

Time Series Forecast

Essential of Time-Series forecast, Forecasting models



January 7, 2020

EISTI

Arjun SINGH Murail Mopidevi krishna

Contents

[Time-Series Forecast 2](#_Toc29588064)

[1. Fundamentals: 2](#_Toc29588065)

[2. Definition and Examples 2](#_Toc29588066)

[3. Uses of Time Series 2](#_Toc29588067)

[4. Components for Time Series Analysis 3](#_Toc29588068)

[5. Mathematical Model for Time Series Analysis 5](#_Toc29588069)

[6. Trend Measurement Methods 6](#_Toc29588070)

[ Freehand method(Graphical method):- 6](#_Toc29588071)

[ Semi-average method:- 6](#_Toc29588072)

[ Moving Average method:- 7](#_Toc29588073)

[ Method of Least Square 7](#_Toc29588074)

[ Exponential Smoothening 8](#_Toc29588075)

[7. Terminologies associated with Time Series 8](#_Toc29588076)

[ The Lag operators 8](#_Toc29588077)

[ Autocovariance Function 8](#_Toc29588078)

[ Autocorrelation Function 9](#_Toc29588079)

[ Stationarity 9](#_Toc29588080)

[8. Forecasting Models 9](#_Toc29588081)

[ Difference Equation 9](#_Toc29588082)

[ Moving Averages (MA) Models 14](#_Toc29588083)

[ Autoregressive Models (AR) Models 16](#_Toc29588084)

[Time Series Analysis 19](#_Toc29588085)

[ Dataset 1: Murders in USA 19](#_Toc29588086)

[ Dataset 2: Divorce/Annulment in USA 22](#_Toc29588087)

[ Dataset 3: Housing 25](#_Toc29588088)

[ Dataset 4: Brazil Imports 29](#_Toc29588089)

[ Dataset 5: USA exports to Canada 32](#_Toc29588090)

[ Dataset 6: Air passenger 36](#_Toc29588091)

[ Dataset 7: Death 40](#_Toc29588092)

[ Airpass using AR & MA Model 43](#_Toc29588093)

# Time-Series Forecast

## Fundamentals:

An arrangement of statistical data in accordance with time of occurrence or in a chronological order is called a time series. The numerical data, which we get at different points of time-the set of observations is known as time series.

In time series analysis, current data in a series may be compared with past data in the same series. We may also compare the development of two or more series over time. These comparisons may afford important guidelines for the individual firm. In Economics, statistics and commerce it plays an important role.

## Definition and Examples

A time series is a set of observations made at specified times and arranged in a chronological order. For example, if we observe agricultural production, sales, National Income etc., over a period, say over the last 3 or 5 years, the set of observations is called time series.

Thus, a time series is a set of time, quantitative readings of some various recorded at equal intervals of time. The interval may be an hour, a day, a week, a month, or a calendar year. Hourly temperature reading, daily sales in a shop, weekly sales in a shop, weekly sales in a market, monthly production in an industry, yearly agricultural production, population growth in ten years, are examples of time series.

From the comparison of past data with current data, we may seek to establish what development may be expected in future. The analysis of time series is done mainly for the purpose of forecasts and for evaluating the past performances. The chronological variations will be object of our study in time series analysis.

The essential requirements of a time series are:

* The time gap, between various values must be as far as possible, equal.
* It must consist of a homogeneous set of values.
* Data must be available for a long period.

Symbolically if “*t*” stands for time and “*Yt*” represents the value at time t then the paired values (t, yt) represents a time series data.

## Uses of Time Series

The analysis of time series is of great significance not only to the economists and businessman but also to the scientists, astronomists, geologists, sociologists, biologists, research worker etc. In the view of following reasons

* It helps in understanding past behavior.
* It helps in planning future operations.
* It helps in evaluating current accomplishments.
* It facilitates comparison.

## Components for Time Series Analysis

The various reasons or the forces which affect the values of an observation in a time series are the components of a time series. The four categories of the components of time series are

* Trend
* Seasonal Variations
* Cyclic Variations
* Random or Irregular movements



Seasonal and Cyclic Variations are the periodic changes or short-term fluctuations.

