|  |
| --- |
| Sharath Chandra, Nabha Subramanya  9-16-2020 |

Contents

[1. Objective 3](#_Toc51174865)

[2. Why Forecasting? 3](#_Toc51174866)

[3. Exploring Data Patterns 5](#_Toc51174867)

[4. Data Prep 7](#_Toc51174872)

[5. Time Series Imputation 8](#_Toc51174875)

[6. Different Time Series Approaches 8](#_Toc51174881)

[7. Accuracy Measures 13](#_Toc51174899)

[8. Prediction Intervals 13](#_Toc51174903)

[9. OSG Time Series Dashboard 14](#_Toc51174904)

[10. Pros & Cons of Forecast 15](#_Toc51174905)

[11. References 16](#_Toc51174906)

# Objective

The objective of this white paper is to increase awareness of forecasting approaches mostly used in OSG, when to use a model, how to tune parameters & how to measure forecast accuracy.

# Why Forecasting?

Forecasting is frequently used in predictive analytics and has wide applications including demand planning. It has significant impact on both topline and bottom line of an organization. Different forecasting problems include demand/supply planning for service or product, price forecasting, manpower planning etc.

Assumptions and estimations generally produce reasonably accurate forecasts for the immediate future, but assumptions become weaker predictors of the future as the time horizon lengthens.

However, forecasts can be made more reliable if the assumptions and estimations used are supported by preexisting data that is based on solid *market behavior drivers*. By using data that considers why consumers of your products behave as they do, the reliability of forecasts increase.

Therefore, it is important for businesses to establish a forecast early, refine it as intelligence is gathered regarding their understanding of customer behavior drivers, and iterate to achieve a higher degree of accuracy. Assessing the quality of past forecasts with respect to their accuracy and consistency with gathered intelligence improve the quality of future forecasts.

Hint: While starting a project, having a good understanding of business requirement, what factors drive behavior and impact the independent variable is crucial. Also collecting data and ensuring validity of the data is important factor.

Below are a few examples of Forecasting done at OSG:

# Exploring Data Patterns

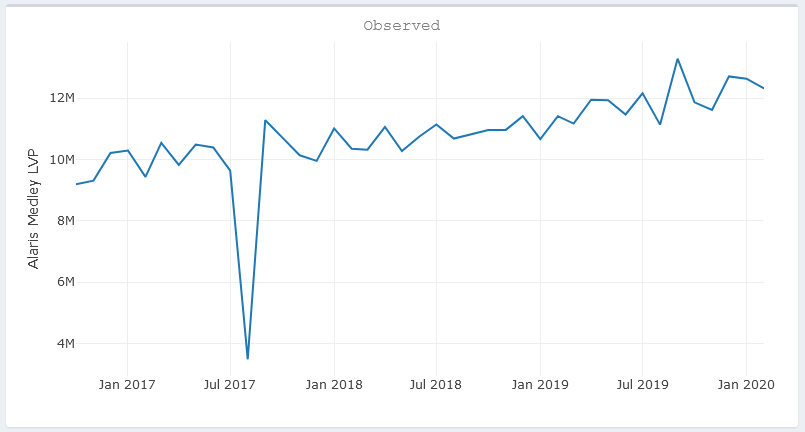
Studying the patterns in the data improves the forecaster’s chances of successfully [modelling data for forecasting applications](http://bit.ly/2oMZbiq). Through exploratory data analysis (EDA), a demand forecaster can start the important task of finding factors (drivers of demand) that are generally quantitative in nature. A planned forecasting and modelling effort that does not include provisions for exploratory data analysis often miss the most interesting and important results; but it is only a first step, not the whole story

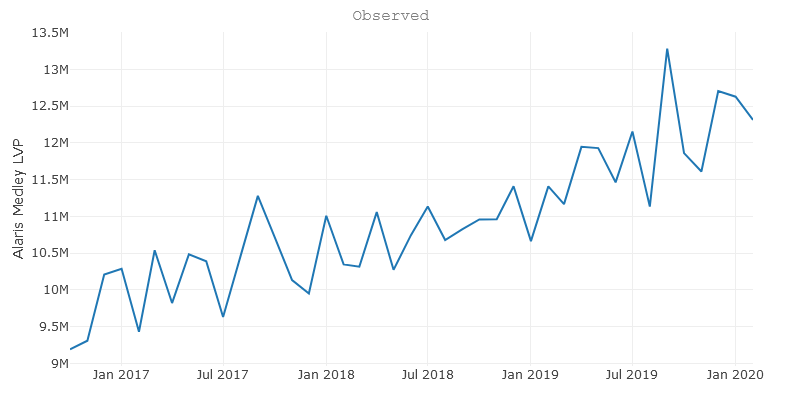
Exploratory data analysis means looking at data, absorbing what the data are suggesting, and using various summaries and display methods to gain insight into the process generating the data.

