Pragya

the best FRM revision course!

FRM 2017
Part 1
Book 2 – Quantitative Analysis
Pragya
the best FRM revision course!
1. **Definitions:**
   a. Random Variable: An uncertain quantity or number
   b. Outcome: An observed value of a random variable
   c. Event: A single or set of outcomes
   d. Mutually Exclusive: Event that both cannot happen at the same time
   e. Exhaustive Events: Events that include all possible outcomes

2. **Discrete Random Variable (DRV):** Random variable for which the number of possible outcomes can be counted and for each possible outcome there is a measurable and positive probability e.g. Money

3. **Continuous Random Variable (CRV):** Random variable for which number of possible outcomes is infinite. The probability of any single value is zero. E.g. Time

4. **Joint Probability:** Probability that two events will both occur. It is denoted as \( P(AB) \) i.e. Probability of event A and B and is calculated as \( P(AB) = P(A/B) \times P(B) \) where \( P(A/B) \) is read as probability of A given B and is called Conditional Probability

5. **Addition Rule:** \( P(A \text{ or } B) = P(A) + P(B) - P(AB) \) Note: The word “and” indicates multiplication i.e. Joint probability and “or” indicates addition i.e. addition rule

6. **Bayes Theorem:** \( P(A/B) = \frac{P(B/A) \times P(A)}{P(B)} \)

7. **Probability Density Function (PDF):** Denoted as \( f(x) \), it is used to generate probability that outcomes of a continuous variable lie within a particular range of outcomes. Remember that in a continuous distribution, probability of any particular outcome is zero.

8. **Cumulative Density Function (CDF):** Defines a random variable \( x \) that takes a value less than or equal to a specific value.

9. **Discrete Uniform Random variable:** Probabilities of all variables is equal
BASIC STATISTICS


1. Definitions:
   a. Variance: Variance is defined as the expected value of the difference between the variable and its mean squared. \( \sigma^2 = \text{E} [(X - \mu)^2] \)
   b. Standard Deviation: Square root of variance is called Standard Deviation (\( \sigma \))
   c. Return of portfolio with 2 assets: Average value of return e.g. if there are two portfolios with returns as \( r_1 \) and \( r_2 \) with weights as \( w_1 \) and \( w_2 \), then mean return of portfolio is \( r = (w_1 \times r_1) + (w_2 \times r_2) \). Note that \( w_1 + w_2 = 1 \) always
   d. Variance of Portfolio with 2 assets: \( \sigma_{1,2}^2 = (w_1 \times \sigma_1)^2 + (w_2 \times \sigma_2)^2 + 2\rho_{1,2}w_1w_2\sigma_1\sigma_2 \)
   e. Covariance: Covariance is analogous to variance, but instead of looking at the deviation from the mean of one variable, we are going to look at the relationship between the deviations of two variables.
      i. \( \text{Cov}_{x,y} = \rho_{x,y}\sigma_x\sigma_y \)
   f. Correlation is \( \rho \) in the above equation. Correlation has the nice property that it varies between −1 and +1

2. Skewness: Skew refers to the extent to which the distribution are not symmetrical.

3. Kurtosis: Degree to which a distribution is peaked than a normal distribution. Average Kurtosis for a normal curve is 3.
   a. Skewness is degree 3 and Kurtosis is degree 4.
   b. \( \text{Skewness} = \frac{\Sigma_{i=1}^{n}(X_i - \mu)^3}{\sigma^3n} \) and \( \text{Kurtosis} = \frac{\Sigma_{i=1}^{n}(X_i - \mu)^4}{\sigma^4n} \)
Reading: Distributions (Chapter 4, Michael Miller, Mathematics and Statistics for Financial Risk Management (Hoboken, NJ: John Wiley & Sons, 2013))

1. **Definitions:**
   a. Parametric Distribution: A parametric distribution can be described by a mathematical function. E.g. Normal Curve
   b. Non-parametric distribution: A nonparametric distribution cannot be summarized by a mathematical formula. In simplest form, a nonparametric distribution is just a collection of data. E.g. Historical data

