TIME SERIES AND FORECASTING

1. **Probability & Statistics:**
   1. **Probability distributions**

Probability distributions that describe the likelihood of a set of random variables having a particular set of values. Normal distribution is a symmetrical distribution around the mean, t-distribution is used to model data with heavier tails than normal, and F-distribution is used in hypothesis testing.

* + 1. **Normal distributions**
    2. **t distributions**
    3. **F distributions:**
  1. **Hypothesis testing:**

A statistical procedure used to test a claim or hypothesis about a population parameter. **The objective is to determine if there is enough evidence to support the claim.**

1. **Simple return vs Log returns:**
   1. **Simple returns:**

A measure of the percentage change in an investment's value over a given period of time. It is calculated as (price at the end of the period - price at the beginning of the period) / price at the beginning of the period.

* 1. **Log returns:**

A measure of the percentage change in an investment's value over a given period of time, calculated as the natural logarithm of the end-of-period price divided by the beginning-of-period price.

1. **Basic Trading Models:**
   1. **MACD (Moving Average Convergence Divergence):**

A momentum indicator used in technical analysis to identify changes in the strength, direction, and momentum of a stock or other asset.

* 1. **Bollinger Bands:**

A volatility indicator used in technical analysis to identify whether a stock or other asset is overbought or oversold.

1. **Basic Time Series concepts:**
   1. **White noise:**

A sequence of random variables with constant mean and constant variance.

White noise is a random signal in which the values at any given time step are independent and identically distributed with a mean of zero. It is a commonly used concept in time series analysis and forecasting.

White noise can be expressed mathematically as follows:

In practice, white noise is used as a benchmark for evaluating other time series forecasting models. If a time series model is no better than white noise, it is considered to be a poor model for the data.

The presence of white noise in a time series can be measured using various statistical tests, such as the Ljung-Box test, which tests the null hypothesis that the residuals of a time series model are white noise. The accuracy of a time series model in modeling white noise can be evaluated using standard time series evaluation metrics such as mean absolute error, mean squared error, and root mean squared error.

* 1. **Stationary/Non-stationary series:**

A time series is stationary if its mean, variance, and autocorrelation structure do not change over time. A non-stationary time series has a mean, variance, or autocorrelation that changes over time.

* 1. **Measures of the correlation between a time series and its lagged values.**
     1. **ACF (Autocorrelation Function)/**
     2. **PACF (Partial Autocorrelation Function):**
  2. **ADF (Augmented Dickey-Fuller) test:**

A statistical test used to determine whether a time series is stationary or not.

1. **Extrapolation models:**
   1. **Methods used to smooth out fluctuations in a time series and to forecast future values based on past data.**
      1. **Moving Average Filter**

Moving Average Filter (MA) is a simple time series forecasting method used for univariate time series data. It is a weighted average of past observations, with equal weights given to all observations.

The moving average filter can be expressed as follows:

In practice, the moving average filter is used to smooth a time series and remove noise, making it easier to identify trends and patterns. It can also be used for short-term forecasting, with the most recent observations being used to calculate the average for the next time step. The accuracy of the forecast can be measured using standard time series evaluation metrics such as mean absolute error, mean squared error, and root mean squared error.

The moving average filter is simple to implement and has a low computational cost, making it a popular choice for many time series applications. However, its limitations include its inability to handle trends and seasonality, and its dependence on the choice of .

* + 1. **EWMA (Exponentially Weighted Moving Average):**

Exponentially Weighted Moving Average (EWMA) is a time series forecasting method used for univariate time series data. It is a weighted average of past observations, with more recent observations receiving higher weights.

The EWMA model can be expressed as follows:

The smoothing parameter, , determines the amount of weight given to recent observations, with a larger value of corresponding to more weight given to recent observations.

In practice, EWMA is used to smooth a time series and remove noise, making it easier to identify trends and patterns. It can also be used for short-term forecasting, with the most recent observation being used as a prediction for the next time step. The accuracy of the forecast can be measured using standard time series evaluation metrics such as mean absolute error, mean squared error, and root mean squared error.

EWMA is simple to implement and has a low computational cost, making it a popular choice for many time series applications.