## Basic Analysis

The first and foremost analysis is to start with a simple line chart which will give insights into multiple aspects of the pattern i.e. Trend, Seasonality, cycle-city, Peaks & Troughs and outliers.

Hint: Spend time understanding the data, the causality of peaks and troughs that occur. Understand the outlier e.g. if there is a huge spike in sales – was it driven by a discount or a promotion or a data glitch and what additional factors are influencing that peak.





Understand Data Errors, make corrections

Understand Cause for peak and the dip following it

Observe the trend

*🡨 Aug 17 outlier value Imputed, with Average of previous & next month values, explained in Section 5.1*

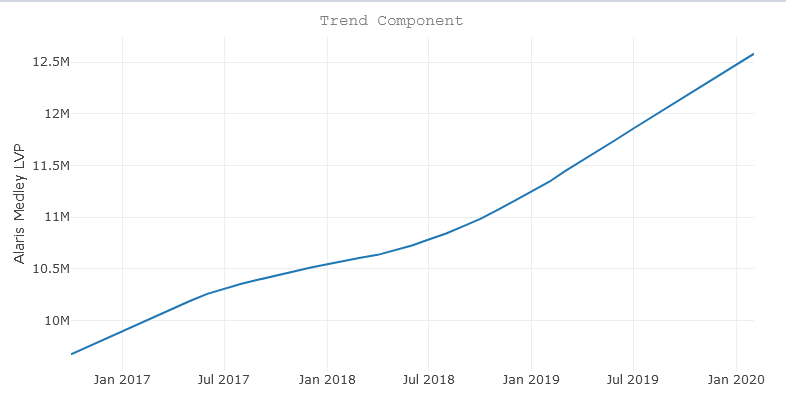
## How to Decompose Time Series Data into Trend and Seasonality

Time series decomposition involves thinking of a series as a combination of level, trend, seasonality, and noise components. There could also be a component of cycle city i.e. a cycle that might occur in a timeframe greater than a year.

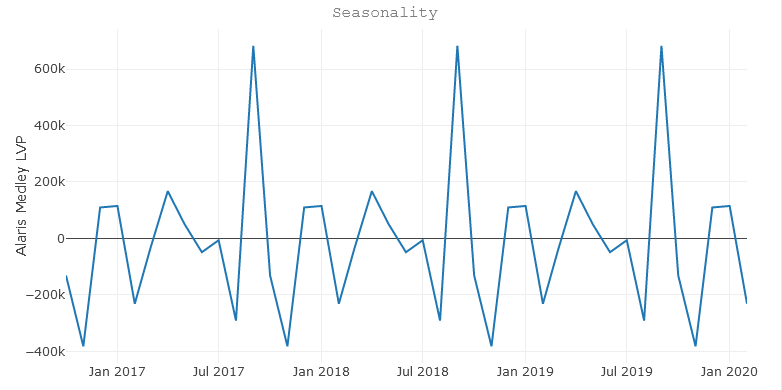
Decomposition provides a useful abstract model for thinking about time series generally and for better understanding problems during time series forecasting.

### Time Series Components

* Systematic: Components of the time series that have consistency or recurrence and can be described and modelled.
* Non-Systematic: Components of the time series that cannot be directly modelled
* Level: The average value in the series. (can be considered as intercept)
* Trend: The increasing or decreasing value in the series over time.

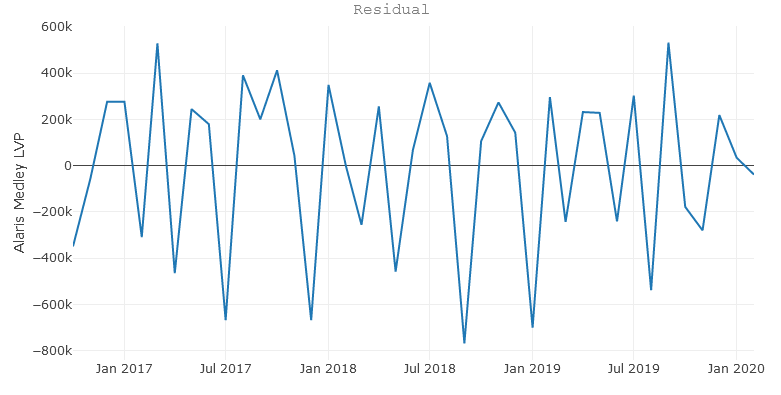


*🡨 Trend in the chart show the long-term direction of the metric, in this case sales of a product*

* Seasonality: The repeating short-term cycle in the series

*🡨 Seasonality shows patterns that occur on a cyclical basis within a year, in the example September observes higher sales, as this is the financial year end and sales team would be pushing for closures*

* Noise: The random variation in the series. Also referred to as residual (after removing trend, seasonal and level components from the time series.)