2. **Uniform Distribution, Bernoulli Distribution, Binomial Distribution**

<table>
<thead>
<tr>
<th>Uniform Distribution</th>
<th>Bernoulli Distribution</th>
<th>Binomial Distribution</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="Uniform Distribution" /></td>
<td><img src="image" alt="Bernoulli Distribution" /></td>
<td>A binomial distribution can be thought of as a collection of Bernoulli random variables. If we have two independent bonds and the probability of default for both is 10%, then there are three possible outcomes: no bond defaults, one bond defaults, or both bonds default</td>
</tr>
<tr>
<td>Probability between points a and b is equal and is zero at every other point</td>
<td>A Bernoulli random variable is equal to either zero or one. E.g. coin toss. (\mu = P) and (\sigma = P(1-P))</td>
<td></td>
</tr>
</tbody>
</table>

3. **Poisson distribution**: The Poisson distribution is often used to model the occurrence of events over time. E.g. the number of bond defaults in a portfolio or the number of crashes in equity markets.
   a. \(P[X = n] = \frac{(\lambda t)^n}{n!} e^{-\lambda t}\) where \(n\) is the number of default events and \(\lambda\) is the constant rate of decay.

4. **Normal Distribution**: It is more common to refer to the normal distribution as the Gaussian distribution. The distribution is described by two parameters, \(\mu\) and \(\sigma\); \(\mu\) is the mean of the distribution and \(\sigma\) is the standard deviation. Because a linear combination of normal distributions is also normal, standard normal distributions are the building blocks of many financial models. When a normal distribution has a mean of zero and a standard deviation of one, it is referred to as a standard normal distribution. The skew of a normal distribution is always zero. The kurtosis of a normal distribution \(Z = \frac{(x-\mu)}{\sigma}\) is always 3.
5. **Lognormal Distribution**: If a variable has a lognormal distribution, then the log of that variable has a normal distribution. Unlike the normal distribution, which ranges from negative infinity to positive infinity, the lognormal distribution is undefined, or zero, for negative values. Using lognormal distribution, we avoid returns lower than -100%.

6. **Students –t Distribution, Chi Square, F-Test**

<table>
<thead>
<tr>
<th>Students T</th>
<th>Chi Square</th>
<th>F Test</th>
</tr>
</thead>
<tbody>
<tr>
<td><img src="image" alt="T Distribution" /></td>
<td><img src="image" alt="Chi Square" /></td>
<td><img src="image" alt="F Test" /></td>
</tr>
<tr>
<td>Used when n &lt; 30 in place of normal distribution</td>
<td>Used to test if Variance of a normally distributed population is equal to some value</td>
<td>Used to compare equality of 2 variances.</td>
</tr>
</tbody>
</table>

\[
Z = (x - \mu) \times \left(\frac{\sqrt{n}}{\sigma}\right) \quad \chi = \frac{\left[\left(n - 1\right)\sigma_{\text{Actual}}^2\right]}{\sigma_{\text{Hypo}}^2} \quad F = \frac{\sigma_1^2}{\sigma_2^2}
\]
**Bayesian Analysis**

Reading: Bayesian Analysis (Chapter 6, Michael Miller, Mathematics and Statistics for Financial Risk Management (Hoboken, NJ: John Wiley & Sons, 2013))

1. **Bayes Theorem**: Bayes theorem for two variables is defined as

\[ P(A/B) = \frac{P(B/A) \cdot P(A)}{P(B)} \]

   a. The numerator of the above is same \( P(AB) \) which is same as \( P(A|B) \cdot P(B) \)
   
   b. Unconditional probabilities are given as \( P(A) \) and \( P(B) \)
   
   c. Bayes’ theorem provides a framework for determining the probability of one random event occurring given that another random event has already occurred.
1. **Sample Mean and Sample Variance**: As \( n \) gets larger, the sample mean (\( \bar{x} \)) approaches the population mean (\( \mu \)).

The sample standard deviation is given by \( \sigma_{\text{sample}} = \frac{\sigma}{\sqrt{n}} \) and variance is \( \sigma^2 \).

2. **I.I.D (Independent and Identically Distributed Variables)**: If each random variable has the same probability distribution as the others and all are mutually independent, then they are said to be i.i.d. Observations in a sample are often assumed to be effectively i.i.d. for the purposes of statistical inference.