* 1. **Linear Trend:**

A method of forecasting future values by extrapolating a linear relationship between past values and time.

Linear Trend is a time series forecasting method used for univariate time series data that involves fitting a line to the observed data. It assumes that the underlying trend in the data is linear and uses this assumption to make predictions for future time steps.

The linear trend model can be expressed as follows:

In practice, linear trend is used to model a trend in the data, such as an increase or decrease over time. It can be used to make predictions for future time steps by extrapolating the line. The accuracy of the forecast can be measured using standard time series evaluation metrics such as mean absolute error, mean squared error, and root mean squared error.

Linear trend is simple to implement and can handle a linear trend in the data, but its limitations include its inability to handle non-linear trends, seasonality, and other complex patterns. It is important to carefully inspect the data and assess the presence of any non-linear trends or patterns before using a linear trend model for forecasting.

1. **Forecasting models:**
   1. **Naive model:**

A simple method of forecasting future values by using the most recent observation as the forecast for all future periods.

The Naive model is a simple time series forecasting model that assumes that the next value in a time series will be equal to the last observed value. It is often used as a benchmark model or as a simple baseline to compare the performance of more complex models.

Mathematical formula:

Let's consider a time series with observations is the next value to be forecasted. The Naive model can be written as:

Example:

Let's consider a time series data of the number of sales of a product. To forecast the sales for next month, the Naive model would assume that the sales for next month will be equal to the sales in the current month.

Usage:

The Naive model is useful in several cases such as:

* When there is no historical data available
* When there is no trend or pattern in the time series
* When the time series is not stationary
* As a baseline to compare the performance of more complex models.

However, the Naive model is not useful in most real-world scenarios as it does not take into account the trend or seasonality in the time series.

* 1. **ARIMA (AutoRegressive Integrated Moving Average):**

A time series model that uses a combination of autoregression, difference, and moving average to forecast future values.

ARIMA (AutoRegressive Integrated Moving Average) is a time series forecasting method that models the temporal structure in time series data to make predictions. It consists of three components:

1. **Autoregression (AR)**: The dependence of an observation on past observations is modeled using AR terms, denoted by AR(p), where p is the number of lagged observations used. Mathematically, it can be represented as:

The equation models the value of the time series at time t as a linear combination of past values and a residual term. The coefficients φ\_1 to φ\_p determine the importance given to each past value, while the residual term captures any remaining variability in the time series. The AR component captures any long-term patterns in the time series that persist across multiple time steps.

1. **Integration (I):** The difference between an observation and its previous value is taken to make the time series stationary, denoted by I(d), where d is the number of times the difference is taken. Mathematically, it can be represented as:

The difference between consecutive observations is taken to make the time series stationary, as stationarity is a necessary assumption for many time series models, including ARIMA. By taking the difference, trends and seasonality in the original time series are removed, resulting in a stationary time series that is easier to model and make predictions for.

1. **Moving Average (MA):** The current observation is modeled as a function of past residuals or errors, denoted by MA(q), where q is the number of lagged errors used. Mathematically, it can be represented as:

The equation models the value of the time series at time t as a linear combination of past residuals and a constant term. The coefficients to determine the importance given to each past residual, while the constant term represents the overall mean or expected value of the time series. The MA component captures any short-term patterns in the residuals that were not captured by the autoregression component.

In practice, ARIMA is used to model time series data such as stock prices, sales, and weather patterns. The process involves selecting appropriate values of p, d, and q, fitting the model to the data, and using the model to make predictions. ARIMA is widely used in various applications such as sales forecasting, demand forecasting, and financial analysis.

* 1. **Holts & Winters:**

A method of time series forecasting that uses exponential smoothing to handle both trend and seasonality in the data.

Holt's and Winters' method are exponential smoothing techniques for time series forecasting.

Holt's method uses the following equations:

Winters' method extends Holt's method to handle seasonality by including an additional seasonal component:

In practice, Holt's and Winters' methods are used to forecast future values of a time series based on past observations. They are simple and easy to implement, and can handle time series data with both trend and seasonality. The parameters , and are estimated by minimizing the sum of the squared errors between the observed and forecasted values. The accuracy of the forecast can be measured using standard time series evaluation metrics such as mean absolute error, mean squared error, and root mean squared error.

