BHARATHIDASAN UNIVERSITY

Tiruchirappalli- 620024

Tamil Nadu, India.

Programme: M.Sc. Statistics

Course Title: Financial Statistics

Course Code: 23ST09DEC

Unit-III

Time Series Analysis

Dr. T. Jai Sankar
Associate Professor and Head
Department of Statistics

Ms. S. Soundarya
Guest Faculty
Department of Statistics

TIME SERIES ANALYSIS

The collection of observations obtained through repeated measurements over an interval of time is called time series analysis. The time series analysis is used to get meaningful statistics and other data characteristics. The 'Time Series Analysis' is distinct from other analyses because the ordering of observations is natural in this process.

Some examples of time series analysis include:

- Measurements of rainfall
- Monitoring the heart rate (Electrocardiograph)
- Recording the readings of temperatures
- Recording the data of weather
- Analysis of stock prices
- Recording the quarterly sales

Components of time series analysis

The components of time series analysis are the reasons which affect the value of an observation. Its components are:

The main components of time series analysis are trend, seasonality, and cyclic.

- **Trends of time series analysis** – During a long interval of time, the increase or decrease of the data shows the trend of time series analysis. The increase or decrease of observations in data is not always in the same direction in the given time period. The trend is a long-term, smooth and general tendency. Examples of trends are the number of deaths and births in a population, number of schools and colleges in an area, number of industries or factories in an area, etc. There are two types of trends: linear and non-linear.
- **Periodic Fluctuations –** The fluctuations that repeat themselves over an interval of time are called periodic fluctuations. There are two types of periodic fluctuations:

1. **Seasonal fluctuations –** These variations show the same pattern over a period of twelve months. The seasonal data are recorded hourly, monthly, daily, quarterly, or annually in time series analysis.
2. **Cyclic fluctuations –** These variations show the same pattern during a period of more than one year. In time series analysis, the oscillatory movement has a period of oscillation of more than a year. Cyclic fluctuations are sometimes also called the 'business cycle'.

- **Random or irregular movements –** The variation which is not regular or purely random is known as a random or irregular movement. Random movements are unpredictable or uncontrollable, like earthquakes, floods, disasters etc.
- **Mathematical model of time series analysis –** Mathematically, time series is expressed as Xt = f(t). Mathematical models of time series are of three types:

1. **Additive model for time series analysis.-** In the additive model, we represent a particular observation in a time series as the sum of these four components. where O represents the original data, T represents the trend. S represents the seasonal variations, C represents the cyclical variations and I represents the irregular variations.

$$y(t)=g(t)+s(t)+h(t)+\epsilon(t).$$

2. **Multiplicative model for time series analysis-** In the multiplicative model, the original time series is expressed as the product of trend, seasonal and irregular components. Under this model, the trend has the same units as the original series, but the seasonal and irregular components are unitless factors, distributed around 1.

$$Y(t) = Trend * Seasonal * Error.$$

3. **A mixed model for time series analysis**- Linear Mixed Models and Time Series analysis. where b_0 is the fixed-average intercept, j is the random (subject-specific) intercept for subject j , b is the vector of regression parameters, X_{ij} is the matrix of explanatory variables and ε_{ij} is the error for subject j at time i.

$$yt = Tt + St \times Ct \times Rt$$

Types of Time Series Analysis

The time series analysis includes variations of data. Its models of analysis include:

- Classification – It categorises the data and identifies the data
- Curve fitting – The data is plotted along the curve to study the relationships within the data
- Descriptive analysis identifies the time series data like tends, seasonal variations, etc
- Explanative analysis – It understands the data and its relationship with it
- Exploratory analysis – It tells the characteristics of time series data
- Forecasting – It predicts the future trend
- Intervention analysis – It describes how an event can change data
- Segmentation – It splits the data into two or more segments to show basic properties

Notation for time series data

- Y_t = value of Y in period t.
- Data set: $Y_1,..., Y_T$ = T observations on the time series random variable Y
- We consider only consecutive, evenly-spaced observations (for example, monthly, 1960 to 1999, no missing months) (missing and non-evenly spaced data introduce technical complications)

STATIONARY PROCESS

A stationary process in time series is a process where the statistical properties of the data do not change over time. This means that the mean, variance, covariance, and standard deviation are constant.

Stationary series is easier for statistical models to predict effectively and precisely.

