Introduction to time series analysis and ARIMA model

Time series analysis is a statistical technique used to analyze and understand data points that are ordered chronologically over time. It involves modeling and analyzing the patterns, trends, and behaviors of the data over time. This can help identify important features or characteristics of the data, make predictions about future values, and uncover relationships between different variables.

Some common methods of time series analysis include:

Trend analysis: This involves analyzing the long-term trend or pattern in the data. This can help identify whether the data is increasing, decreasing, or staying relatively constant over time.

Seasonal analysis: This involves analyzing the pattern of data over the course of a seasonal cycle. This can help identify regular patterns or cycles that occur within the data.

Autocorrelation analysis: This involves analyzing the correlation between a time series and a lagged version of itself. This can help identify whether there are any repeating patterns or cycles in the data.

Forecasting: This involves using statistical models to make predictions about future values of the time series. One commonly used method for forecasting is the ARIMA model (AutoRegressive Integrated Moving Average), which is based on combining autoregressive and moving average models.

Overall, time series analysis can be a powerful tool for understanding and predicting patterns in data over time. It has applications in a wide range of fields, including finance, economics, meteorology, and engineering, among others.

Stationary and non-stationary time series:

A stationary time series is one whose statistical properties such as mean, variance, and autocorrelation remain constant over time. In other words, the distribution of the data does not change over time. A stationary time series is considered to be more predictable and easier to model than a non-stationary time series.

On the other hand, a non-stationary time series is one whose statistical properties change over time. In other words, the distribution of the data changes over time. Non-stationary time series often exhibit trends, seasonal patterns, or cycles. These patterns can make it difficult to model and predict future values.

Example of non-stationary time series data:

- A time series of monthly airline passenger traffic over a 10-year period. This series is likely to exhibit a trend and seasonality (e.g., more passengers in the summer months), making it non-stationary.

Example of stationary time series data:

- A time series of daily temperatures in a specific location over a 30-day period. If the temperature is generally stable over time with no significant upward or downward trend, and the variance of temperature is roughly constant over time, the series can be considered stationary.

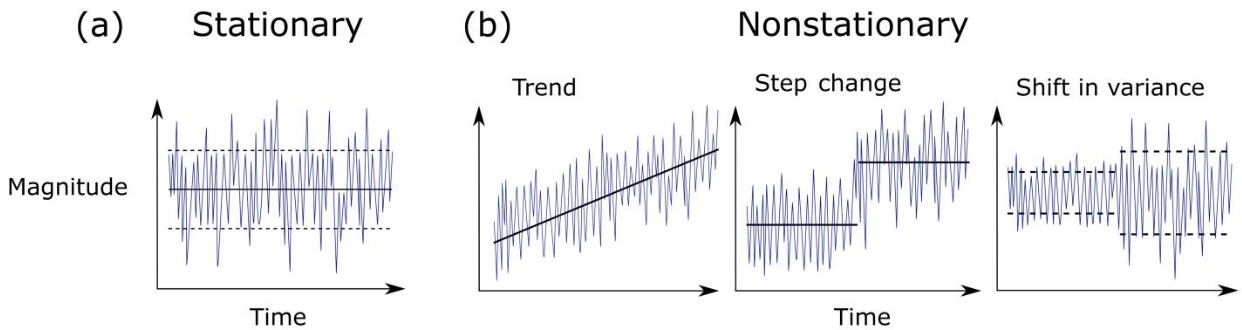

Figure 1: Examples of (a) a stationary time series with constant mean and variance and (b) three nonstationary time series in the form of a shift in mean (trend and step change) and a shift in variance. Solid and dashed black lines

To transform a non-stationary time series into a stationary time series, we can use techniques such as differencing, detrending, and seasonal adjustment. These techniques remove the trends and seasonality from the time series, leaving behind a stationary time series.

ARIMA Model

ARIMA (Autoregressive Integrated Moving Average) is a popular time series model used to analyze and forecast time-dependent data. It is a versatile model that can handle a wide range of time series data, including those with trends, seasonality, and other complex patterns.

The ARIMA model is based on three key components: autoregression (AR), differencing (I), and moving average (MA). Autoregression refers to the use of past values in the series to predict future values. Differencing involves taking the difference between consecutive observations to remove trends or seasonality. Moving average involves using past forecast errors to predict future values.

ARIMA models are commonly used in fields such as finance, economics, and engineering for a variety of applications, including stock market forecasting, weather forecasting, and demand forecasting.

Autoregression (AR) Model:

In an AR model, the value of the variable at a given time point is modeled as a linear combination of its past values. Specifically, the value at time t, denoted as Yt, can be expressed as:

$Y_t = \beta_0 + \beta_1 Y_{t-1} + \beta_2 Y_{t-2} + ... + \beta_p Y_{t-p} + \varepsilon_t$

where β_0, β_1, β_2, ..., β_p are the coefficients of the model and εt is an error term. The parameter p is the order of the model, representing how many past values are used to predict the current value.

Moving Average (MA) Model:

In an MA model, the value of the variable at a given time point is modeled as a linear combination of its past error terms. Specifically, the value at time t, denoted as Yt, can be expressed as:

$Y_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + ... + \theta_q \varepsilon_{t-q}$

where μ is the mean of the time series, εt is the error term at time t, and θ_1, θ_2, ..., θ_q are the coefficients of the model. The parameter q is the order of the model, representing how many past error terms are used to predict the current value.

The difference between AR and MA models is that AR models use past values of the variable itself to make predictions, while MA models use past error terms. AR and MA models can be combined to create more complex models, such as the ARMA (autoregressive moving average) model.

Differencing is a key component of ARIMA models that is used to transform non-stationary time series data into stationary data.

The process of differencing involves computing the difference between consecutive observations in the time series. This removes any trends or seasonality that may be present in the data, leaving behind a series of "differences" that is hopefully stationary.

For example, if we have a time series of stock prices that is increasing over time, we might take the first difference of the series (i.e., the difference between each consecutive observation) to remove the trend. If the resulting differenced series is stationary, we can use it as input to an ARIMA model.

It's important to note that not all non-stationary time series can be made stationary through differencing alone. In some cases, more complex transformations may be necessary, such as taking the logarithm or seasonal differencing. It's also important to visually inspect the differenced series to ensure that it is stationary before using it in an ARIMA model.

As a summary:

Steps for Building an ARIMA Model:

- Preprocess the data to ensure it is stationary (if not already)
- Determine the values of p, d, and q for the ARIMA model
- Fit the ARIMA model to the data
- Evaluate the model by examining the residuals and using statistical measures like AIC and BIC
- Use the model to make forecasts for future time periods

To learn more, please refer to the link below.

https://machinelearningmastery.com/arima-for-time-series-forecasting-with-python/