**Trend**

The trend shows the general tendency of the data to increase or decrease during a long period of time. A trend is a smooth, general, long-term, average tendency. It is not always necessary that the increase or decrease is in the same direction throughout the given period of time.

It is observable that the tendencies may increase, decrease or are stable in different sections of time.  But the overall trend must be upward, downward or stable. The population, agricultural production, items manufactured, number of births and deaths, number of industry or any factory, number of schools or colleges are some of its example showing some kind of tendencies of movement.



* **Linear and Non-Linear Trend**

If we plot the time series values on a graph in accordance with time t. The pattern of the data clustering shows the type of trend. If the set of data cluster more or less round a straight line, then the trend is linear otherwise it is non-linear (Curvilinear).

* **Periodic Fluctuations**

There are some components in a time series which tend to repeat themselves over a certain period of time. They act in a regular spasmodic manner.

**Seasonal Variations**

These are the rhythmic forces which operate in a regular and periodic manner over a span of less than a year. They have the same or almost the same pattern during a period of 12 months. This variation will be present in a time series if the data are recorded hourly, daily, weekly, quarterly, or monthly.

These variations come into play either because of the natural forces or man-made conventions. The various seasons or climatic conditions play an important role in seasonal variations. Such as production of crops depends on seasons, the sale of umbrella and raincoats in the rainy season, and the sale of electric fans and A.C. shoots up in summer seasons.

The effect of man-made conventions such as some festivals, customs, habits, fashions, and some occasions like marriage is easily noticeable.  They recur themselves year after year. An upswing in a season should not be taken as an indicator of better business conditions.

**Cyclic Variations**

The variations in a time series that operate themselves over a span of more than one year are the cyclic variations. This oscillatory movement has a period of oscillation of more than a year. One complete period is a cycle. This cyclic movement is sometimes called the ‘Business Cycle’.

It is a four-phase cycle comprising of the phases of prosperity, recession, depression, and recovery. The cyclic variation may be regular are not periodic. The upswings and the downswings in business depend upon the joint nature of the economic forces and the interaction between them.

**Random or Irregular Movements**

There is another factor, which causes the variation in the variable under study. They are not regular variations and are purely random or irregular. These fluctuations are unforeseen, uncontrollable, unpredictable, and are erratic. These forces are earthquakes, wars, flood, famines, and any other disasters.

## Mathematical Model for Time Series Analysis

In classical analysis, it is assumed that some type of relationship exists among the four components of time series. Analysis of time series requires decomposition of a series, to decompose a series we must assume that some type of relationship exists among the four components contained in it.

yt = f (t)

Here, ytis the value of the variable under study at time t. If the population is the variable under study at the various time period t1, t2, t3, … , tn. Then the time series is

t: t1, t2, t3, … , tn

yt: yt1, yt2, yt3, …, ytn

or, t: t1, t2, t3, … , tn

yt: y1, y2, y3, … , yn

The value Yt of a time series at any time t can be expressed as the combinations of factors that can be attributed to the various components. These combinations are called as models and these are primarily two types.

**Additive Model for Time Series Analysis**

If ytis the time series value at time t. Tt, St, Ct, and Rt are the trend value, seasonal, cyclic and random fluctuations at time t respectively. According to the Additive Model, a time series can be expressed as

**yt = Tt + St + Ct + Rt.**

This model assumes that all four components of the time series act independently of each other.

**Multiplicative Model for Time Series Analysis**

The multiplicative model assumes that the various components in a time series operate proportionately to each other. According to this model **yt = Tt × St × Ct × Rt**

## Trend Measurement Methods

* Freehand method (Graphical, method)
* Semi-average method
* Moving Average method
* Method or Least Square
* Exponential Smoothening

### Freehand method(Graphical method):-

The freehand method is the simplest of all the methods for measuring the trend. Under this method, the original data are plotted on a graph paper and a trend line is fitted by inspection. The trend line or curve should be drawn through the data in such a way that the areas below and above the trend are equal. They should be exactly equal for the series as a whole and approximately equal for the first half and last half of the series separately and as per as possible for each major cycle.

Advantages

* The freehand method is the simplest of all the methods for measuring the trend. It is a non-mathematical method of trend measurement and as such, it can be easily understood by most of the people.
* No calculations are involved in this method. Therefore, it saves time and can be employed when a quick result is desired.