3. **Confidence Interval**: It is a range of values within which the actual value of the parameter will lie, given the probability of 1-\( \alpha \) where \( \alpha \) is called the significance level. The calculation of confidence interval is as follows:

\[ \bar{x} \pm (\text{Reliability factor} \times \text{Standard Error}) \]

where reliability factor is the Z value and Standard Error is \( \frac{\sigma}{\sqrt{n}} \)

Note: For the entire population, replace \( \bar{x} \) with \( \mu \) and \( \sigma_s \) with \( \sigma \)

4. **Hypothesis Testing**: It is a statistical assessment of a statement or idea regarding the population. A hypothesis is a statement about the value of a population parameter developed for the purpose of testing a theory or belief.

   a. **Null Hypothesis (H_0)**: Generally the hypothesis which the researcher wants to reject.

   b. **Alternate Hypothesis (H_A)**: Hypothesis which is true if we are able to reject the null hypothesis.

5. **Two tail test or One tail test**:

<table>
<thead>
<tr>
<th>Hypothesis</th>
<th>Test Type</th>
<th>Decision</th>
</tr>
</thead>
<tbody>
<tr>
<td>( H_0: \mu = \mu_{\text{Hypo}} )</td>
<td>Use a two tail test</td>
<td>Reject ( H_0 ) if test statistic &gt; upper critical or &lt; lower critical values</td>
</tr>
<tr>
<td>( H_0: \mu \leq \mu_{\text{Hypo}} )</td>
<td>Use a one tail test</td>
<td></td>
</tr>
<tr>
<td>( H_0: \mu \geq \mu_{\text{Hypo}} )</td>
<td>Use a one tail test</td>
<td></td>
</tr>
</tbody>
</table>

6. **Chebyshev’s Inequality**: For a random variable, \( X \), with a standard deviation of \( \sigma \), the probability that \( X \) is within \( n \) standard deviations of \( \mu \) is less than or equal to \( \frac{1}{n^2} \). For a given level of variance, Chebyshev’s inequality places an upper limit on the probability of a variable being more than a certain distance from its mean. For a given distribution, the actual probability may be considerably less. Take, for example, a standard normal variable. Chebyshev’s inequality tells us that the probability of being greater than two standard deviations from the mean is less than or equal to 25%. Used when we do not know the distribution of the variable.
LINEAR REGRESSION

Reading: Linear Regression with One Regressor (Chapter 4, James Stock and Mark Watson, Introduction to Econometrics, Brief Edition (Boston: Pearson Education, 2008))

1. **Definition:** Regression analysis has the goal of measuring how change in one variable, called ‘Dependent’, can be explained by changes in one or more variables called ‘Independent Variables’. The parameters of an equation indicate relationship. A generic form of an equation is: \( Y = B_0 + B_1X_1 + \epsilon \) where \( B_0 \) is called Intercept, \( B_1 \) is called Slope, \( X_1 \) is the independent variable and \( \epsilon \) is error term.

2. **Properties of Linear Regression:**
   a. Independent variables (X) enter into the equation without any transformation such as \( X^2 \)
   b. Dependent variable (Y) is a linear function of parameters but does not require that there be linearity in variables

3. **OLS (ordinary Least Squares):** It is a method to determine the values of \( B_0 \) and \( B_1 \) in the above equation. The method tries to minimize the sum of squares of error term i.e. \( \sum (Y - B_0 - B_1X_1)^2 \) or \( \sum \epsilon_i^2 \). OLS computes the values of the parameters \( B_1 \) as:
   \[
   \frac{\sum(x_i - \bar{x})(y_i - \bar{y})}{\sum(x_i - \bar{x})^2}
   \]

4. **Coefficient of Determination:** The coefficient of determination (\( R^2 \)) measures the goodness of fit. It is interpreted as % of variation in the dependent variable explained by the independent variable. Note that \( R^2 = \frac{SSR}{SST} \)

5. **R Square:** It is used to show the percentage variation explained by all the independent variables. A typical problem with R Square is that it increases as number of variables increases. To overcome this problem, we use Adjusted R Square: \( R_{Adj}^2 = \left[ \left( \frac{n-1}{n-k-1} \right) \times (1 - R^2) \right] \) where \( k \) is no. of independent variables