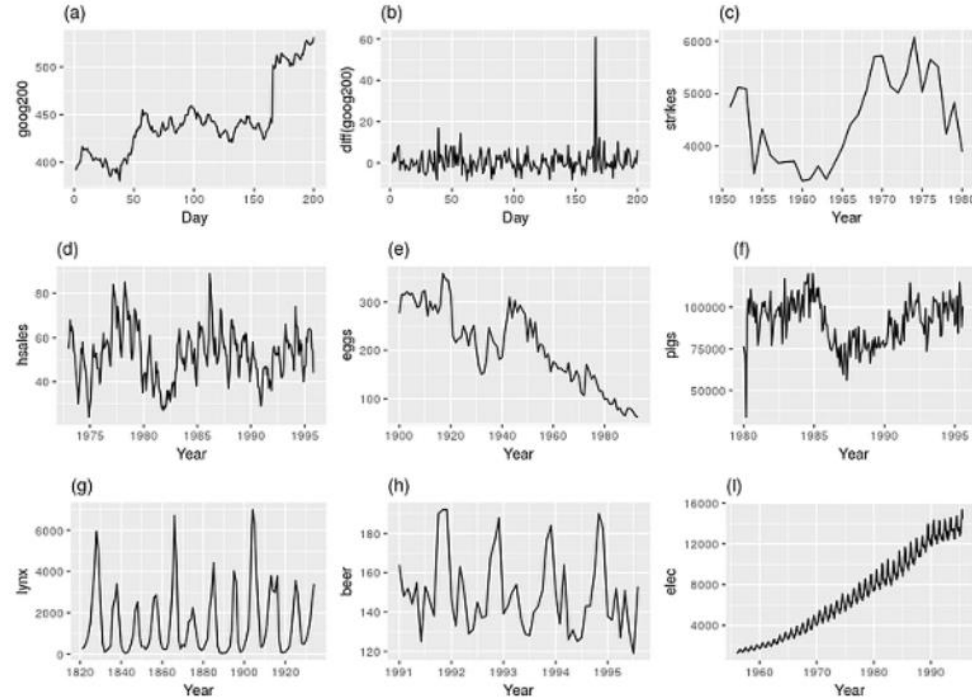
In data science, it is important to know about statistical tests, just as it is important to know about deep learning and machine learning algorithms. It helps us understand the data better and select forecasting models, like the ARIMA (Auto Regressive Integrated Moving Average) model or the SARIMA (Seasonal ARIMA) model.

Types of Stationary Series

1. **Strict Stationary** – Satisfies the mathematical definition of a stationary process. Mean, variance & covariance are not a function of time.
2. **Seasonal Stationary** – Series exhibiting seasonality.
3. **Trend Stationary** – Series exhibiting trend.

Visualizations

The most basic methods for stationarity detection rely on plotting the data and visually checking for trend and seasonal components. Trying to determine whether a stationary process generated a time series just by looking at its plot is a dubious task. However, there are some basic properties of non-stationary data that we can look for.



Nine examples of time series data; (a) Google stock price for 200 consecutive days; (b) Daily change in the Google stock price for 200 consecutive days; (c) Annual number of strikes in the US; (d) Monthly sales of new one-family houses sold in the US; (e) Annual price of a dozen eggs in the US (constant dollars); (f) Monthly total of pigs slaughtered in Victoria, Australia; (g) Annual total of lynx trapped in the McKenzie River district of north-west Canada; (h) Monthly Australian beer production; (i) Monthly Australian electricity production. [Hyndman & Athanasopoulos, 2018]

Seasonality can be observed in series (d), (h), and (i)

The trend can be observed in series (a), (c), (e), (f), and (i)

Series (b) and (g) are stationary

Conditions for stationarity

- **Constant mean**: The mean of the data does not change over time
- **Constant variance**: The dispersion of data points does not change over time
- **Constant autocorrelation**: The degree of similarity between the data and a lagged version of itself does not change over time.
- **No trend or seasonality**: The data does not have long-term trends or repeating patterns at fixed intervals.

Benefits of stationarity

- Stationary data is easier to model and predict
- Stationary data behaves predictably, which improves model accuracy

Non-stationary data

- Non-stationary data is data where the values and associations between variables change over time
- Non-stationary data is often transformed to become stationary

WEAK WHITE NOISE

Time series analysis, "weak white noise" refers to a sequence of random variables that are uncorrelated with each other, meaning there is no statistically significant relationship between any pair of observations, while also having a constant mean and variance; essentially, it's a time series with no discernible pattern or structure, where each data point is independent from the others and follows the same distribution.

Key points about weak white noise:

No autocorrelation:

The most important characteristic is that the autocorrelation function (ACF) of the series is zero at all lags, indicating no correlation between observations at different time points.

Constant mean and variance:

The mean and variance of the time series remain constant throughout the data.

Independence not required:

While "strong white noise" implies that the observations are completely independent and identically distributed (IID), "weak white noise" only requires the lack of correlation, not necessarily full independence.