Disadvantages

* Since no mathematical formula is used so in order to fit the trend line, different people may draw a different trend line from the same data.
* The trend line depends on the judgment of the investigator. So, it can be affected by personal bias.

### Semi-average method:-

According to this method, the original data are divided into two equal parts the values of each part are summed up and averaged. The average of each part is centered in the period of time of the part from which it has been calculated and plotted on the graph. A straight line shall then be drawn to pass through the plotted points. This line constitutes the semi-average trend line. When the numbers of years are odd the middle years are not considered while the data divided into two equal parts and obtained averages. The semi-average method is sometimes employed when a straight line appears to be an inadequate explanation of the trend.

Advantages

* This method is simple and there is no probability of personal prejudice and bias affecting the result.
* For a particular series, there will be only one trend line.

Disadvantages

* It assumes that there is a linear relationship between the plotted points which may not be true in all cases.
* Since we have to secure averages, so this method this affected by extreme values.
* This method does not eliminate seasonal and cyclic fluctuation.
* It does not work well if data is given for a long period.

### Moving Average method:-

Moving Average method is a simple device of reducing fluctuations and obtaining trend values with a fair degree of accuracy. When a trend is to be determined by this method, the average value for a number of years (months or week) is secured. Then this average is taken as the normal or trend value for the unit of time falling in the middle of the period covered in the calculation of the period. In the moving average method determination of the period of moving average (i.e 3 years, 4years, 7or 8 years etc.) is very important as trend values are affected by the period of the moving average. The basic principle should be that the period of moving average is equivalent to the period of the cycle.

Advantages

* The greatest advantage of this method is that it eliminates the short time fluctuation that may be present in the time series.
* This method is simple to understand and does not require the use of any complex mathematical calculation.
* If a few more values are added to the time series then it simply results in a few more trend values which can be easily obtained without distributing the previous calculation.
* This method is not subjective because the choice of the period of moving average is determined by the oscillatory movements of the data and not by the whims of the statistician.

Disadvantages

* The main disadvantage of this method is that it does not provide trend values for all the terms. There are not trend values for some time periods in the beginning and at the end of the series.
* If the fluctuation in the time series is irregular then it is difficult to determine the period of the moving average.
* The method of moving average is developed under the assumption that the trend line is linear. Thus, for time series with a linear trend (which is generally the case in the economics and business) these methods either overestimate or underestimate the trend values.
* This method cannot be used for predicting or forecasting, which is the main objective for trend analysis because it does not put forward any mathematical relationship between the variate under study and time.

### Method of Least Square

The method of least squares is a mathematical device which places a line through a series of plotted points in such a way that the sum of the square of the deviations of the actual points above and below the trend line is at a minimum. If we sum up the positive and negative deviations on either side of the fitted line, the sum will be zero. Thus, the sum of the squares of these deviations obtained will be least compared to the sums of the squares of the deviations obtained by using another line. It is due to the reason that the method is known as the method of least squares.

Advantages

* This method is based on the fitting of a mathematical equation, which makes it highly objective, and different people calculating the trend will reach the same result.
* In this method, we determine the relationship in such a way that the sum of the deviation is minimum.
* This method can be used for estimating future or past values.
* This method can be used for computing the trend values for all the given time periods in the series.
* In this method, the estimates of the constants (i.e. a and b) are unbiased and have minimum variance.

Disadvantages

* This method is relatively difficult to understand and involves complex calculation, especially when an exponential or quadratic trend equation is to be fitted.
* With the inclusion of one or more values in the series all the calculation have to be done afresh.
* Though the method is used for prediction, such prediction will be successful only if the same situation continuous to prevail during the period of estimation.
* Though the method is objective, the choice of the type of curve to be fitted is a subjective matter. So, different people may choose different types of curve namely, linear, exponential, quadratic etc. for analysis of trend.

### Exponential Smoothening

Exponential smoothing is a technique used to detect significant changes in data by considering the most recent data. Also known as averaging, this method is used in making short-term forecasts.

Given that there are many other ways to make forecasts, what makes exponential smoothing better in certain cases compared to others? Also, what makes it not ideal for certain scenarios?