6. **SST, SSR and SSE:**
   a. SST: Sum of Squares total is the sum of squares of predicted values from their average
   b. SSR(ESS): Sum of square regression, also known as Explained sum of squares of deviation from its average

![Diagram of Linear Regression](image-url)
1. **Slope Hypothesis Test:**
   a. A frequently asked question is whether an estimated slope coefficient is statistically different from zero. i.e. $H_0: B_1 = 0$ and $H_A: B_1 \neq 0$.
   b. The confidence interval for the regression coefficient is given by $B_1 \pm (t_{\text{stat}} + \sigma_{\hat{b}_1})$ where $B_1$ is the mean of slope coefficient, $\sigma_{\hat{b}_1}$ is standard error of slope and $t_{\text{stat}}$ is a two-tailed t test with degrees of freedom (df) as $n-2$.
   c. Alternatively, we might also use the t-test to determine if the slope is equal to some hypothesized value ($B_{\text{Hypo}}$). The critical value is given by $t_{\text{critical}} = \frac{(B_1 - B_{\text{Hypo}})}{\sigma_{\hat{b}_1}}$ If the value of $t_{\text{critical}}$ is not in the range above, we reject $H_0$.
   d. For testing whether an independent variable explains the variation in the dependent variable, the hypothesis test is whether slope is zero.

2. **Dummy Variable:** An independent variable, binary in nature, is called a dummy variable. They are assigned values of 0 or 1.

3. **Homoskedasticity & Heteroskedasticity:**
   a. If the variance of the residuals(ε) is constant across all observations in the sample, the regression is Homoscedastic.
   b. If variance of the residual(ε) increases as independent variable increases, then it's called Conditional Heteroskedasticity.

4. **OLS as unbiased estimator:** OLS places no restrictions on conditional variance of the residual term thus OLS remains unbiased and consistent despite the residual term.

5. **Gauss-Markov Theorem:** It states that in a linear regression model in which the errors have expectation zero and are uncorrelated and have equal variances, the best linear unbiased estimator (BLUE) of the coefficients is given by the ordinary least squares (OLS) estimator. Here residual is assumed to be homosekdastic and . Limitations of the theorem are:
   a. In economic applications, residual or error terms are heteroskedastic.
   b. Under certain conditions, other estimators are more efficient than OLS.
MULTIPLE REGRESSOR

Reading: Linear Regression with Multiple Regressors: Hypothesis Tests and Confidence Intervals (Chapter 6, James Stock and Mark Watson, Introduction to Econometrics, Brief Edition (Boston: Pearson Education, 2008))

1. **Omitted Variable Bias**: An omitted variable bias is present when
   a. Omitted variable is correlated with the movement of the independent variable
   b. Omitted variable is a determinant of the dependent variable

2. **Multiple regression** takes more than one independent variable (x), thus equation is of the form: \[ Y = B_0 + B_1X_1 + B_2X_2 + B_3X_3 + \ldots + B_nX_n + \epsilon \]

3. **Assumptions of Multiple regression**:
   a. \( \epsilon \) has a mean of zero
   b. \( X_1, X_2, X_3, \ldots, X_n \) are i.i.d
   c. Large outliers are unlikely
   d. No perfect multicollinearity

4. **Standard Error of Regression (SER)**: It measures the standard deviation of the error term.

5. **Goodness of Fit**: \( R^2 \) is not a good measure of goodness of fit as it increases with increase in the number of independent variables. Thus, it overestimates regression accuracy. To overcome this bias, we use adjusted \( R^2 \)
   which is given as: \[ R_{Adj}^2 = \left( \frac{n-1}{n-k-1} \right) \times (1 - R^2) \] where \( n \) is the number of observations and \( k \) is number of independent variables.
   a. Note that \( R_{Adj}^2 \leq R^2 \)

6. **Multicollinearity**: It refers to the condition that two or more variables or linear combination of the independent variables in a multiple regression are highly correlated with each other. If correlation is perfect, then OLS estimation is not possible.
   a. Test is to check If \( R^2 \) value is high but t-test shows that none of the independent variable are significant, then multicollinearity exists
MULTIPLE REGRESSOR HYPOTHESIS

Reading: Hypothesis Tests and Confidence Intervals in Multiple Regression (Chapter 7, James Stock and Mark Watson, Introduction to Econometrics, Brief Edition (Boston: Pearson Education, 2008))

