# 1 Modeling Time Series Data for Forecasting

All business and organizations do forecasting. Starting from the roadside vendors and peddlers to major airlines and hotels as well as drug manufacturers and governments forecast for the next several years to plan and optimize their outputs, maximize profits and minimize losses, plan for manpower resources and overall smoothly run their businesses. Of course forecasting methods are not the same for small retailers and pharmaceuticals running drug trails all over the world. In fact in many cases, even sizable organizations do not employ statistical methods for forecasting, but go by their gut-feeling or at best employ naïve methods.

A good forecasting method takes into account the systematic changes in the data. There are many different methods of forecasting. Forecasting involves predicting for one or more important variables, or responses, for future. Regression techniques can also be used as a forecasting method, if it may be assumed that the predictors are all known for future. However, possibly the most used method of forecasting is time-series, where one or more variables are predicted for future, depending on their past performances. In this chapter we will deal with univariate time-series only, where a single variable is studied and forecasted for future.

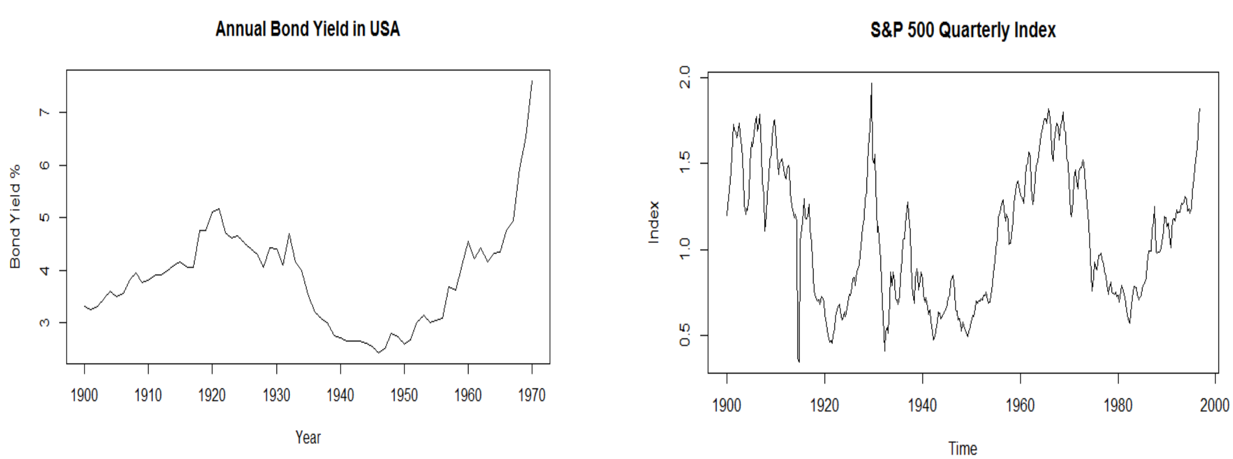
Forecasting will never be perfect; the challenge is to make a business forecast model better than the competition. Better forecasts result in better customer service and lower costs, as well as better relationships with suppliers and customers. The forecast can, and must, make sense based on the big picture, economic outlook, market share, and so on. The best way to improve forecast accuracy is to focus on reducing forecast error. Bias is the worst kind of forecast error. Aim of all forecasters is to make accurate forecasts with less cost and effort. To achieve the best possible predictions one has to learn about present scenario, make realistic assumptions about future and change assumptions and study their impact on future values. Forecasts are to be benchmarked against naïve method forecasts as well as industry standards.

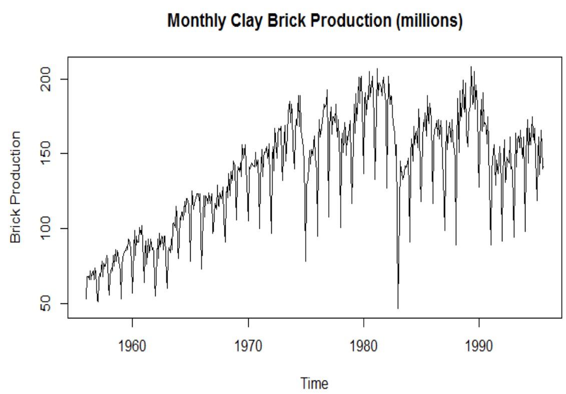
## 2 Characteristics and Components of Time Series Data