Advantage

* It is easy to learn and apply.
* It produces accurate forecasts..
* It gives more significance to recent observations.

Disadvantage

* It produces forecasts that lag behind the actual trend.
* It cannot handle trends well.

## Terminologies associated with Time Series

### The Lag operators

Lxt = xt-1 , yt = xt-1, In other words Lxt is the value of the time series at time t-1

### Autocovariance Function

Originally, autocorrelation/auto covariance function is used to estimate the dominant periods in the time series. The autocovariance is the covariance of a variable with itself at some other time, measured by a time lag (or lead) τ.



### Autocorrelation Function

The Autocorrelation function is the normalized auto covariance function

It shows us how much a time series is similar to a time shifted version of itself



### Stationarity

Stationarity Stationarity requires that the process is an a particularly state of equilibrium, that is when its statistical property are non dependent from the time. Weakly stationarity: the properties involved only the rst (mean) and the second moment (variance and autocovariance strictly stationarity: when the join probability distribution at any set of times t1,t2, . . . ,tm must be the same as the joint probability distribution at time t1 + k,t2 + k, . . . ,tm + k

Sample moments

* E(x1)=E(x2)=. . .=E(xt)=µ
* V(x1)=V(x2)=. . .=V(xt)=σ 2
* Cov(x1, x1+k )=Cov(x2, x2+k )
* autocorrelation: ρk = γk/γ0

## Forecasting Models

* Difference Equation: is an expression relating a variable Y(t) to its previous values

Yt = ɸYt-1 + Wt ----D(1.1)

**First order linear difference equation**

It is because only the first lag of variable Y(t) (i.e Y(t-1)) appears in the equation and it expresses Yt as linear function of Y(t-1) & w(t)

**Goldfield’s Example**



Where





It = The log of Aggregate real Income

rbt = The log of the interest rate of bank account

rct = The log of the interest rate of the commercial paper

With the description that D(1.1) governs the behavior of Y for all the dates t, we have an equation which relates Yt with previous value Yt-1 and the current value of Wt



That means that y taken on a date t can be described as fundamentals of the intial value Y(-1) and the history of the inputs from 0………………..1

So if we know the starting value of y at t = -1

If we know the starting value of у for date t = -1 and the value of w for dates t = 0, 1, 2, . . . , then it is possible to simulate this dynamic system to find the value of у for any date. For example, if we know the value of у for t = -1 and the value of w for t = 0, we can calculate the value of у for t = 0 directly from y(0). Given this value of y0 and the value of w for t = 1, we can calculate the value of у for t = 1

У(1) = ФУ(0) + W1 = Ф(ФУ-1 + w1) + w,

or

У(1) = Ф2\*У(-1) + ФWo + W1

Continuing recursively in this fashion, the value that у takes on at date t can be described as a function of its initial value y(-1) and the history of w between date 0 and date t:



This procedure is known as solving the difference equation by recursive substitution.

**Dynamic Multipliers**

Expressing y(t), as a linear function of the initial value y^1 and the historical values of W. This makes it very easy to calculate the effect of W(0) on y(t). If W(0) were to change with y(-1) and w1,w2, . . . , w, taken as unaffected, the effect on y, would be given by



Dynamic calculations were to remain same if we were to start at time t





Now the effect on y(t+j) is

Dynamic multiplier considered against Goldfields money demand specification

For instance we want to know what will happen to money demand two quarters from now if current income It were to increase by 1 unit & future income I(t+1) & I(t+2) remains unaffected

Now change Dynamic Multipliers

🡪 

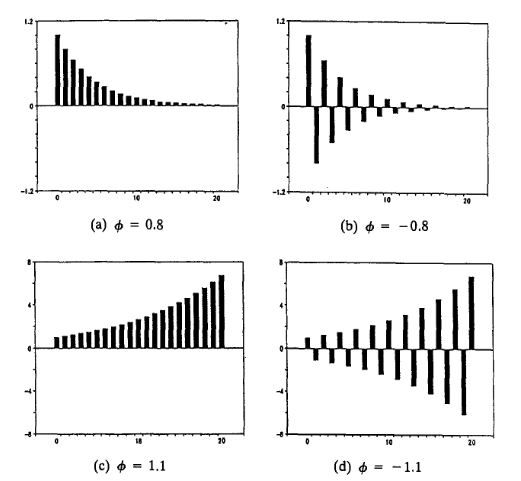
Since I(t) is the log of income 0.001 means 1% increase in income. An increase in mt of 1% is 0.01\* 0.098 ~ 0.001 coressponds to 0.001 or 0.1% increase in income