1. **Hypothesis Testing:**
   a. The t-statistic used for the single regressor can also be used here.
   b. The degree of freedom (df) changes to \( n - k - 1 \) in place of \( n - 1 \)

2. **P-values:** The p-values are the smallest level of significance for which the null hypothesis can be rejected.
   a. Decision: \( p < \alpha \); reject \( H_0 \)
   b. For statistical significance, we can either use the confidence interval or the p-values. Both will result in the same decision being taken

3. **Joint Hypothesis Testing:** It sets two coefficients as zero i.e. \( H_0 : B_1 = 0 \) and \( B_2 = 0 \). \( H_A \) is that one of them is not zero. If even one of them is not equal to zero, we can reject the null hypothesis. For joint hypothesis testing, we always use the F-Test
   a. F-test is always one tail test with \( F = \frac{SSR/k}{SSE/(n-k-1)} \)
   b. The degree of freedom (df) in numerator is \( k \) and df in denominator is \( n - k - 1 \)
   c. The F-value obtained is tested against the critical one tail value
   d. Rejection indicates that at least one of the values is not equal to 0

4. **Specification Bias:** It refers to the fact that slope coefficients and other statistics for a given variable are usually different in simple regression when compared to multiple regression of the same coefficients.

5. **Model Misspecifications:**
   a. Functional Form:
      i. Important variables are omitted
      ii. Variable should be transformed
      iii. Data is improperly pooled
   b. Correlation with Error Term
      i. Lagged dependent variable is taken as independent
      ii. Function is used as independent variable
# Modelling & Forecasting Trend

**Reading:** Modeling and Forecasting Trend (Chapter 5, Francis X. Diebold, Elements of Forecasting, 4th Edition)

1. **Definitions:**
   
a. **Time Series:** A set of observations for a variable over successive periods of time

<table>
<thead>
<tr>
<th>Trend</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Linear Trend</td>
<td>( y_t = \beta_0 + \beta_1(t) )</td>
</tr>
<tr>
<td></td>
<td>where ( y_t ) is value of time series, ( \beta_0 ) is regression intercept, ( \beta_1 ) is slope and ( t ) is time</td>
</tr>
<tr>
<td></td>
<td>The formula is same for OLS (Ordinary Least Squares)</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trend</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinear trend (Quadratic)</td>
<td>( y_t = \beta_0 + \beta_1(t) + \beta_1(t)^2 )</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Trend</th>
<th>Equation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Nonlinear Trend (Exponential)</td>
<td>( y_t = \beta_0 e^{\beta_1(t)} )</td>
</tr>
<tr>
<td></td>
<td>Financial data can bet better modelled with Exponential trends</td>
</tr>
</tbody>
</table>
2. Selecting a model:
   a. Linear trend Model: if data points appear equally distributed above and below the regression line, use Linear trend (e.g. Inflation Data)
   b. Log-linear trend: If data plot is curved, and data is persistently negative or positive for a period of time, use log-linear model.

3. Model Selection:
   a. Mean Squared Error (MSE): Select one model with least MSE where \( \text{MSE} = \frac{\sum_{t=1}^{T} e_{t}^2}{T} \)
   b. Unbiased MSE(s square): \( S^2 = \frac{\sum_{t=1}^{T} e_{t}^2}{T-k} \) where MSE is adjusted for degrees of freedom (k)
   C. AIC (Akaike Information Criterion): \( \text{AIC} = e^{\left(\frac{2k}{T}\right)} \frac{\sum_{t=1}^{T} e_{t}^2}{T-k} \)
   d. SIC (Schwarz Information Criterion): \( \text{SIC} = T^{\left(\frac{k}{T}\right)} \frac{\sum_{t=1}^{T} e_{t}^2}{T-k} \)

4. Criteria Consistency:
   a. Probability of approaching true model increases as sample size increases
   b. Adjusting for degrees of freedom is important. SIC (Schwarz Information Criterion) is the most consistent
MODELING & FORECASTING SEASONALITY

Reading: Modeling and Forecasting Seasonality (Chapter 6, Francis X. Diebold, Elements of Forecasting, 4th Edition)