1. **Measuring Forecasting Errors:**
   1. **Measures of the difference between actual and forecast values.**
      1. **Bias**

Bias refers to the systematic error or deviation of the estimated value from the true value in statistical analysis. In mathematical terms, bias is defined as the expected difference between the estimated value and the true value, expressed as follows:

In time series forecasting, bias can be calculated as the average difference between the predicted values and the actual observed values. The objective is to minimize the bias in order to produce accurate predictions. A function to calculate the bias in a time series forecasting model could be expressed as:

* + 1. **MSE (Mean Squared Error)**

Mean Squared Error (MSE) is a commonly used metric to measure the accuracy of a forecast or prediction model. The MSE is the average of the squared differences between the predicted values and the actual observed values.

Function:

Given a prediction model with predicted values ŷ and actual observed values y, the MSE is calculated as:

where n is the number of observations.

* + 1. **RMSE (Root Mean Squared Error)**

Root Mean Squared Error (RMSE) is a widely used metric to evaluate the accuracy of a forecasting model. It measures the difference between the actual and predicted values of a time series.

The function to calculate RMSE is:

The RMSE is expressed in the same units as the original time series and it provides an aggregate measure of the magnitude of the error in the forecast. A lower RMSE value indicates a better fit of the model to the data.

* + 1. **SE (Squared Error):**

SE (Squared Error): The squared error is a commonly used measure of the difference between two continuous variables. It is defined as the difference between the actual value of the target variable and the predicted value, squared. The squared error is used as a performance metric in regression analysis to assess the accuracy of the model.

**,**

The relationship between SE, RMSE and MSE is:

The formula for MSE is:

The formula for RMSE is:

* 1. **MAE (Mean Absolute Error)**

Mean Absolute Error (MAE) is a measure of forecast error in time series forecasting. It calculates the average magnitude of the errors in the predictions, without considering their direction.

The mathematical function for MAE is as follows:

Let's say we have n observations in our time series, and the observed value is and the predicted value is for .

Then, the MAE can be defined as:

* 1. **Measures of the difference between actual and forecast values expressed as a percentage**
     1. **MPE (Mean Percentage Error)**

Mean Percentage Error (MPE) is a forecasting error metric that measures the difference between the actual and forecasted values as a percentage of the actual value. It is calculated as follows:

The MPE gives a measure of the percentage error in the forecast and is a commonly used error metric in finance and economics. Positive MPE values indicate that the forecast overestimates the actual value, while negative MPE values indicate that the forecast underestimates the actual value. A zero MPE indicates that the forecast and actual values are equal.

* + 1. MAPE (Mean Absolute Percentage Error)

MAPE (Mean Absolute Percentage Error) is a measure of the forecast accuracy that represents the average absolute percentage difference between actual and predicted values. It can be calculated as follows:

MAPE is widely used in forecasting to measure the percentage deviation of the forecast from the actual value. It provides a clear representation of the accuracy of the forecast, allowing analysts to determine the degree of error in the predictions. The lower the MAPE value, the better the forecast accuracy.

However, it is important to note that MAPE has some limitations. For instance, it is not symmetrical, meaning that an absolute percentage error of 10% is not equal to an absolute percentage error of -10%. Additionally, MAPE can become undefined or misleading when is equal to zero, making it unsuitable for some forecasting scenarios.

* 1. **Theil's U1/U2:**

Measures of the accuracy of a time series forecast that take into account the size of the errors as well as their distribution.

Theil's U1 and U2 are measures of forecast accuracy used in economics and econometrics. Theil's U1, also known as the mean logarithmic deviation (MLD), measures the average percentage deviation of the forecasts from the actual values. Theil's U2, also known as the mean squared logarithmic deviation (MSLD), measures the average squared percentage deviation of the forecasts from the actual values.

Mathematically, Theil's U1 is given by:

Theil's U2 is given by:

Theil's U1 and U2 are used as a measure of forecast accuracy, especially in comparison to other forecasts or models. A lower value of U1 or U2 indicates a more accurate forecast.