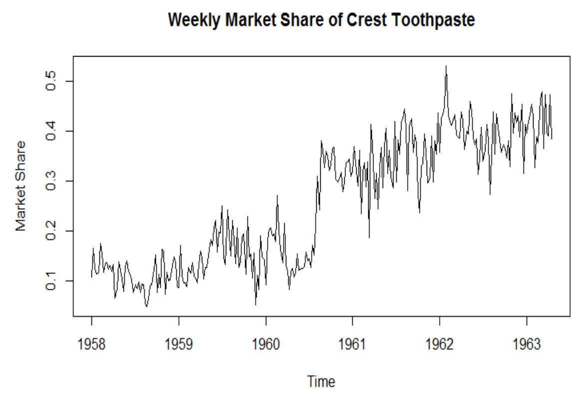
A time series is defined as a set of quantitative observations arranged in chronological order. We generally assume that time is a discrete variable. Time series observations may be collected for all time points. To measure a country's growth Gross National Product is measured at yearly level. Profits and losses of companies' are measured at every quarter. Sales volumes are measures monthly. Stock market movements are followed daily. All of the above examples are time series data. Even minute-by-minute movement of stock vale can be modelled using time series. The two most important characteristics of time series data are that, all observations must be collected at the same interval and observations must be contiguous. In the same time series yearly and monthly observations must not be combined. Neither there be any gaps in the range of the time points for which the data are collected.

Following are several examples of time series data collected at various time intervals – yearly, quarterly, monthly and weekly. Time series data can also be collected daily or hourly or at any other precisely measurable intervals. Time is assumed to be discrete for collection of time series data.

FIGURE 1:






Naturally chronologically observed variables are not independent. Hence standard procedures for independent set of observations will not work for time series. It is also clear from the above figures that over time, the observations show definitive movements – increasing or decreasing- or cyclical movements that are repeated yearly. These distinct patterns are characteristics for time series data.

The major components of Time Series are

- Trend
- Seasonality
- Cyclical and
- Irregular component

Trend is the long term increasing or decreasing movement in a time series. Every series shows a tendency to either move in a certain direction or stagnate over time. The series of Annual Bond Yield shows a definite increasing trend post 1950. The series has an increasing trend, whereas the S&P 500 Quarterly series does not show any trend over the years considered. The Weekly Market Share proportion of Crest Toothpaste shows an upward movement all through the six years under consideration.

Seasonality indicates fluctuations within a year which is not random in nature. The series on monthly clay brick production shows similar movement within a year, irrespective of whether there is a trend or not. The repetitive fluctuation within a year is known as seasonal movement of a time series, since this regular pattern is associated with seasons within a year. Seasonal patterns are for time series with intervals less than a year.

Cyclicality of a time series is associated with ups and downs associated with a business across years. Schematically a business may go through a sequence of starting-up, prosperity, decline and finally, recovery. Such long-term movements are known as cyclical patterns. This may also be noticed in retail business where demands for particular items may show come-back after regular intervals. Study of cyclical pattern is more difficult than the other two and we will omit this component from our discussion.

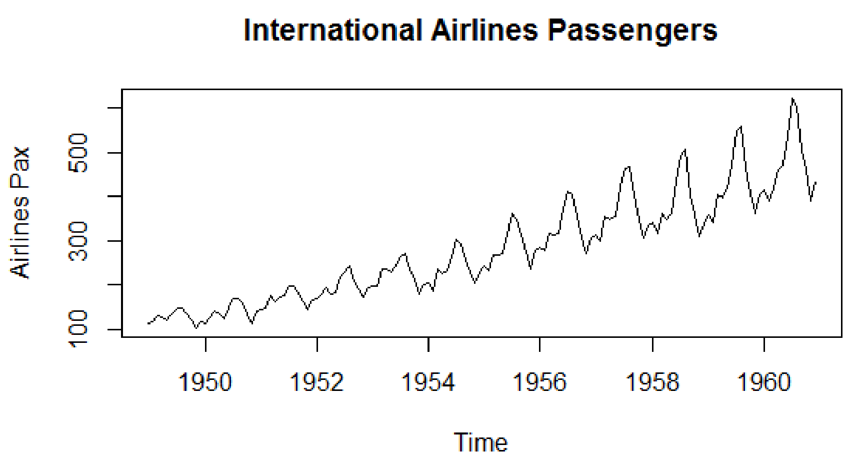
Irregular or random component of a time series is in addition to these three components, which are caused by sudden shocks to the series. These patterns are not repetitive and do not show any pattern. This component is similar to the error component in a regression model and is also known as the noise.

If Y(t) is the value of a time series at time t, then Y(t) may be expressed as a sum or as a product of the three components.