*🡨 Check if any patterns are seen or special scenarios that would need to be explored and could lead to understanding of exogenous variables that impact sales like a deep discount, Tax raise etc. Speculation can also be a exogenous variable linked to an event.*

### Combining Time Series Components

A series is thought to be an aggregate or combination of these four components. All series have a level and noise. The trend and seasonality components are optional. It is helpful to think of the components as combining either **additively** or **multiplicatively**.

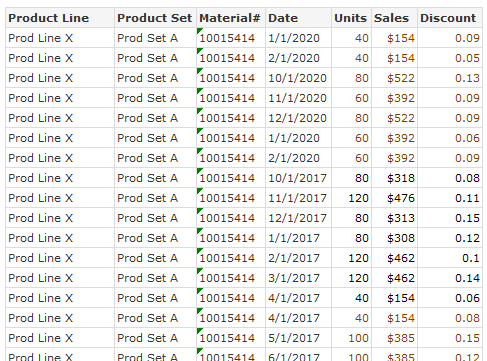
## Understanding Exogenous Variables

Plot the independent and exogenous variables to understand if there is an impact of the independent variables – these variables that show a strong relationship can be use in ARIMAX or Regression based forecasting models.

Figure 2: Chart shows a -ve relationship between the CPO Price and CPO Closing Stocks

# Data Prep

## Data Format

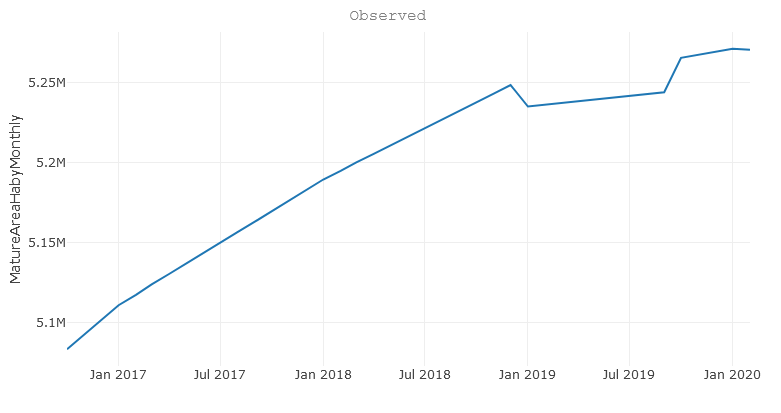
Time Series data can be handled best if data is stored in a flat format (example below) 

## Variables with different periodicity

It often happens that the data we are trying to integrate with i.e. the exogenous variables, or the dependent variables are not in the same periodicity as the dependent variable. E.g. to predict the price of Palm Oil, we use dependent variables among which few are monthly, and few are yearly. To be able to build the model, we need to bring the data into similar periodicity.

Depending of the data availability 2 approaches can be taken

* Summarize Low Granularity data to higher level periods:
  + Sum the monthly variables to be as same periodicity as the yearly variables. Then all variables are yearly and in same periodicity
  + However, the limitation with this model is that we lose granularity and variability in the data and might not help with the overall model accuracy or understanding seasonal components
  + Can be used when trend and level are mostly seen in data with very low monthly fluctuations.



* Linear Growth split the yearly data into months:
  + In this scenario, the yearly variables are split into monthly, using linear growth rate between the 2 years
  + In this case, the split-out data is going to be linear, and might not have a statically significant impact in a few cases

# Time Series Imputation

In timeseries data when there are missing values for a month or a few months, it would be challenging to run the forecasting models. If data is missing in a time series – there are multiple methods to impute the data

## Avg of previous & next period

In situations where a value is an outlier and is more likely a measurement system error or a data error, we use a replacement technique where the Outlier month value is replaced with the mean of previous period and next period.

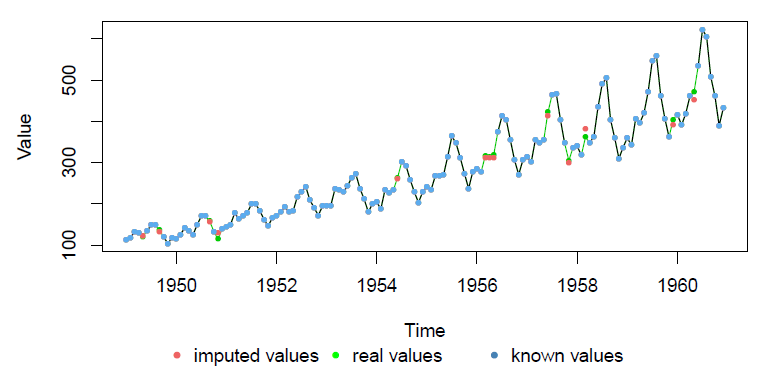
## Previous Value Imputation

If values in a time period is missing, take the past value to impute (e.g. last month value taken to this month). While the simplest of all methods, it is not accurate as it excludes seasonality and trend components