Different value of Ф produces value of dynamic responses of Y to w



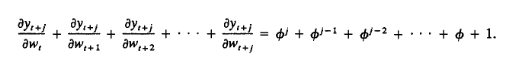
That means a One unit increase will cause a permanent one-unit increase in value of y





In general if |Ф| <1 system is stable, if |Ф| >1 system is explosive

Sometimes we might be interested in the consequences of permanent change

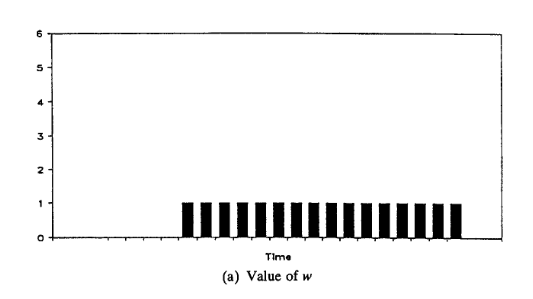
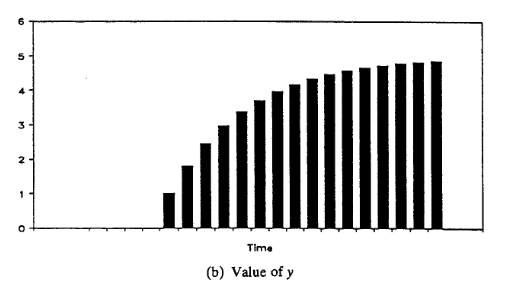


When |Ф| <1 the limit of this expression as j goes to infinity is sometimes described as the long run effects of w on y:

1+ Ф2+ Ф3+ Ф4 + Ф5+ …………………. = 1/(1- Ф)

In Goldfields case Ф = 0.72 hence for long run income elasticity of money demand in the system is given by 1/(1- Ф) \* 0.19(change in I(t)) = 0.68

A permanent 1% increase will lead to 0.68% increase in money demand



**Pth order linear difference equation**

Generalize the dynamic system of 1.1.1 by allowing the value of y at date t to depend on p of its own lags among with the current value of input variable w(t)



A linear Pth difference equation

It is convenient to write pth order difference equation in scalar to vector first order difference equation

|  |  |
| --- | --- |
| First element the value Y(t) took on date t  Second Element the value Y took on date t-1 and so on |  |
| P x P matrix F |  |
|  | |