- Additive time series: $Y(t) = T(t) + S(t) + I(t)$
- Multiplicative time series: $Y(t) = T(t) \times S(t) \times I(t)$

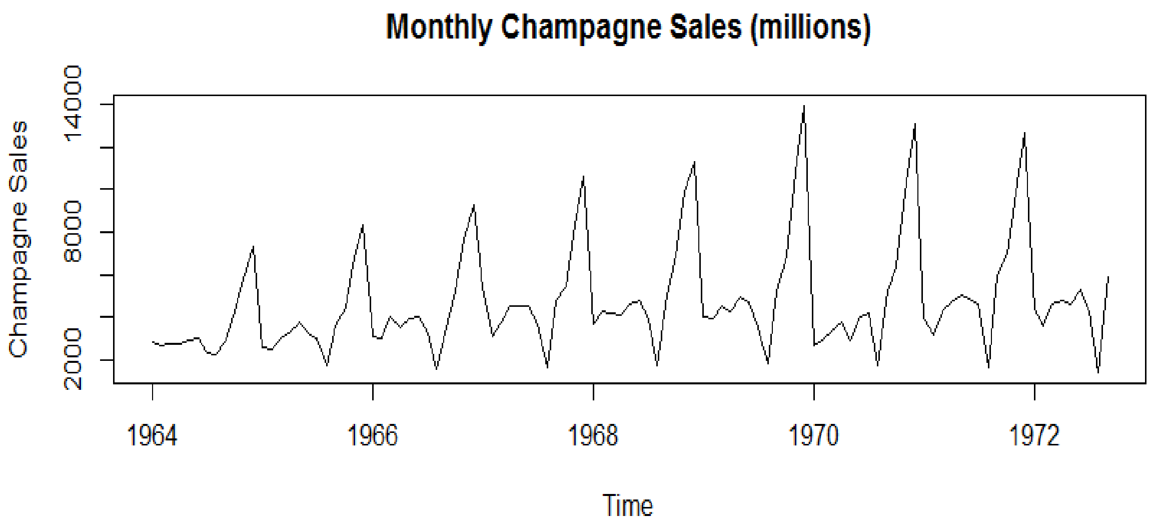
where T(t) is the trend value of the series at time t, S(t) is the seasonality index at time t and I(t) is the value of the irregular or error component associated with the observation at t-th time point. Multiplicative model is based on the assumption that the components trend and seasonality are not independent, but show an interaction effect. Typically this is noticed when seasonality factors are higher with larger trend values. The example below indicates that the total number of international airlines passengers is increasing over the years and along with the variation in the number of passengers across months is also showing higher fluctuations.

FIGURE 2:



In comparison to that, the sales of champagne seem to be constant over the years, as shown in the figure below. The seasonality effect, however, shows higher fluctuations in recent years. The series on air traffic is a possible case for multiplicative time series while the series of champagne sales would possibly be better modelled as an additive series.

FIGURE 3:




## 3 Time Series Analysis Techniques

Two distinct features of time series are clear from the above discussion. Each observation in a time series may be comprised of trend, seasonality and error component. Each observation in a time series depends on the previous one or more observations. Depending on these two characteristics, two different techniques may be employed to analyse time series. Main goal in analysing a time series is to obtain an understanding of the underlying forces and structure that produced the observed data and to fit a model and proceed to forecasting, monitoring or even feedback and feed-forward control. The classical time series analysis procedure decomposes the series into its constituent components. The regression approach models each observation as a function of previous observations.

Whatever may be the ultimate analysis techniques, the important questions to consider when first looking at a time series are:

- Is there a trend, meaning that, on average, the measurements tend to increase (or decrease) over time?
- Is there a seasonality, meaning that there is a regularly repeating pattern of highs and lows related to calendar time such as seasons, quarters, months, days of the week, and so on?

We have already considered these questions for Airline Passengers data and for Champagne Sales data. The former shows both trend and seasonality while the latter shows only seasonal effect.

- Are their outliers? In regression, outliers are way off from the fitted line. In time series data, outliers are suddenly different from the other data points

As in regression, outliers are difficult to identify. In the S&P 500 Quarterly series there may be an outlier for 1929 Q3. For the series of Clay Brick production January 1981 may be an outlier. Outliers are sometimes incorporated in the history of the time series. Consider example of passenger traffic through an airport, which suddenly closed down on a certain day due to heavy fog or snowfall. That day passenger volume will be extraordinarily low, if not exactly zero. This observation is definitely an outlier and must be treated as such from the beginning.

- Is the variance constant over time?

This is also a difficult assumption to test. Variation of the irregular component, which is the random error component, causes the fluctuations in the time series. At the same time, fluctuations are caused by seasonality also. Typically daily movement of stock exchanges or such high volatile data show signs of non-constant variance. Special models (ARCH or GARCH) have been developed for those series, which are out of the purview of our discussion here.

- Are there any abrupt changes to either the level of the series or the variance?

This is a very real problem and has a large implication in the course of the course of time series analysis. Consider a production process and output of that process is the series of interest. If the process is modernized at some point of time and as a result has become more efficient, a sudden change in the level of outputs is expected to be noted. If this information is not available to the analyst, and if the information before the modernization is used to model the series, the forecasts may not be the most accurate. Weekly Market Share series of Crest toothpaste seems to show a clean break in the middle part of 1980. However, it is difficult to detect exactly at which point the break occurred. Similarly it is possible that around 1973-73 there is a change in the nature of the Clay Brick production series.