Check for weak white noise:

Visual inspection: Plot the time series to see if there are any obvious patterns or trends.

Autocorrelation plot: Examine the ACF to see if the correlation coefficients are close to zero at all lags.

Statistical tests: Use statistical tests like the Ljung-Box test to formally assess whether the series can be considered white noise.

Example: A sequence of random numbers generated from a standard normal distribution (mean 0, variance 1) would be considered weak white noise.

Weak white noise in time series is a sequence of random variables that are uncorrelated and have a zero mean and finite variance. It's a mean-covariance process with a constant expectation and covariance over time.

Explanation

- **Uncorrelated**: The variables in a weak white noise sequence are not correlated with each other.
- **Zero mean**: The mean of the variables in a weak white noise sequence is zero.
- **Finite variance**: The variance of the variables in a weak white noise sequence is finite.

White noise is a key concept in time series analysis and forecasting. It's used as a benchmark to compare against, and to evaluate if there are patterns or predictable behavior in the data.

Examples of white noise

- **Gaussian white noise**

  A sequence of independent and identically distributed (IID) normal random variables with a zero mean and variance of σ^2

- **Binary white noise**

  A sequence of IID variables where n=1 with probability ½.

  Predicting white noise

White noise time series cannot be predicted because it is defined by zero correlation, constant variance, and zero mean.

Explanation

- A white noise time series is a series of independent and identically distributed (i.i.d.) elements with zero mean and variance.
- The autocorrelation function (ACF) of a white noise series has no significant spikes, except at lag zero.
- The plot of a white noise series shows no apparent pattern or direction.

How to test for white noise

- **Plot the time series**: Check that the average value is zero, the standard deviation is constant, and there are no distinct patterns.
- **Check the autocorrelation function**: Check that the ACF has no significant spikes, except at lag zero.
-

improve a model

- If forecast residuals are not white noise, you may be able to improve your model.
- Experiment with different techniques, evaluate the results, and refine your approach.
- Model the noise statistically to account for its presence and improve the accuracy of your models.

Estimating the parameters of a stationary process

To estimate the parameters of a stationary process time series, you can analyze the time series' stable properties, such as its mean, variance, and autocorrelation function. You can then use these properties to build a time-series model and make predictions.

Steps

- **Check for stationarity**: Use statistical tests like the Augmented Dickey-Fuller (ADF) test or the Kwiatkowski-Phillips-Schmidt-Shin (KPSS) test to check if the time series is stationary.
- **Transform the data**: If the time series is not stationary, you can transform it into a stationary series. For example, you can take the natural logarithm of the data.
- **Analyze the stable properties**: Analyze the mean, variance, and autocorrelation function of the stationary series.
- **Build a time-series model**: Use the stable properties to build a time-series model.
- **Make predictions**: Use the time-series model to make predictions.
- **Reverse the transformation**: If you transformed the data, reverse the transformation to get predictions for the original series.

Parameters

- **Alpha**: Used when the data doesn't have seasonality
- **Gamma**: Used when the data has a trend
- **Delta**: Used when the data has seasonality cycles

MOVING AVERAGE (MA) PROCESSES

Data is often collected with respect to time, whether for scientific or financial purposes. When data is collected in a chronological order, it is referred to as time series data. Analyzing time series data provides insights into how the data behaves over time, including underlying patterns that can help solve problems in various domains. Time series analysis can also aid in forecasting future values based on historical data, leading to better production, profits, policy planning, risk management, and other fields. Therefore, analysis of time series data becomes an important aspect of data science.

In this article, we will discuss Moving Average Models, which are essential for time series analysis and forecasting trends.

Moving Average Model

Moving Average Models are a type of time series analysis model usually used in econometrics to forecast trends and understand patterns in time series data. In moving average models the present value of the time series depends on the linear combination of the past white noise error terms of the time series. In time series analysis moving average is denoted by the letter "**q**" which represents the order of the moving average model, or in simple words we can say the current value of the time series will depend on the past q error terms. Therefore, the moving average model of order q could be represented as:

$$\mathbf{X_t} = c + \epsilon_t + \theta_1.\epsilon_{t-1} + \theta_2.\epsilon_{t-2} + \ldots + \theta_q.\epsilon_{t-q}$$

Here,

- **Xt** is the value of time series at time t
- **c** is a constant or the mean of the time series
- $\epsilon_t, \epsilon_{t-1}, \epsilon_{t-2}, \ldots, \epsilon_{t-q}$ are the white noise terms associated with the time series at time t, t-1, t-2, ... , t-q.
- $\theta_1, \theta_2, \ldots, \theta_q$ are the moving average constants.