|  |  |
| --- | --- |
| In order to find dynamic multiplier | |
|  | |
| Recursively Proceeding | |
|  | |
| Considering the first equation of this system which categorises value of Y(t). Let f11t  denotes (1,1) of matrix Ft , f12t denotes (1,2) of matrix Ft . Hence equation reduces to    In First order only one initial value but in pth difference equation p initial values for y | |
| Similarly we can do recursive subsitution to arrive at date t+j as well    From which we have      Thus Dynamic Multiplier  where f11t = 1st element(1,1) of F i.e ɸ1 | |
| If j =1 | If j = 2 f112 🡺 F2 🡺[] [] resulting in new f11 from matrix |
| Moving Averages (MA) Models The moving average process of finite order is considered an approximation to the Wold representation that happens to be a moving average process of infinite order. Various sorts of shocks in a time series drive all variations.  frm-Moving-Averages-MA-Models  **The First-Order Moving Average (MA(1)) Process**  The process is defined as:    In the general MA process, and particularly the MA(1) process, a function of current and lagged unobservable shocks expresses the current value of the observed series. This is an important feature that generally defines the MA process.  The following is the equation for the unconditional mean:    And the unconditional variance is:    An increase in the absolute value of θ causes the unconditional variance to increase, given that the value of σ is constant.  The next step is to calculate the autocorrelation of the MA(1) process. We start by calculating the autocovariance function. That is:    Therefore, the autocorrelation function is defined as:    This function happens to be the autocovariance function scaled by the variance.  The Finite-Order Moving Average Process of order q, MA(q), Process  For MA(q) process, we have that: | | |
|  | | |
| Autoregressive Models (AR) Models This another approximation to the World representation. The autoregressive process is a simple stochastic difference equation. In discrete time-stochastic dynamic modeling, the natural vehicle is the stochastic difference equations.  The AR(1) process  The following equation is the AR(1) for short, in the AR(1) process:    It can also be expressed in the lag operator form as follows:    It is also important to note that a finite-order moving average process is always covariant stationary. However, for invertibility, certain conditions have to be met. But for autoregressive process invertibility always exist. However, covariance stationarity in the autoregressive process requires some conditions to be satisfied.  For the AR(1) process:    Then on the right hand side backward substitution for the lagged yy’s is done to obtain:    And this can be expressed in the following manner in the lag operator form:    For convergence to exist in this moving average representation for yy, then |φ|<1. Therefore, in the AR(1) process, the condition for covariance stationarity is |φ|<1.  The unconditional mean can be calculated as:    And the unconditional variance is calculated as:    For the autocovariances we have:    Both sides of the equation are multiplied by yt−τyt−τ, such that:    For τ≥1,when we take expectations of both sides we obtain:    **AR(p) Process**  The following is the equation of a general ppth order autoregressive process, AR(p):    This can also be expressed in the following way, as a lag operator:    Covariance stationarity in the AR(pp) process occurs iff all the roots of the autoregressive lag operator polynomial Φ(L)Φ(L) have inverses that fall inside the unit circle. Here, the process can be written in the form of a convergent infinite moving average:    At displacement p, the cutoff for the AR(p) partial autocorrelation is sharp. | | |

# Time Series Analysis

## Dataset 1: Murders in USA

**The following table represents the number of Murders in the United States(in Thousands) for the years intervals from 1985 to 1995**

In this its evidently clear that Polynomial trend of degree 2 fits my data better as it has higher value of R^2 we can go higher with degree but to avoid the problem of overfitting , we evaluated till degree 2 only

|  |
| --- |
|  |
| This is evaluation of different techniques for trend analysis by doing Moving Average, centered one and the thing is of all the techniques applied these three MA(4) CMA(4) & MA(5) performed quite similar to data |
|  |
| The Weighted Moving Average(5) and Centered MA(4) are exactly same trend(overlapping) it is quite clear once you look at the data and the curve as well. For Exponential Smoothening, The Lower Values almost result in Linear trend but does not represent the trend on other hand higher value result in almost same trend as actual trend |
|  |
| Semi Average Mean gives a better result which better represents my data. Since the arithmetic mean is greatly affected by extreme values, it is subjected to misleading values, and hence the trend obtained by plotting by means might be distorted. However, if extreme values are not apparent, this method may be successfully employed. |
|  |
| Polynomial degree 2 trend does justice to the data but we can also use MA(5) and otherCMA,WMA but problem with Moving Averages method is loosing values at extreme ends and the incline/decline in the data might already happen before moving average takes that into account |

## Dataset 2: Divorce/Annulment in USA

The following dataset shows the number of divorces and annulments (in Thousands) in the United State for the Years 1986-1995

|  |
| --- |
|  |
| In this Analysis, we had to do go until Polynomial trend deg5 to find that curve which is oscillating in nature |
|  |
| The Inference of the following moving Average is that as wee increase the Size of Moving Average we smoothen the curve to obtain a new trend but the problem is fluctuating data in this case higher order moving average flattens it completely into almost straight line and we lose lot of information on both sides which is not good with trend to lose that much data, we want to retain information from original as well as smoothen it as well |
|  |
| in this its evidently clear the centered Moving average is giving me better result but compared to exponential smoothening which looses no data at from other end its recursive formulation allows the new forecast to be determined |

|  |
| --- |
|  |
| Since the arithmetic mean is greatly affected by extreme values, it is subjected to misleading values, and hence the trend obtained by plotting by means might be distorted. However, if extreme values are not apparent, this method may be successfully employed. |
|  |
| Comparing this with the Moving Averages & Exponential smoothening, latter gives better result as it smooths the curve without losing data and smoothening the trend in process |