## Rolling Mean Imputation

Rolling Mean imputation is a method in which the missing value for a certain period is replaced by the rolling mean based on last X records (X could vary based on data e.g 6 month rolling or 4 weeks rolling etc.). Depending on the nature of the data and what fits best – there are 3 approaches that can be used here a simple moving average, a linear weighted moving average (gives higher weights to the most recent month) or an exponential weighted moving average.

## Kalman Smoothing

Kalman Smoothing on structural time series models (or on the state space representation of an ARIMA model) for imputation have a higher accuracy of prediction as compared to some of the previous models.

This approach was used in Beryl project – data has more than 50K hospitals with missing values for few hospitals in different years. The Kalman smoothing algo for TS imputation was implemented here.

## Auto ARIMA

Auto Arima models can be used for TS imputation however requires high computation time. ARIMA is a complex modeling technique which is explained in detail in section *6.5*

# Different Time Series Approaches

[Time series models](https://store.aptech.com/gauss-applications-category/time-series-mt.html) are used for a variety of reasons - predicting future outcomes, understanding past outcomes, making policy suggestions, and much more. These general goals of time series modelling don’t vary significantly from modelling cross-sectional or panel data. However, the techniques used in [time series models](https://store.aptech.com/gauss-applications-category/time-series-mt.html) must account for time correlation

* Basic Models
  + Moving Average
  + Simple Exponential Smoothing
  + Holt Method (Double Exponential Smoothing)
  + Holt Winters Method (Triple Exponential Smoothing)
* Advanced Models
  + Regression Models
  + SARIMAX
* AI Based Models
  + RNN & LSTM

## Moving Average

A simple moving average is the most basis time series approaches that could be used to understand demand on the next period. It takes a simple average of last N observations e.g. 3 month moving average, 13 week moving average etc.

Ft+1 – Forecast for time t+1  
N – No. of observations  
value of Y at time k

## Single Exponential Smoothing

One of the drawbacks of simple moving average technique is that it gives equal weight to all the previous observations used in forecasting the future value. This can be overcome by assigning differential weights to the past observations. Parameter α in below Eq. is called the *smoothing constant* and its value lies between 0 and 1. Here, the weights assigned to past data decline exponentially with the most recent observations assigned higher weights.

Ft+1 = αYt + α (1- α) Yt-1 + α (1- α)2 Yt-2 + α (1- α)3 Yt-3 +…

α – Smoothing Constant and its value lies between 0 and 1.  
Ft+1 – Forecast for time t+1  
Yt - value of Y at time t

Single Exponential Smoothing also assumes a steady time-series data with no significant trend, seasonal or cyclical component.

*Hint: Whenever the data is smooth (without much fluctuations), we may choose higher value of* α*. However, when the data is highly fluctuating, then it is better to choose lower value of* α*.*

## Holt Method

One of the drawbacks of single exponential smoothing is that the model does not do well in the presence of **trend**. This can be improved by introducing an additional equation for capturing the trend in the time-series data. Double exponential smoothing (Holt Winters) uses **two equations** to forecast the future values of the time series, **one for forecasting the level (short term average value) and another for capturing the trend**.

**Level (or intercept)**: Lt = α Yt + (1 − α) Ft

Lt is the level which represents the smoothed value  
Ft is the forecast for period t. Where Ft = Lt-1 + Tt-1Yt is value of Y at time t  
α is smoothing constant for level, 0 < α < 1

**Trend**: Tt = β × (Lt – Lt−1) + (1 − β) × Tt−1

Tt is the slope of the line or the rate of increase or decrease at period t  
β is the smoothing constant for trend, 0 < β < 1

**Forecast**: Ft+1 = Lt + Tt

## Holt Winters (Triple Exponential Smoothing)

Triple exponential smoothing is used when the data has trend as well as seasonality.

Holt ([1957](https://otexts.com/fpp2/holt-winters.html#ref-Holt57)) and Winters ([1960](https://otexts.com/fpp2/holt-winters.html#ref-Winters60)) extended Holt’s method to capture seasonality. The Holt-Winters seasonal method comprises the forecast equation and three smoothing equations — one for the **level Lt, one for the trend Tt, and one for the seasonal component St**, with corresponding smoothing parameters α, β and γ. We use c to denote the frequency of the seasonality, i.e., the number of seasons in a year. For example, for quarterly data c=4, and for monthly data c=12.