For example, if we consider MA(1) model, in this model the present value of the time series will only depend on a single past error term and the time series becomes:

$$X_t = c + \epsilon_t + \theta_1.\epsilon_{t-1}$$

From this observation we can also conclude one of the most important aspects of moving average models that the higher the value of the order of moving average model (**q**), the model will have longer memory and dependence on the past values.

Interpretation of MA model:

There is a difference in the shock wave that is seen in the MA model and AR model that we can mention which might help us get a better understanding how MA and AR model differ. For a better understanding let's look at the AR model's general form as well:

$$X_t = c + \phi_1.X_{t-1} + \phi_2.X_{t-2} + ... + \phi_p.X_{t-p} + \epsilon_t$$

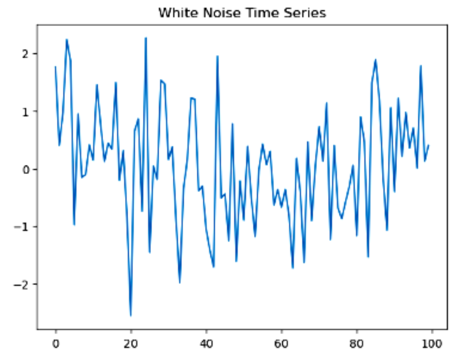
- First that the past noise term ϵ_{t-1} affects the MA model's present value Xt directly as we can see in the above equation of the MA model but in AR model the past noise term(ϵ_{t-1}) have an indirect influence on the present AR model value(Xt) since the AR model equation depends on the previous value of the model(X_{t-1}) and the previous model value depends on it's noise term(ϵ_{t-1}).
- The MA model works a a finite impulse model, which means that the current noise value affects the present value of the model as well as "q" further values, as the moving average models only depend on q terms in the past. Whereas AR models acts as infinite impulse model since the current noise affects infinite values of the model in the future. In AutoRegressive model ϵ_t value affects the X_t term which affects the X_{t+1} term and so on.

Concept Related to Moving Average:

Now let's discuss about some of the concepts that can help us in understanding the moving average model in a better way:

- **Stationarity**: Stationarity is the principle of time series data that conveys that the statistical properties of the data doesn't change with time, the mean of the data remains the same or we can also say that the data fluctuates around a certain value, the standard deviation of the time series data nearly remains constant, and there must not be any seasonality in the time series data or there is no periodic behavior in the data. We can check for the stationarity of the dataset visually as well as through Augmented Dickey-Fuller(ADF) Test**.** We consider stationarity to be one of the most important aspect that the time series data must possess in order to be accepted by the models that are applied to time series data for accurate modelling.
- **Differencing**: Differencing is one of the most important steps to consider during time series analysis, after taking a peek at the original time series data, if the data is not stationary and contains a lot of trends then differencing must be considered since for accurate time series data analysis the data must be stationary. In regular differencing the current time series data is subtracted by the previous data point. $\Delta y_t = y_t - y_{t-1}$, this method removes trends from the data, making it suitable for modelling.
- **White Noise**: White noise is the error term which has the mean of zero and a constant standard deviation with no correlation of the data points with each other. White noise acts

as a benchmark in the forecasting process through time series modelling, if the forecast error is nor white noise further modifications could be performed on the model, but if it reaches a state such that the forecast errors are white noise then the model would need no further improvements. The value of white noise series are random and unpredictable therefore if any time series data is a white noise then there is no method to model or forecast it.



White noise time series with mean = 0 and standard deviation = 1

- **ACF Plot**: Autocorrelation Function plot or the ACF plot is the plot of correlation between the time series and its lagged version. It shows how similar the time series is with it's different **lagged** values. Here the lag term is a fixed time displacement, in the ACF plot the x-axis is the lagged time series and the y-axis is the correlation which ranges from -1 to 1.

ARIMA PROCESSES

- ARIMA stands for Autoregressive Integrated Moving Average and it's a technique for time series analysis and for forecasting possible future values of a time series.
- Autoregressive modeling and Moving Average modeling are two different approaches to forecasting time series data. ARIMA integrates these two approaches, hence the name. Forecasting is a branch of machine learning using the past behavior of a time series to predict the one or more future values of that time series. Imagine that you're buying ice cream to stock a small shop. If you know that sales of ice cream have been rising steadily as the weather warms, you should probably predict that next weeks order should be a little bigger than this weeks order. How much bigger should depend on the amount that this weeks sales differ from last weeks sales. We can't forecast the future without a past to which to compare it, so past time series data is very important for ARIMA and for all forecasting and time series analysis methods.