1. Definitions:
   a. Seasonality: A pattern in time series data that tends to repeat from year to year
      i. Seasonally adjusted series: Created by removing seasonal variation from data

2. Modeling Seasonality: It assigns season dummy variables a value of either 1 or 0. The number of dummy variables to be included is s-1 where s is the number of seasons. The intercept accounts for the missing season.
   a. E.g. $\text{EPS}_t = \beta_0 + \beta_1 D_{1,t} + \beta_2 D_{2,t} + \beta_3 D_{3,t}$
      i. Where EPS is quarterly earnings per share
      ii. $D_{1,t}$ is 1 if period is first quarter else 0
      iii. $D_{2,t}$ is 1 if period is second quarter else 0
      iv. $D_{3,t}$ is 1 if period is third quarter else 0
      v. Intercept $\beta_0$ represents EPS for fourth quarter
Characterizing Cycles

Reading: Characterizing Cycles (Chapter 7, Francis X. Diebold, Elements of Forecasting, 4th Edition (Mason, Ohio: Cengage Learning, 2006))

1. Definitions:
   a. Time Series: A set of observations for a variable over successive periods of time
   b. Autoregressive: Past values are used to predict future values
   c. Covariance Stationary: Mean, Variance and Covariance do not change over time
   d. Autocovariance Function: Tool used to quantify stability of the covariance
   e. White Noise Process: A time series with zero mean, constant variance and no serial correlation is referred to white noise
   f. Q statistic: Used to measure the degree to which autocorrelations vary from zero and whether white noise is present in the data set
   g. Box-Pierce Q Statistic: Reflects magnitude of the correlations as it uses the sums squared autocorrelations
   h. Ljung-Box Q Statistic: Replaces the sum of squared autocorrelations with weighted sum of squared autocorrelations
   i. World’s Theorem: Evaluates covariance stationary as pre-requisite for time series modelling
1. **MA(1)**[Moving Average First Order Process]: It is a linear regression of the current values of a time series against both the current and previous unobserved white noise error terms which are random shocks. It is defined as: $y_t = \varepsilon_t + \theta \varepsilon_{t-1}$
   a. It has a mean of zero and a constant error term
   b. The autocorrelation cutoff ($\rho$) is zero for any value beyond the first error term and can be computed as $\rho = \frac{\theta^2}{1+\theta^2}$

2. **MA(q)** process: It increases the lagged terms to q and all other values after that have zero autocorrelation.

3. **AR(1)** [Autoregressive First Order]: It is specified in the form of a variable regressed against itself in the lagged form. It is given as: $y_t = \varepsilon_t + \phi y_{t-1}$
   a. The absolute value of the coefficient must be less than 1 i.e. $|\phi|<1$
   b. Yule Walker: $\rho_t = \phi^t$
   c. Moving averages exhibit autocorrelation cut-off while Autoregressive series exhibit autocorrelation decay but never become zero

4. **AR(p)**: It increases the lagged terms to p in the AR series

5. **ARMA**[Autoregressive Moving Averages]: It combines MA and AR models and is defined as:
   a. $y_t = \varepsilon_t + \theta \varepsilon_{t-1} + \phi y_{t-1}$
   b. The autocorrelations decay gradually like in AR process
   c. We can have ARMA(p,q) model which is p terms of AR and q terms of MA
   d. This model enables modeling of more complex observations as compared to MA or AR processes
VOLATILITY


1. Definitions:
   a. Volatility: Standard deviation of a variables continuously compounded return
   b. Square root of time rule: Extending daily volatility to a number of T days is achieved by multiplying daily volatility with square root of T
   c. Variance: Square of volatility. Variance is linear in contrasts to volatility which increases with square root of time. Thus, if Volatility is 1.5% daily, Variance is \((1.5\%)^2 = 0.0225\).
      i. Volatility for 10 days is: \((1.5\%) \times \sqrt{10} = 0.047\)
      ii. Variance for 10 days is: \(0.0225 \times 10 = 0.225\)
   d. Implied volatility; Volatility level that will equate an options market price to its model price (Black Scholes Merton model – Covered in Book 4)