Note: ln is the natural logarithm function.

* 1. **Decomposition (Bias, Variance, Covariance proportion):**

A method of breaking down the forecast error into its components of bias, variance, and covariance proportion.

Decomposition is a statistical method used to break down a time series into three parts: bias, variance, and covariance proportion. The goal of decomposition is to isolate and understand the underlying components that make up a time series, so that they can be effectively modeled and forecasted.

Bias: Bias is the difference between the average forecast and the actual values. Mathematically, it can be represented as the mean of the difference between the actual and forecasted values.

Mathematical function for Bias:

Variance: Variance measures the spread of the data points around the mean. It is a measure of the randomness or unpredictability of the time series.

Mathematical function for Variance:

Covariance Proportion: Covariance proportion measures the proportion of the forecast error that is due to covariance between the variables. It is calculated as the ratio of the covariance between the forecast error and the actual values to the variance of the actual values.

Mathematical function for Covariance Proportion:

The decomposition method is used to help identify and understand the underlying factors affecting a time series, such as seasonality, trend, and residuals. By breaking down a time series into its component parts, it is possible to diagnose and address any issues with the forecast and improve the accuracy of future predictions.

1. **Bivariate Time Series Models**
   1. **Spurious regression:**

A type of regression that occurs when two non-stationary time series are used to create a false relationship between them.

Spurious regression refers to a false relationship that appears to exist between two non-stationary time series. This can occur when the two time series are used to estimate a regression model, even though there is no actual causal relationship between them.

The presence of a spurious relationship is typically due to the presence of common trends in the time series, which can cause the regression coefficients to be biased. This bias can cause the estimated relationship to appear significant, even though it is not.

To detect and avoid spurious regression, it is necessary to first test for the stationarity of the time series and apply appropriate techniques, such as differencing or co-integration, to make the time series stationary.

Spurious regression is a common problem in financial time series analysis and can lead to incorrect conclusions and unreliable predictions. Therefore, it is important to be cautious when interpreting the results of regression models involving non-stationary time series.

**NOTE:**

The presence of a spurious relationship can be determined by performing statistical tests such as the Augmented Dickey-Fuller (ADF) test for unit roots or the Johansen test for cointegration. If the two time series are found to be non-stationary, the relationship between them may be spurious and needs to be further investigated. To remove the effect of non-stationarity, the time series can be made stationary by taking first differences or applying a transformation such as the natural log. After this pre-processing step, a regression model can be fit to the stationary time series to determine if the relationship is real or not.

* 1. **Dynamic model:**

A time series model that accounts for the relationships between the current and past values of the series. Dynamic models in bivariate time series models describe the relationships between the current and past values of two or more time series. These models are used to analyze how changes in one time series affect changes in another time series. The objective of a dynamic model is to capture the underlying causal relationships between the variables and to forecast the future values of the series.

A common example of a dynamic model in bivariate time series is the Vector Autoregression (VAR) model, where the relationship between the time series is described by a set of equations. In a VAR(p) model, the current values of each time series are modeled as a linear combination of their past p values and the past values of all other time series in the model. The coefficients in the model represent the strengths of the relationships between the series, and can be estimated using maximum likelihood estimation or other statistical methods. The impulse response function can also be used to analyze how a shock in one variable affects the other variable over time.

Dynamic models can be used to identify causal relationships between variables, to analyze the effects of exogenous variables, and to make forecasts based on the relationships between the time series. They are widely used in econometrics, finance, and other fields to analyze time series data.

* 1. **Granger Causality:**

A statistical concept used to determine whether one time series is a useful predictor of another time series. It is determined through the use of lagged values of the predictor time series in a regression analysis. The objective is to determine the causality relationship between two time series.

Granger Causality is a concept in econometrics and time series analysis that is used to determine whether a time series is influenced by another time series. The main idea behind this concept is that if a time series X causes another time series Y, then past values of X should be able to provide information about the future values of Y that is not contained in past values of Y alone.

Granger causality can be tested using a regression analysis. The basic idea is to fit two regression models: one using only past values of Y as predictors and one using both past values of X and Y as predictors. If the second model provides a better fit, it is said that X Granger-causes Y.