## Dataset 3: Housing

|  |
| --- |
|  |
| Seasonal Index graph for Housing dataset, Ratio-to-trnd Median gives a little spikes but rest of index are almost similar |
|  |
| The Trend for the Housing Data, an incremental Trend |

|  |
| --- |
|  |
| We remove the seasonality from the time series to (Y/S = T\*C\*I) and the deseasonalize data is there and we remove |
|  |
| After De trending(Y/(S\*T) ) we have still the spikes in data. We use Moving Averages to smoothen the data and get rid of spikes |
|  |
| MA(3), MA(7) to smoothen the data to remove spikes and get the best |
|  |
| MA(3),MA(7) smoothening technique to smoothen the cyclicity and irregularities |
|  |
| The best forecast is obtained using Linear trend and seasonality obtained by Ratio-to-Trend(mean) |

## Dataset 4: Brazil Imports

|  |
| --- |
|  |
| The Dataset for Imports to Brazil, lot of spikes and smoothened using various trend measurement methods |
|  |
|  |
|  |
| Seasona index with consistent seasons in my data |
|  |
| Removing the seasons but trends are still very much evident |
|  |
| Detrending the data gives a very nice smooth curve and we can work with this data without smoothening |
|  |
| The forecast of this is not very good, but the best one in the models selected. We obtained the best one for Ratio-to-trend Mean with a linear trend |

## Dataset 5: USA exports to Canada

The following Dataset shows the monthly new Housing starts (in Thousands) for the United States from January 1990 all the way to December 1995

|  |
| --- |
|  |
| The different Seasonal Indexes Obtained using different methods |
|  |
| The two trend we are using for this case, Linear and degree2 trend and they have a very close accuracy as well |
|  |
| Deseasonalized data but the trend is visible, incremental trend. |
|  |
| We detrend the deseasonalized data and obtain a curve (Y/T\*S = C\*I) and we smoothen C\*I |
|  |
|  |
|  |
| We smoothen the detrended data to remove the irregularities and flatten cyclicity of the curve |
|  |
| The Best Method for forecast is obtained by using Y = (~T(linear) \* S(MA(12))) |

## Dataset 6: Air passenger

|  |
| --- |
|  |
| Airline Passenger Data for 12 Years with increasing trend and various seasons |
|  |
| Seasonal indexes for the Airpassengers |
|  |
| Deseasonalization is done but the trend still remans and its and increasing trend that needs to be balanced |
|  |
| Detrending the data and we have a very good stable curve but the existence of spikes because of remaining two components of Time-Series (cyclicity & irregularities) |
|  |
|  |
|  |
| Smoothening the detrended data to remove the spikes(flatten cyclicities) |
|  |
| The new forecast fits very close to actual data and we obtain it using ( Y = ~T(Linear)\*S(ratio-to-Trend mean)) |

## Dataset 7: Death

|  |
| --- |
|  |
| The trend for the death database is declining due to improved medical healthcare facilities(declining trend but seasonal) |
|  |
| Seasonal index for Death dataset, High during Summer( warm Summers could be one reason) |
|  |
| Deseasonalized but the trend exists we need to remove that as well |
|  |
| Detrending the data and that parabolic curve trend disappears but cyclicity is there and we have to remove that using MA(3) & MA(7) |
|  |
| The final forecast its almost the same and a very close approximate obtained using Y = (~T(Polynomial)\*S(ratio-to-tren median)) |

## Airpass using AR & MA Model

|  |
| --- |
| data("AirPassengers")  print(AirPassengers)  is.ts(AirPassengers)  # yep it is a TimeSeries data  summary(AirPassengers)  Min. 1st Qu. Median Mean 3rd Qu. Max.  104.0 180.0 265.5 280.3 360.5 622.0  ts.plot(AirPassengers, xlab="Year", ylab="Number of Passengers", main="Monthly totals of international airline passengers, 1949-1960")  # This will fit in a line  abline(reg=lm(AirPassengers~time(AirPassengers)))  #check for Stationarity  adf.test(diff(log(AirPassengers)), alternative="stationary", k=0)  # Augmented Dickey-Fuller Test  # data: diff(log(AirPassengers))  # Dickey-Fuller = -9.6003, Lag order = 0, p-value = 0.01  # alternative hypothesis: stationary |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\airpass_time_series.png |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\Airpass_linearT.png |

|  |  |
| --- | --- |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\acf.png | C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\pacf.png |
| ACF (Auto Correlation) | PACF (Partial Auto Correlation) |