There are **two** variations to this method that differ in the nature of the seasonal component namely **additive & multiplicative** method

The additive method is preferred when the seasonal variations are roughly constant through the series, while the multiplicative method is preferred when the seasonal variations are changing proportional to the level of the series.

With the additive method, the seasonal component is expressed in absolute terms in the scale of the observed series, and in the level equation the series is seasonally adjusted by subtracting the seasonal component. Within each year, the seasonal component will add up to approximately zero.

With the multiplicative method, the seasonal component is expressed in relative terms (percentages), and the series is seasonally adjusted by dividing through by the seasonal component. Within each year, the seasonal component will sum up to approximately c.

### Holt-Winters’ Additive method

The component form for the multiplicative method is

**Forecast**: Ft+1 = Lt + Tt + St+1-c

Level or Intercept equation:

Trend Equation:

Seasonal Equation:

Lt is the level which represents the smoothed value  
Yt is value of Y at time t  
Tt is the slope of the line or the rate of increase or decrease at period t  
St is the seasonal index   
c is the number of seasons (c=12 for monthly data, 4 for quarterly, 7 for daily etc.)  
α is smoothing constant for level, 0 < α < 1  
β is the smoothing constant for trend, 0 < β < 1  
γ is the smoothing constant for seasonality, 0 < γ < 1

### Holt-Winters’ Multiplicative method

The component form for the multiplicative method is

**Forecast**: Ft+1 = [Lt + Tt] \* St+1-c

Level or Intercept equation:

Trend Equation:

Seasonal Equation:

Lt is the level which represents the smoothed value  
Yt is value of Y at time t  
Tt is the slope of the line or the rate of increase or decrease at period t  
St is the seasonal index   
c is the number of seasons (c=12 for monthly data, 4 for quarterly, 7 for daily etc.)  
α is smoothing constant for level, 0 < α < 1  
β is the smoothing constant for trend, 0 < β < 1  
γ is the smoothing constant for seasonality, 0 < γ < 1

## Regression Model for Forecasting

Regression is probably more appropriate method for forecasting when the data has values of predictor (explanatory) variables in addition to the dependent variable Yt.

The forecasted value at time t, Ft, can be written as a regression equation as follows:

Ft is the forecasted value of Yt  
X1t, X2t, etc. are the predictor variables measured at time t

## ARIMA

ARIMA models are an expansion of ARMA models. Auto-regressive (AR) and moving average (MA) or ARMA models are basically regression models.

Before getting into ARIMA models, we need to understand 2 concepts

**Stationary Process:**

If a time-series data, Yt, is stationary, then it satisfies the following conditions:

1. The mean values of Yt at different values of t are constant.
2. The variances of Yt at different time periods are constant (Homoscedasticity)
3. The covariances of Yt and Yt-k for different lags depend only on k and not on time t.

**White noise:**

White noise is a process of residuals εt that are uncorrelated and follow normal distribution with mean 0 and constant standard deviation.

### Integrated (I)

Represents the differencing of raw observations to allow for the time series to become stationary, i.e., data values are replaced by the difference between the data values and the previous values.

The lag operator shifts the data and the errors either forward (when k < 0) or backward lags

### Autoregressive (AR)

A type of a stochastic model which depends on its time-lagged regressions of the series is named auto-regression (AR). Essentially, Auto-regression means **regression of a variable on itself** measured at different time periods.

Fundamental Assumptions of AR Models

* Time series is assumed to be a **stationary process**
* The errors εt follow a **white noise** process

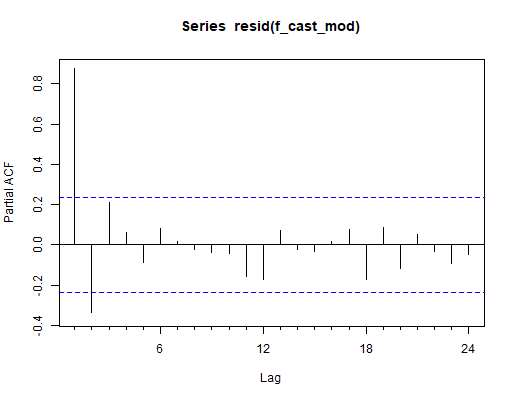
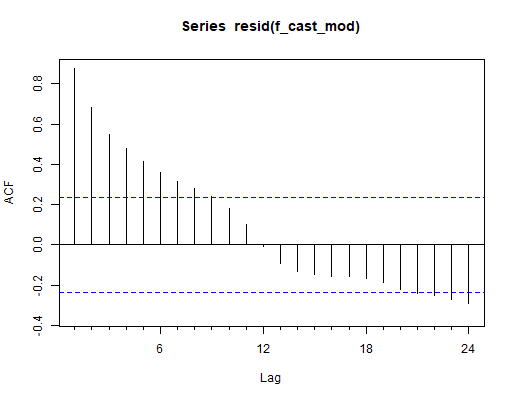
When the time series data is not stationary, then we must **convert the non-stationary times-series data to stationary data before applying AR models**.