- ARIMA is one of the most widely used approaches to time series forecasting and it can be used in two different ways depending on the type of time series data that you're working with. In the first case, we have create a Non-seasonal ARIMA model that doesn't require accounting for seasonality in your time series data. We predict the future simply based on patterns in the past data. In the second case, we account for seasonality which is regular cycles that affect the time series. These cycles can be daily, weekly, or monthly and they help define patterns in the past data of the time series that can be used to forecast future values. Like much of data science, the foundation of forecasting is having good time series data with which to train your models. A time series is an ordered sequence of measurements of a variable at equally spaced time intervals. It's important to remember that not all data sets that have a time element to it is actually time series data because of this equally spaced time interval requirement.

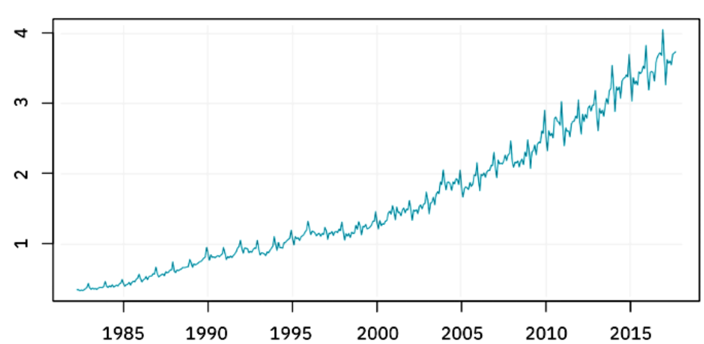
The Box-Jenkins Method

- In 1970 the statisticians George Box and Gwilym Jenkins proposed what has become known as the The Box-Jenkins method to fit any kind of time series model.[1] The approach starts with the assumption that the process that generated the time series can be approximated using a model if it is stationary. It consists of four steps:
- **Identification**: Assess whether the time series is stationary, and if not, how many differences are required to make it stationary. Then generate differenced data for use in diagnostic plots. Identify the parameters of an ARMA model for the data from auto-correlation and partial auto-correlation.
- **Estimation**: Use the data to train the parameters of the model (i.e. the coefficients).
- **Diagnostic Checking**: Evaluate the fitted model in the context of the available data and check for areas where the model may be improved. In particular this involves checking for overfitting and calculating the residual errors.
- **Forecasting**: Now that you have a model, you can begin forecasting values with your model.

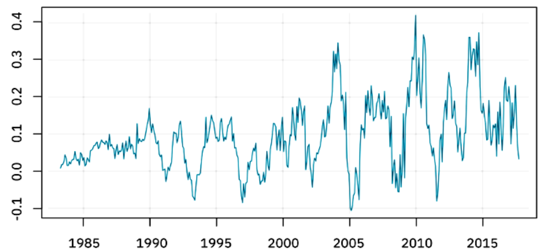
Once you’ve confirmed that your model fits your data correctly, you’re ready to begin ARIMA forecasting. We'll examine each of these steps in detail.

Characteristics of Time Series Data

- A time series can be stationary or non-stationary. A stationary time series has statistical properties that are constant over time. This means that statistics like the mean, variance, autocorrelation, don't change over the data. Most statistical forecasting methods, including ARIMA, are based on the assumption that the time series can be made approximately stationary through one or more transformations. A stationary series is comparatively easy to predict because you can simply predict that the statistical properties will be about the same in the future as they were in the past. Working with non-stationary data is possible but difficult with an approach like ARIMA.
- Another key feature of time series data is whether it has a trend present in the data. For instance, the prices of basic staples in a grocery store from the last 50 years would exhibit a trend because inflation would drive those prices higher. Predicting data that contains trends can be difficult because the trend obscures the other patterns in the data. If the data has a stable trend line to which it reverts consistently it may be trend-stationary, in which case the trend can be removed by just fitting a trend line and subtracting the trend from the data before fitting a model to it. If the data isn't trend-stationary, then it might be difference-stationary in which case the trend can be removed by differencing. The simplest way of differencing is to substract the previous value from each value to get a measure of how much change is present in the time series data. So for instance, if Y_t is the value of time series Y at period t, then the first difference of Y at period t is equal to Y_t-Y_{t-1}.
- Here we can see a plot of time series that's not stationary. It has an obvious trend upwards and exhibits seasonality.



The seasonality here is a regular 12 month cycle. This could be addressed by differencing the time series by 12 units so that we difference April 1990 with April 1989. After we apply differencing with a 12 unit lag to the time series, we can see a more stationary time series. The variance of this time series still changes but an ARIMA model could be fit to this time series and forecasts made using it.



Stationarity can be confusing, for instance, a time series that has cyclic behaviour but no trend or seasonality is still stationary. As long as the cycles are not of a fixed length when we observe the series we can't know where the peaks and troughs of the cycles will occur. Generally a stationary time series will have no predictable patterns in the long-term. If you were to plot the time series data in a line chart, it would look roughly horizontal with a constant variance and no significant spikes or drops.