AR Fit

|  |
| --- |
| ***(p, d, q)* are the AR order, the degree of differencing, and the MA order.**  **AR <- arima(AirPassengers, order = c(1,0,0))**  print(AR)  # Call:  # arima(x = AirPassengers, order = c(1, 0, 0))  #  # Coefficients:  # ar1 intercept  # 0.9646 278.4649  # s.e. 0.0214 67.1141  #  # sigma^2 estimated as 1119: log likelihood = -711.09, aic = 1428.18  **#Using predict() to make a 1-step forecast**  predict\_AR <- predict(AR)  **#Obtaining the 1-step forecast using $pred[1]**  predict\_AR$pred[1]  [1] 426.5698  **#ALternatively Using predict to make 1-step through 10-step forecasts**  predict(AR, n.ahead = 10)  # $`pred`  # Jan Feb Mar Apr May Jun Jul Aug Sep Oct  # 1961 426.5698 421.3316 416.2787 411.4045 406.7027 402.1672 397.7921 393.5717 389.5006 385.5735  #  # $se  # Jan Feb Mar Apr May Jun Jul Aug Sep Oct  # 1961 33.44577 46.47055 55.92922 63.47710 69.77093 75.15550 79.84042 83.96535 87.62943 90.90636 |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\airpass_ARfit.png |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\airpass_AR_model_forecast.png |

MA Fit

|  |
| --- |
| ***(p, d, q)* are the AR order, the degree of differencing, and the MA order.**  **MA <- arima(AirPassengers, order = c(0,0,1))**  print(MA)  # arima(x = AirPassengers, order = c(0, 0, 1))  #  # Coefficients:  # ma1 intercept  # 0.9642 280.6464  # s.e. 0.0214 10.5788  #  # sigma^2 estimated as 4205: log likelihood = -806.43, aic = 1618.86  **#Forcasting using MA model**  **#Making a 1-step forecast based on MA**  predict\_MA <- predict(MA)  **#Obtaining the 1-step forecast using $pred[1]**  predict\_MA$pred[1]  # [1] 425.1049  **#ALternatively Using predict to make 1-step through 10-step forecasts**  predict(MA, n.ahead = 10)  #$`pred`  # Jan Feb Mar Apr May Jun Jul Aug Sep Oct  # 1961 425.1049 280.6464 280.6464 280.6464 280.6464 280.6464 280.6464 280.6464 280.6464 280.6464  # $se  # Jan Feb Mar Apr May Jun Jul Aug Sep Oct  # 1961 64.84895 90.08403 90.08403 90.08403 90.08403 90.08403 90.08403 90.08403 90.08403 90.08403 |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\airpsass_MA(1).png |
| C:\Users\Administrator\AppData\Local\Microsoft\Windows\INetCache\Content.Word\airpass_MA(predict).png |

**Best Model**

#Choosing AR or MA Goodness of fit

cor(AR\_fit, MA\_fit)

#[1] 0.954995

|  |  |
| --- | --- |
| # Find AIC of AR  AIC(AR)  ## [1] 1428.179 | # Find AIC of MA  AIC(MA)  ## [1] 1618.863 |
| # Find BIC of AR  BIC(AR)  ## [1] 1437.089 | # Find BIC of MA  BIC(MA)  ## [1] 1627.772 |

#Given the lower value of AIC and BIC in AR model,

#we should go with that for the time series analysis of AirPassenger data.

**Best predicted Model**

library(forecast)

AutoArimaModel=auto.arima(AirPassengers)

AutoArimaModel

Series: AirPassengers

ARIMA(2,1,1)(0,1,0)[12]

# Coefficients:

# ar1 ar2 ma1

# 0.5960 0.2143 -0.9819

# s.e. 0.0888 0.0880 0.0292

# sigma^2 estimated as 132.3: log likelihood=-504.92

# AIC=1017.85 AICc=1018.17 BIC=1029.35