2. Power Law: An alternative to using probabilities from a normal distribution. It allows for fatter tails

3. EWMA (Exponentially Weighted Moving Average): it generates volatility estimates based on weightings of previous estimates of volatility. Advantage is that it requires few data points
   a. \(\sigma_n^2 = \lambda \sigma_{n-1}^2 + (1 - \lambda) u_{n-1}^2\) where \(\lambda\) is the decay rate and \(u\) is the return

4. GARCH (p,q) [Generalised Autoregressive Conditional Heteroskedasticity]: GARCH (1,1) is the most commonly used model and it incorporates not only the most recent estimates of variance but also variable that accounts for long-run average level of variabce.
   a. \(\sigma_n^2 = \omega + \beta \sigma_{n-1}^2 + \alpha u_{n-1}^2\) where \(\omega\) is weighted long run variance, \(\beta\) is weighting of previous volatility estimate, \(\alpha\) is weighting of previous period return.
   b. \(\omega = \gamma V_L\) where \(V_L\) is the long run variance
   c. \(V_L = \frac{\omega}{1 - \alpha - \beta}\) and \(\alpha + \beta + \gamma = 1\)
   d. \(\alpha + \beta\) is called the persistence rate. It describes the rate at which volatility will revert to its long term average.
Correlations and Copulas


1. Definitions:
   a. Correlation and Covariance: Measure linear relationship between the co-movements over time. Correlation is between -1 to 1 and Covariance is between -∞ to ∞
   b. $\rho_{x,y} = \frac{\text{Cov}_{x,y}}{\sigma_x \sigma_y}$
   c. Independent Variables: Where one variable does not impact the probability distribution of another variable
   d. Copula creates a joint probability distribution between two or more variables while maintaining their individual marginal distributions

2. EWMA (Exponentially Weighted Moving Average): Gives current observations more weight that previous observations.
   a. $\text{Cov}_n = \lambda \text{Cov}_{n-1} + (1 - \lambda)X_{n-1}Y_{n-1}$
   b. Covariance of an asset with itself is the same as Variance i.e. $\text{Cov}_x = \text{Var}_x = \sigma^2_x$

3. GARCH(1,1) [Generalised Autoregressive Conditional Heteroskedasticity]:
   a. $\text{Cov}_n = \gamma V_L + \beta \text{Cov}_{n-1} + \alpha X_{n-1}Y_{n-1}$; in the equation $\alpha+\beta+\gamma=1$ always. Also, $\omega = \gamma V_L$

4. Copula: Steps for finding the correlation between two marginal distributions:
   a. Plot each marginal distribution to a normal distribution (Percentile wise)
   b. Assume that the two Normal distributions are bivariate normal and calculate correlation
   c. Gaussian Copula: Maps marginal distribution to a standard normal distribution
   d. Student’s $t$ Copula: Maps to Student’s $t$ distribution
   e. Multivariate Copula: For more than two variables like Factor Copula models

5. Tail Dependence: Student’s $t$ distribution exhibits better tail values than Gaussian distribution as during times of stress, most of the tail values are similar like in Student’s $t$ distribution thus its better at determining correlations

6. Positive-semidefinite matrix: A matrix is positive-semidefinite if it is internally consistent. The covariance matrix of a one factor model is positive-semidefinite. One factor model also requires $N$ estimates for correlations where each variable $N$ is correlated to factor $F$
SIMULATION METHODS


1. **Steps for Monte Carlo Simulation**
   a. Specify the DGP (Data Generating Process) and generate the data
   b. Do the regression and calculate the test statistic
   c. Save the statistic or parameter of interest
   d. Go back to step 1 and repeat N times

2. **Sampling Error in Monte Carlo**: If we run the same simulation twice, we are likely to get different values for the parameter of interest. There are two methods to reduce the sampling error in Monte Carlo:
   a. Antithetic Variables: Take complement of a set of random numbers and run a parallel simulation
   b. Control Variates: Employ a similar variable whose outcomes are known and simulate it with parameter

3. **Bootstrapping**: Random sampling with replacement i.e. the original value is not taken out for the next round.
   a. Disadvantages: If there are outliers, they may impact the conclusions of the test. Use of bootstrap implicitly assumes that data are independent of each other.

4. **Pseudo-random draws**: Computer generated random number draws.

5. **Disadvantages of Simulation**
   a. It might be computationally expensive
   b. Results might not be precise
   c. Results are hard to replicate
   d. Results are experiment specific