Auto-regression model which is regressed with Lag 1 is generalizes as AR(1) or AR(p), where p represents the no. of lags.

One of the important tasks in using auto-regressive model in forecasting is the model identification, which is, identifying the value of p (the number of lags). One of the standard approaches used for model identification is auto-correlation function (ACF) and partial auto-correlation function (PACF).

**ACF** is an (complete) auto-correlation function which gives us values of correlation of any series with its lagged values. We plot these values along with the confidence for ACF plot. In simple terms, it describes how well the present value of the series is related with its past values. A time series can have components like trend, seasonality, cyclic and residual. ACF considers all these components while finding correlations hence it’s a ‘complete auto-correlation plot’  
**PACF** is a partial auto-correlation function. Basically instead of finding correlations of present with lags like ACF, it finds correlation of the residuals (which remains after removing the effects which are already explained by the earlier lag(s)) with the next lag value hence ‘partial’ and not ‘complete’ as we remove already found variations before we find the next correlation. So, if there is any hidden information in the residual which can be modeled by the next lag, we might get a good correlation and we will keep that next lag as a feature while modeling. Remember while modeling we don’t want to keep too many features which are correlated as that can create multicollinearity issues. Hence, we need to retain only the relevant features

Figure 3 ACF & PACF plots  
Understanding ACF & PACF plots play a crucial role in identifying the apt model parameters. Looking at the ACF that is gradually declining in this case, and PACF that is +ve at Lag(1) and PACF that cuts off sharply – this is a AR model



Note that the model identification is an iterative process and may require additional inputs. The model identification using ACF and PACF cannot be taken as conclusive evidence for the number of lags in AR process.

### Moving average (MA)

Incorporates the dependency between an observation and a residual error from a moving average model applied to lagged observations

### Construction of an ARIMA model

1. Stationarize the series, if necessary, by differencing (& perhaps also logging, deflating, etc.)
2. Study the pattern of autocorrelations and partial autocorrelations to determine if lags of the stationarized series and/or lags of the forecast errors should be included in the forecasting equation
3. Fit the model that is suggested and check its residual diagnostics, particularly the residual ACF and PACF plots, to see if all coefficients are significant and all of the pattern has been explained.
4. Patterns that remain in the ACF and PACF may suggest the need for additional AR or MA terms

### AR & MA Signatures

**AR signature**: ACF that dies out gradually and PACF that cuts off sharply after a few lags

* An AR series is usually positively autocorrelated at lag 1(or even borderline nonstationary)

**MA signature**: ACF that cuts off sharply after a few lags and PACF that dies out more gradually

* An MA series is usually negatively autocorrelated at lag 1(or even mildly overdifferenced)

### Which model should we choose?

That depends on the assumptions we are comfortable making with respect to the constancy of the trend in the data. The model with only one order of differencing assumes a constant average trend--it is essentially a fine-tuned random walk model with growth--and it therefore makes relatively conservative trend projections. It is also fairly optimistic about the accuracy with which it can forecast more than one period ahead. The model with two orders of differencing assumes a time-varying local trend--it is essentially a linear exponential smoothing model--and its trend projections are somewhat fickle. As a general rule in this kind of situation, I would recommend choosing the model with the lower order of differencing, other things being roughly equal. In practice, random-walk or simple-exponential-smoothing models often seem to work better than linear exponential smoothing models.

### Can Mixed models be used i.e. have AR and MA components:

In most cases, the best model turns out a model that uses either only AR terms or only MA terms, although in some cases a "mixed" model with both AR and MA terms may provide the best fit to the data. However, care must be exercised when fitting mixed models. It is possible for an AR term and an MA term to cancel each other's effects, even though both may appear significant in the model (as judged by the t-statistics of their coefficients). Thus, for example, suppose that the "correct" model for a time series is an ARIMA(0,1,1) model, but instead you fit an ARIMA(1,1,2) model--i.e., you include one additional AR term and one additional MA term. Then the additional terms may end up appearing significant in the model, but internally they may be merely working against each other. The resulting parameter estimates may be ambiguous, and the parameter estimation process may take very many (e.g., more than 10) iterations to converge.