Autocorrelation

We can see the degree to which a time series is correlated with its past values by calculating the auto-correlation. Calculating the auto-correlation can answer questions about whether the data exhibit randomness and how related one observation is to an immediately adjacent observation. This can give us a sense of what sort of model might best represent the data. The autocorrelations are often plotted to see the correlation between the points, up to and including the lag unit.

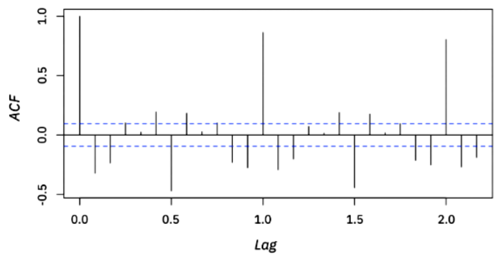
Each lag in the autocorrelation is defined as:

$$r_k = \frac{\sum_{t=k+1}^{T} \left(y_t - \bar{y}\right) - \left(y_t - k - \bar{y}\right)}{\sum_{t=1}^{T} \left(y_t - \bar{y}\right)^2}$$

r is any lag in the autocorrelation, T is the length of the time series, and y is the value of the time series. The autocorrelation coefficients make up the *autocorrelation function* or ACF.

In ACF, the correlation coefficient is in the x-axis whereas the number of lags

(referred to as the lag order) is shown in the y-axis. An autocorrelation plot can be created in python using plot_acf from the statsmodels library and can be created in R using the acf function.

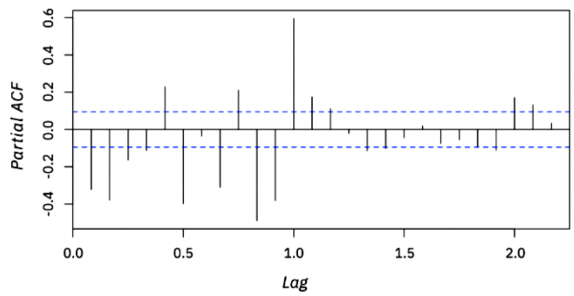


In this ACF plot of a time series differenced with a lag of 12 time units, the zero lag correlates perfect with itself. The first lag is negative, the second lag is slightly positive, the third lag is negative, and so on. You'll notice that the 12th lag is strongly correlated with itself. Since we were looking at monthly data, this makes sense. We can see that the auto-correlation maintain roughly the same cycle throughout the time series, an indication that our time series still contains significant seasonality. ACF plots are also useful for helping to infer the parameters of the ARIMA model that will best fit this data.

PARTIAL AUTOCORRELATION FUNCTION (PACF)

Another important plot in preparing to use an ARIMA model on time series data is the Partial Autocorrelation Function. An ACF plot shows the the relationship between yt and y_{t-k} for different values of k. If y_t and y_{t-1} are correlated, then y_{t-1} and y_{t-2} will also be correlated. But it's also possible for yt and yt−2 to be correlated because they are both connected to y_{t-1}, rather than because of any new information contained in y_{t-2} that could be used in forecasting y_t. To overcome this problem, we can use partial autocorrelations to remove a number of lag observations. These measure the relationship between y_t and y_{t-k} after removing the effects of lags 1 to k. So the first partial autocorrelation is identical to the first autocorrelation, because there is nothing between them to remove. Each partial autocorrelation can be estimated as the last coefficient in an autoregressive model.

Whether you're working in R or Python or another programming language or library, you'll have a way to calculate the PACF and create a PACF plot for easy inspection. An autocorrelation plot can be created in python using plot_pacf from the statsmodels library and can be created in R using the pacf function.



This PACF uses the same data as the above ACF plot. The PACF plot starts from 1 rather than 0 as in the ACF plot and shows strong correlations until the 1.0 lag, which correlates with same month of the previous year. After that first year, we see a decreasing amount of autocorrelation as the number of lags increases. Since we were looking at monthly data with a variance that changes year to year, this makes sense.

AUTOREGRESSION AND MOVING AVERAGE

As its name indicates, the acronym ARIMA integrates Autoregression and Moving Average models into a single model depending on the parameters passed. These two ways of modeling change throughout the time series are related but have some key differences. In an autoregression model, we forecast the variable of interest using a linear combination of past values of the variable. The term auto-regression indicates that it is a regression of the variable against itself. This technique is similar to a linear regression model in how it uses past values as inputs to the regression. Autoregression is defined as:

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \ldots \phi_q y_{t-q} + \epsilon_t$$

where ε_t is white noise. This is like a multiple regression but with lagged values of y_t as predictors. We refer to this as an AR(p) model, an autoregressive model of order p.