The mathematical notation for the models can be represented as follows:

* 1. **Co-integration & Engle-Granger test:**

Co-integration refers to a long-term statistical relationship between two or more non-stationary time series. The Engle-Granger test is used to determine the presence of co-integration between two time series.

Co-integration is a statistical property of two or more time series where the differences between their values over time tend to remain constant. The Engle-Granger test is used to determine whether two time series are co-integrated.

The Engle-Granger test consists of two steps:

1. Regression analysis: Fit a linear regression model with one time series as the dependent variable and the other as the independent variable.
2. Augmented Dickey-Fuller (ADF) test: Perform an ADF test on the residuals of the regression model to check for stationarity.

If the residuals of the regression model are stationary, then the two time series are considered to be co-integrated. The co-integration relationship can then be modeled using an Error Correction Model (ECM).

Mathematically, the Engle-Granger test can be expressed as follows:

Let and be two time series. The linear regression model can be represented as:

The ADF test on the residuals of the regression model can be represented as:

If the null hypothesis of non-stationarity in the residuals is rejected, then the two time series are considered to be co-integrated.

* 1. **Error Correction Model (ECM):**

A type of multivariate time series model that captures the long-run relationship between two or more time series, while also accounting for any short-run deviations from that relationship. The objective is to identify and correct any errors in the relationship between the time series.

Error Correction Model (ECM) is a time series model used in econometrics to examine the long-run relationship between two non-stationary time series variables. The ECM framework expresses the long-run relationship between two time series as a linear combination of the first differences of the series, and a linear combination of their lagged levels. The error correction term (ECT) captures the short-run dynamics and the deviation from the long-run relationship. The ECM can be expressed mathematically as:

The ECM is used to study the relationship between two variables in the long run, while capturing the short-run dynamics. The ECM framework provides a powerful tool for studying long-run relationships in the presence of non-stationarity, co-integration, and short-run dynamics.

1. **Multivariate Time Series Models:**
   1. **VAR (Vector Autoregression):**

A type of multivariate time series model that captures the relationships between multiple time series. The model uses lagged values of all the series in a regression analysis to determine the relationship between them. The objective is to understand the effect of one time series on another.

Vector Autoregression (VAR) is a multivariate time series model that uses a set of autoregressive equations to model the interactions between multiple time series.

In a VAR model, each time series is modeled as a linear combination of its own past values and the past values of all other time series in the model. The basic form of a VAR model can be written as:

The objective of VAR is to model the dynamic relationships between multiple time series, and to make predictions about future values based on these relationships. VAR models can be used for forecasting, simulation, and causality analysis. The model can be estimated using maximum likelihood estimation (MLE) or other estimation techniques.

VAR (Vector Autoregression) model:

A VAR model represents the relationships between multiple time series as a set of regression equations. The general form of a VAR model can be written as:

The coefficients in the VAR model can be estimated using ordinary least squares (OLS) or maximum likelihood estimation (MLE) methods, and the residuals can be analyzed to check for stationarity and ensure that the model is a good fit for the data.

* 1. **Impulse Response Function:**

A tool used to evaluate the impact of a shock in one time series on another in a multivariate time series model. It shows the response of one series to a sudden change in another series over time.

The IRF for a VAR or VECM model can be mathematically represented as:

The IRF can be estimated using the estimated coefficients from a VAR or VECM model, and it can be plotted over time to visualize the dynamic impact of the shock.

In the context of a VAR model, the IRF is calculated using the estimated coefficients from the VAR model and represents the impact of a shock in one time series on the other time series.

In the context of a VECM model, the IRF is calculated using the estimated coefficients from the VECM model and represents the impact of a shock in one time series on the other time series after accounting for any co-integration between the series.

Impulse Response Function (IRF) is a tool used in time series analysis to understand the dynamic relationship between a time series and its response to a one-time shock or impulse. IRF measures the effect of a shock in one time series on another time series over time.

In the context of a VAR or VECM model, IRF provides a graphical representation of the impact of a shock in one time series on the other time series. The IRF plots the response of a time series to a one-standard-deviation shock in another time series over a specified number of periods. This allows the analyst to understand how changes in one time series propagate and affect the other time series over time.