To get detailed understanding and learn about the rules to apply and choosing the best models and considerations, visit the site <http://people.duke.edu/~rnau/411arim.htm>

## SARIMA

The addition of S to the ARIMA model is used for factoring Seasonality (S). Seasonality makes it so that the mean of the observations is not constant, but instead evolves according to a cyclical pattern we say that the series has seasonality of periods

## ARIMAX

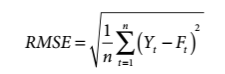
The addition of X to the ARIMA model is used for factoring is Exogenous input (X). Integrates an ordinary regression model that uses external variables into the ARIMA model. Important note is that to be able to get the forecast, the independent variables must have predictions for the equal number of periods for which the independent variables needs to be predicted.

# Accuracy Measures

RMSE & MAPE are most used accuracy measures for forecasting.

### RMSE – Root Mean Squared Error

**Root Mean Square Error**(RMSE) is the [standard deviation](https://www.statisticshowto.com/probability-and-statistics/standard-deviation/) of the [residuals](https://www.statisticshowto.com/residual/) ([prediction errors](https://www.statisticshowto.com/prediction-error-definition/)). Residuals are a measure of how far from the regression line data points are; RMSE is a measure of how spread out these residuals are. In other words, it tells you how concentrated the data is around the [line of best fit](https://www.statisticshowto.com/line-of-best-fit/). Root mean square error is commonly used in climatology, forecasting, and [regression analysis](https://www.statisticshowto.com/probability-and-statistics/regression-analysis/) to verify experimental results.



* Yt – Actual Sales at time t
* Ft – Forecasted Sales for time t
* n – [sample size](https://www.statisticshowto.com/probability-and-statistics/find-sample-size/) or no. of periods

### MAPE – Mean Absolute Percentage Error



* Yt – Actual Sales at time t
* Ft – Forecasted Sales for time t
* n – [sample size](https://www.statisticshowto.com/probability-and-statistics/find-sample-size/) or no. of periods

MAPE benchmarks: MAPE **below** **10%** is an acceptable model

### Power of Forecasting Model:

Theil’s coefficient is the ratio of the mean squared error of the forecasting model to the MSE of the Naïve model. The value of U < 1 indicates that forecasting model is better

Yt is observed value of Y at time t  
Ft is forecasted value of Y at time t  
Fnt is Naïve forecast value of Y at time t

# Prediction Intervals

A prediction interval gives an interval within which we expect Yt to lie with a specified probability. For example, assuming that the forecast errors are normally distributed, a 95% prediction interval for the forecast is

Yt ±1.96 \* σh

σh - standard deviation of residuals at h-step. Higher the h higher the standard deviation and hence an increase in prediction interval as we continue to increase the forecast periods.

For Practical purposes below is the approach used:

* Lower bound: Predicted – 1.5 \* MAPE
* Upper bound: Predicted + 1.5 \* MAPE

# OSG Time Series Dashboard

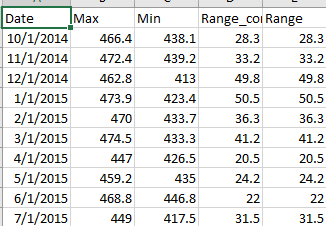
To help in running Time Series Models with ease, a R Shiny Dashboard has been developed which can be run locally and all explained models in this paper can be run including parameter tuning

Pre-requisites

* Install R & R Studio
* Install packages – shiny, shinydashboard, scales, tseries, tidyverse, plotly, lubridate, forecast, zoo, DT, feasts, fpp3, naniar, imputeTS

Data Format

* Save file as ‘TS\_RawData.csv’
* First column should be Date, and names as ‘Date’
  + Date format should be mm/dd/yyyy i.e. 10/31/2020 for Oct 31, 2020
* All other columns should be numeric columns
* Each column represents a variable like Sales, Price etc.