A moving average model on the other hand uses the past forecast errors rather than using past values of the forecast variable in a regression. A moving average simply averages k values in a window, where k is the size of the moving average window, and then advances the window. The forecast values are evaluated using the actual values to determine the error at each step in the time series. A moving average is defined as:

$$y_t = C + \epsilon_t + \theta_1 \epsilon_{t-1} + \theta_2 \epsilon_{t-2} + \ldots \theta_q \epsilon_{t-q}$$

ε_t is white noise. We refer to this as an MA(q) model, a moving average model of order q. Of course, we do not observe the values of ε_t, so it is not really a regression in the usual sense.

Notice that each value of y_t can be thought of as a weighted moving average of the past few forecast errors.

Typically in an ARIMA model you'll use either the Auto-Regressive term (AR) term or the Moving

Average term (MA). The ACF plot and PACF plot are oftentimes used to determine which one of these terms is most appropriate.

Specifying an ARIMA model

Once the time series has been made stationary and the nature of the auto-correlations have been determined, it's possible to fit an ARIMA model. There are 3 key parameters for an ARIMA model which are typically referred to as p, d, and q.

p: the order of the Autoregressive part of ARIMA

d: the degree of differencing involved

q: the order of the Moving Average part

These are typically written in the following order: ARIMA(p, d, q). Many programming languages and packages will provide an ARIMA function that can be called with the time series to be analyzed and these three parameters. Most often the data is split into a train set and a test set so that accuracy of the model can be tested after it has been trained. It is usually not possible to tell just from looking at a time plot what values of p and q will be most appropriate for the data. However it is oftentimes possible to use the ACF and PACF plots to determine appropriate values for p and q and thus those pllots are important terms for working with ARIMA

A rough rubric for when to use AR terms in the model is when:

- ACF plots show autocorrelation decaying towards zero
- PACF plot cuts off quickly towards zero
- ACF of a stationary series shows positive at Lag - 1

A rough rubric for when to use MA terms in the model is when:

- Negatively Autocorrelated at Lag - 1
- ACF that drops sharply after a few lags
- PACF decreases gradually rather than suddenly

There are a few classic ARIMA model types that you may encounter.

ARIMA(1,0,0) = first-order autoregressive model: if the series is stationary and autocorrelated, perhaps it can be predicted as a multiple of its own previous value, plus a constant. If the sales of ice cream for tomorrow can be directly predicted using only the sales of ice cream from today, then that is a first-order autoregressive model.

ARIMA(0,1,0) = random walk: If the time series is not stationary, the simplest possible model for it is a random walk model. A random walk is different from a list of random numbers because the next value in the sequence is a modification of the previous value in the sequence. This is often how we model differenced values for stock prices.

ARIMA(1,1,0) = differenced first-order autoregressive model: If the errors of a random walk model are autocorrelated, perhaps the problem can be fixed by adding one lag of the dependent variable to the prediction equation i.e., by regressing the first difference of Y on itself lagged by one period.

ARIMA(0,1,1) without constant = simple exponential smoothing models: This is used for time-series data with no seasonality or trend. It requires a single smoothing parameter that controls the rate of influence from historical observations (indicated with a coefficient value between 0 and 1). In this technique, values closer to 1 mean that the model pays little attention to past observations, while smaller values stipulate that more of the history is taken into account during predictions.

ARIMA(0,1,1) with constant = simple exponential smoothing models with growth. This is the same as simple exponential smoothing except that there is an additive constant term that makes the Y value of the time series grow as it progresses.

There are many other ways that ARIMA models can be fit of course, which is why we often calculate multiple models and compare them to see which one will provide the best fit for our data. All of these are first order models which means that they map linear processes. There are second order models which map quadratic processes and higher models that map more complex processes.

Comparing ARIMA models

Typically multiple ARIMA models are fit to the data and compared with one another to find which one beset predicts that patterns seen in the time series data. There are three key metrics to assess the accuracy of an ARIMA model:

Akaike’s Information Criterion or AIC. This is widely used to which to select predictors for regression models, and it's also useful for determining the order of an ARIMA model. AIC quantifies both the goodness of fit of the model and the simplicity/parsimony of the model in a single statistic. A lower AIC score is better than a higher one, so we would prefer the model that has a lower score. AIC favors simpler models, more complex models receive higher scores as long as their accuracy is roughly the same as a simpler model. There is also the corrected AIC or AICC which simply has a small correction applied for the sample size.