IRF is commonly used in macroeconomic analysis to understand the impact of monetary policy on the economy, such as the effect of a change in interest rates on inflation and output. It is also used in financial analysis to understand the impact of changes in one market on other markets.

* 1. **VECM (Vector Error Correction Model):**

An extension of VAR that accounts for co-integration among the time series in the model. The objective is to capture the long-run relationships between time series, while also accounting for any short-run deviations from that relationship.

VECM (Vector Error Correction Model) model:

A VECM model extends the VAR model by incorporating co-integration between the time series, which is a long-term relationship between two non-stationary time series. The general form of a VECM model can be written as:

The coefficients in the VECM model can be estimated using maximum likelihood estimation (MLE) methods, and the residuals can be analyzed to check for stationarity and ensure that the model is a good fit for the data. The VECM model allows for a more accurate representation of the relationships between time series, especially when there is a long-term relationship between the series.

* + 1. Explain the VECM while giving examples
       1. VECM (Vector Error Correction Model) is a type of multivariate time series model that accounts for co-integration among the time series being modeled. It is an extension of the VAR (Vector Autoregression) model and it is used to capture the long-run relationships between the time series, while also accounting for any short-run deviations from that relationship.
       2. The model is useful for understanding the impact of one time series on another in the presence of co-integration, which refers to a long-term statistical relationship between two or more non-stationary time series. The VECM model uses a combination of lagged values and error correction terms in its regression analysis to capture this relationship.
       3. An example of the use of VECM would be to understand the relationship between the stock price of a company and its sales over time. If the stock price and sales are co-integrated, a long-term relationship exists between them, and a VECM model can be used to capture that relationship and the short-term deviations from it. The model can then be used to make forecasts and to understand how changes in one time series (such as sales) impact the other time series (such as stock price).

**Maximum Likelihood Estimation (MLE)** is a method for finding the parameter values that maximize the likelihood of observing the data. The likelihood function is defined as the probability of observing the data given the parameter values.

Given a random variable with a probability density function **,** where is the vector of parameters, the likelihood function is defined as:

The goal of MLE is to find the value of that maximizes the likelihood function. This value is found by using optimization techniques to solve for the value of **θ** that maximizes the likelihood.

The log-likelihood function is often used in place of the likelihood function, as it is easier to work with mathematically:

The optimization problem is then to find the value of that maximizes the log-likelihood function. This value is the maximum likelihood estimate (MLE) of **.**

1. **Volatility Models:**
   1. **ARCH (Autoregressive Conditional Heteroscedasticity):**

A type of time series model that captures the volatility (or fluctuations) of a time series over time. It is used to model and forecast the volatility of financial time series.

Autoregressive Conditional Heteroscedasticity (ARCH) is a time series model used to model the conditional volatility of a time series. It is commonly used to model financial time series, where volatility can change over time.

In an ARCH model, the variance of the error terms is modeled as a function of the past errors. The basic form of an ARCH model can be written as:

The objective of ARCH is to model the changing volatility of a time series, and to make predictions about future volatility. ARCH models can be used to estimate the volatility of a time series, and to adjust returns for volatility when analyzing financial data. ARCH models can also be used to develop more advanced models, such as GARCH (Generalized Autoregressive Conditional Heteroscedasticity) models.

* 1. **GARCH (Generalized Autoregressive Conditional Heteroscedasticity):**

An extension of the ARCH model that allows for more flexible modeling of volatility over time. The objective is to improve the accuracy of volatility forecasts.

Generalized Autoregressive Conditional Heteroscedasticity (GARCH) is an extension of the Autoregressive Conditional Heteroscedasticity (ARCH) model. It is used to model the conditional volatility of a time series, with the aim of capturing both short- and long-term volatility dynamics.

In a GARCH model, the conditional variance of the error term at time t is modeled as a function of both past error terms and past conditional variances:

The objective of GARCH is to provide a more complete and flexible model for capturing volatility dynamics than ARCH. GARCH models are widely used in finance, particularly for modeling and predicting volatility in stock, currency, and other financial markets.