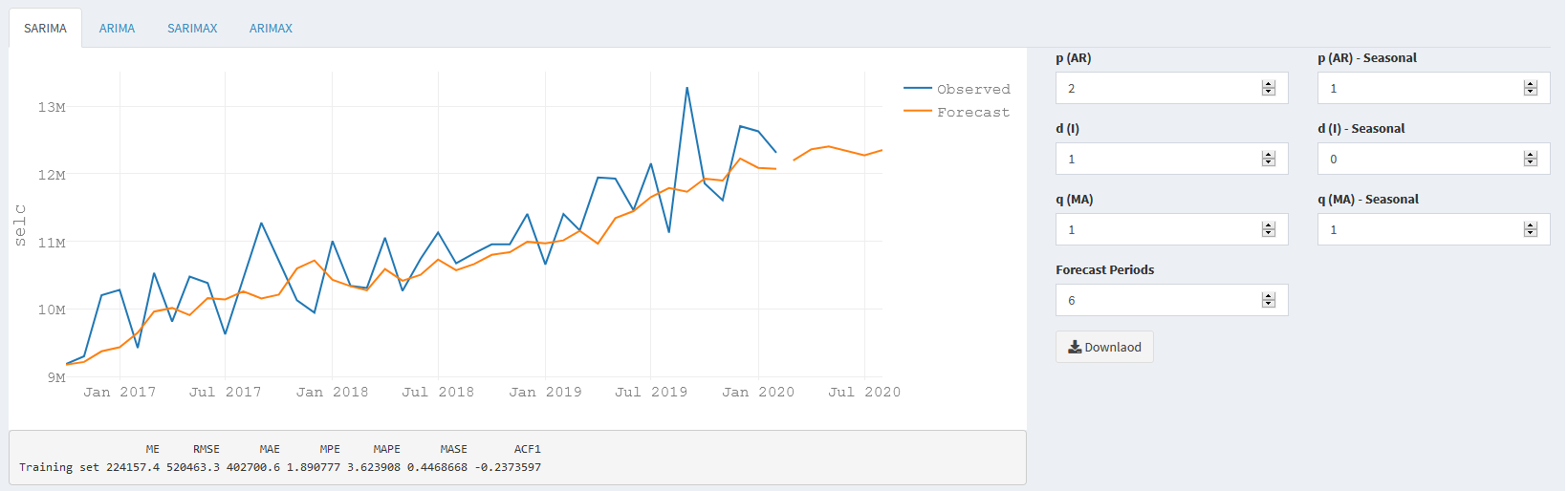


Running the code

* Once all packages are installed
* Ensure the code file ‘TS\_app\_V2.R’ & the raw data file ‘TS\_RawData.csv’ are saved in same folder
* Open the code in R Studio and click on Run App – the dashboard will open in Chrome

Important Note: Dashboard has been designed for Monthly data.

e.g. from TS dashboard on a fitter SARIMA model



# Pros & Cons of Forecast

**Pros**

1. You’ll gain valuable insight

[Forecasting](https://johngalt.com/resources/results/?topic=Forecasting&sortby=popularity)gets you into the habit of looking at past and real-time data to predict future demand. And in doing so, you’ll be able to anticipate demand fluctuations more effectively. But more than that, it’ll give you insight into your company’s health and provide you with an opportunity to course-correct or adjust.

 2. You’ll learn from past mistakes

 You don’t start from scratch after each forecast. Even if your prediction was nowhere close to what ended up coming to pass, it gives you a starting point. It’s common to review where and why things didn’t happen the way you predicted. Your forecasts should eventually improve. But more than that, you’ll get into the habit of reflecting upon past performance. And self-reflection can be a powerful driver of company growth.

 3. It can decrease costs

 When done right, anticipating demand will help you tweak your processes to increase efficiency all along the supply chain. Because you’re better able to predict what customers will want and when they’ll want it, you may also be able to decrease excess inventory levels, thus increasing overall profitability.

**Cons**

1. Forecasts are never 100% accurate

 Let’s face it: it’s hard to predict the future. Even if you have a great process in place and forecasting experts on your payroll, your forecasts will never be spot on. Some products and markets simply have a high level of volatility. And in general, there is just an endless number of factors that influence demand.

 2. It can be time-consuming and resource-intensive

 Forecasting involves a lot of data gathering, data organizing, and coordination. Companies typically employ a team of demand planners who are responsible for coming up with the forecast. But in order to do this well, demand planners need substantial input from the sales and marketing teams. In addition, it’s not uncommon for processes to be manual and labor-intensive, thus taking up a lot of time. Fortunately, if you have the right technology in place, this is much less of an issue.

 3. It can also be costly

 On a related note, hiring a team of demand planners is a significant investment. When you add to that the cost of using good quality tools, upfront costs can add up. But investing in advanced software, high-quality talent and solid forecasting processes is just that: an investment. We’re confident you’ll see a return when all of that is done right.

Forecasting is a business practice that every company engages in to one extent or another. And it can be hugely valuable, providing those companies who have implemented a solid forecasting process with a leg up on their competition. What’s more, even the disadvantages can be overcome with the right people, technology and processes. So, learn how partnering with our Forecast Experts and implementing our Atlas Suite can make a difference.

# References

Kumar, U. Dinesh. *Business Analytics: The Science of Data-Driven Decision Making*. Wiley India, 2017.

Moritz, Steffen. “ImputeTS v3.0.” *ImputeTS Package | R Documentation*, www.rdocumentation.org/packages/imputeTS/versions/3.0.

Hyndman, Rob J, and George Athanasopoulos. “Forecasting: Principles and Practice.” *OTexts*, otexts.com/fpp2/.

Nau, Robert. ‘*Statistical Forecasting: Notes on Regression and Time Series Analysis’*, Duke University,   
people.duke.edu/~rnau/411home.htm.

Makridakis, Spyros, et al. *Forecasting: Methods and Applications*. Wiley, 2005.