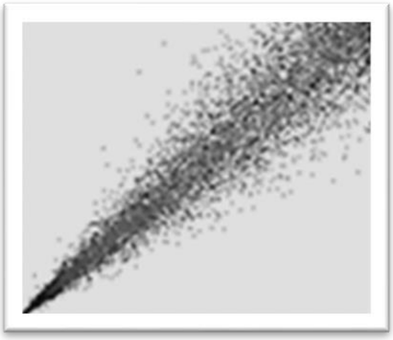
$$\max\left[(x^{-\alpha}+y^{-\alpha}-1)^{-\frac{1}{\alpha}},0\right] \ldots (5.43)$$

Here, α is the parameter of a copula,

The variables x and y have marginal distributions of f(x) and f(y), respectively.

The lower tail dependence (λ_1) is transformed as $2^{-\frac{1}{\alpha}}$ (when, $\alpha = 0$ signifies absence of tail dependence). Below is the diagrammatic representation of the Clayton copula.

Figure 5.5: Clayton copula



Source: Copula methods in finance by Cherubini, Luciano & Vecchiato

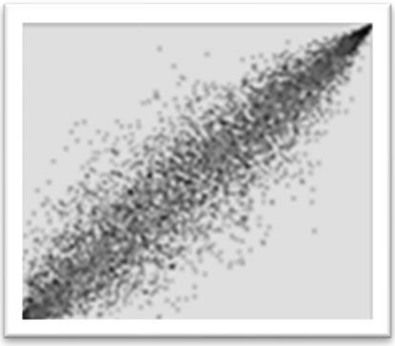
4. ARCHIMEDEAN - GUMBEL COPULA

The Gumbel copula is represented as below with dependence parameter as $\theta \in [0, \infty)$.

$$C(x, y|\Omega) = \exp\left\{-[(-\ln x)^{\theta} + (-\ln y)^{\theta}]^{\frac{1}{\theta}}\right\} \ldots (5.44)$$

Here, exp - exponential function, ln - natural logarithm. Maximum dependencies are achieved when $\theta \rightarrow \infty$, while upper tail dependence is generated when $\theta > 1$. The lower tail dependence (λ_u) is normalized as $2 - 2^{\frac{1}{\theta}}$. Shown below is the diagrammatic representation of the Gumbel copula:

Figure 5.6: Gumbel copula



Source: Copula methods in finance by Cherubini, Luciano & Vecchiato

5. ARCHIMEDEAN - FRANK COPULA

The Frank copula with parameter $\theta \in R\{0\}$ is represented as below:

$$C_\alpha(u,v) = -\frac{1}{\alpha}\ln\left(1 + \frac{(e^{-\alpha u}-1)(e^{-\alpha v}-1)}{e^{-\alpha}-1}\right) \dots (5.45)$$

Here, ln is the natural logarithm and $\alpha \in (-\infty, \infty)/\{0\}$. The following figure shows a diagrammatic representation of a frank copula.

Figure 5.7: Frank copula

Source: Copula methods in finance by Cherubini, Luciano & Vecchiato

Conclusion –

Overall, this chapter discusses various methodologies used for research in the study starting from exploratory data analysis to calculating the beta values, correlation and normality. Further, causality test, i.e., Granger causality is carried out followed by the ARCH and GARCH family models such as exponential GARCH, GJR GARCH, asymmetric power GARCH, univariate fractionally integrated volatility model – FIEGARCH. Further, in order to study the associations between sectors copula models are used – elliptical and archemedian copulas.