Bayesian Information Criterion or BIC. This is another criterion for model selection that penalizes complexity even more than the AIC. As with the AIC, models with lower BIC are generally preferred to those with higher scores. If your model is going to be used for longer term forecasting, the BIC may be preferable, whereas shorter term forecasting may mean that the AIC is preferable.

The sigma squared or sigma2 value is the variance of the model residuals. The sigma term describes the volatility of the hypothesized process. If you have highly volatile data but a very low sigma squared score, or conversely non-volatile data but a high sigma squared score, that is a sign that the model isn't capturing the actual data generating process well.

If we have held back a test data set then we can also compare accuracy metrics like RMSE for different prediction intervals. The ARIMA model can forecast values for a single time step in the future or for multiple steps at a time.

Variations of ARIMA

One other approach to configuring and comparing ARIMA models is to use Auto-ARIMA, which applies automated configuration tasks to generating and comparing ARIMA models. There multiple ways to arrive at any optimal model. The algorithm will generate multiple models and attempt to minimize the AICc and the error of the Maximum Likelihood Estimation to obtain an ARIMA model.

Seasonal Autoregressive Integrated Moving Average, SARIMA or Seasonal ARIMA, is an extension of ARIMA that supports time series data with a seasonal component. To do this, it adds three new hyper parameters to specify the autoregression, differencing and moving average for the seasonal component of the series, as well as an additional parameter for the period of the seasonality. A SARIMA model is typically expressed SARIMA((p,d,q),(P,D,Q)), where the lower case letters indicate the non-seasonal component of the time series and the upper case letters indicate the seasonal component.

Vector Autoregressive Models (or VAR Models) are used for multivariate time series. They are structured so that each variable is a linear function of past lags of itself and past lags of the other variables.

ARIMA models are a powerful tool for analyzing time series data to understand past processes as well as for forecasting future values of a time series. ARIMA models combine Autoregressive models and Moving Average models to give a forecaster a highly parameterizable tool that can be used with a wide variety of time series data.

AIC AND SBC

Akaike Information Criterion (AIC) and Schwarz Bayesian Criterion (SBC) are metrics used to compare and evaluate ARIMA models. Both metrics penalize models with more parameters to prevent over fitting. A lower AIC or SBC value indicates a better model.

Explanation

- **AIC**: Focuses on minimizing prediction error.
- **SBC**: Also known as Bayesian Information Criterion (BIC), this metric places a stronger penalty on model complexity.

Uses of AIC and SBC

1. Fit multiple ARIMA specifications.
2. Calculate the AIC or SBC for each model.
3. Select the model with the lowest AIC or SBC value.

Why use AIC and SBC?

- AIC and SBC help balance goodness-of-fit with model simplicity.
- AIC and SBC help prevent over fitting.
- AIC and SBC help select the best model for forecasting performance.

AIC and SBC

- AIC and SBC are model selection criteria based on the log-likelihood
- **Akaike's information criterion (AIC)**: is defined as $-2 \log(L) + 2(p + q)$;

  L is the likelihood evaluated at the MLE
- **Schwarz's Bayesian Criterion (SBC)**:
  - defined as $-2 \log(L) + \log(n)(p + q)$,
  - n is the length of the time series
  - also called Bayesian Information Criterion (BIC)

- "best" model by either criterion is the model that minimizes that criterion
- Either criteria will tend to select models with a large likelihood value
  - s makes perfect sense since large L means observed data are likely under that model
- term $2(p + q)$ in AIC or $\log(n)(p + q)$ is a penalty on having too many parameters
- therefore, AIC and SBC try to tradeoff
  - good fit to the data measured by L
  - the desire to use few parameters
- which penalizes the most?
  - $\log(n) > 2$ if $n >= 8$
  - most time series are much longer than 8
  - so SBC penalizes $p + q$ more than AIC
- Compared to SBC, with AIC the tradeoff is more in favor of a large value of L than a small value of $p + q$.
- It can be shown that $\text{Log}(L) \approx (-n/2)\log(\sigma^2) + K$ where K is a constant that does not depend on the model or on the parameters.
- Since we only want to minimize AIC and SBC, the exact value of K is irrelevant and we will drop K.
- Thus, you can use the approximations:
  - $\text{AIC} \approx n\log(\sigma^2) + 2(p + q)$
  - $\text{SBC} \approx n\log(\sigma^2) + \log(n)(p + q)$
- σ^2 is called MSE (mean squared error).
- Difference between AIC and SBC is due to the way they were designed.
- AIC is designed to select the model that will predict best and is less concerned with having a few too many parameters.
- SBC is designed to select the true values of p and q exactly.
- In practice the best AIC model is usually close to the best SBC model often they are the same model.
- Models can be compared by likelihood ratio testing when one model is "bigger" than the other.
- Therefore, AIC and SBC are basically LR tests.