Time Series Analysis

Arima and Holt-Winters models for demand at a bike-sharing service

Presented by Yesika Contreras

MS Business Analytics

Some Terminology & Background

What are Time Series?

Definition

An ordered sequence of values of a variable at equally spaced time intervals.

Time series analysis accounts for the fact that data points taken over time may have an internal structure (such as autocorrelation, trend or seasonal variation) that should be accounted for.

Techniques

There are many methods of model fitting to forecasting, monitoring or even feedback and feedforward control, including the following:

- Box-Jenkins ARIMA models
- Box-Jenkins Multivariate Models
- Averaging Methods.
- Exponential Smoothing Techniques.

Cases

- Explaining seasonal patterns in sales
- Predicting the expected number of incoming or churning customers
- Estimating the effect of a newly launched product on number of sold units
- Detecting unusual events and estimating the magnitude of their effect

Univariate Time Series

Consists of single (scalar) observations recorded sequentially over equal time increments.

Ex: Monthly CO2 concentrations.
Southern oscillations to predict el Niño effects.

Approaches

Triple exponential smoothing: Holt-Winters

Decompose the time series into:

- trend,
- seasonal,
- residual component

Autoregressive (AR) Models:

Linear regression of the current value of the series against one or more prior values of the series.

Moving Average (MA) Models:

Linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series.

Fitting the MA estimates is more complicated than with AR models because the error terms are not observable. This means that iterative non-linear fitting procedures need to be used in place of linear least squares.

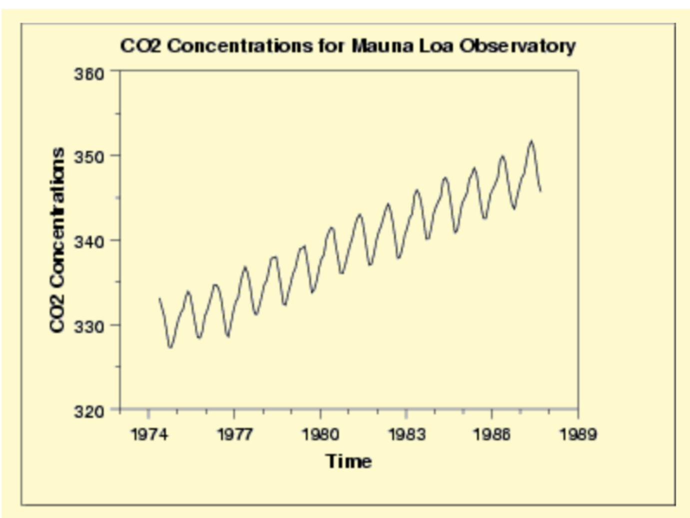
Stationarity

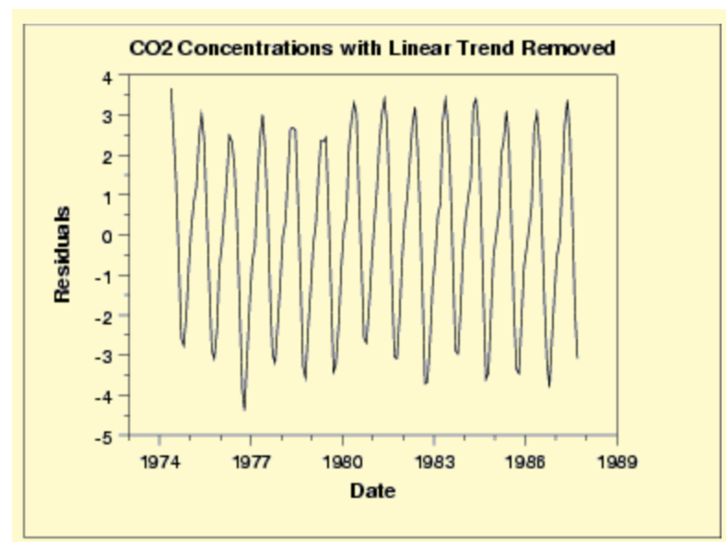
Stationarity

but for our purpose we mean a flat looking series, without trend, constant variance over time, a constant autocorrelation structure over time and no periodic fluctuations (seasonality).

Transformations to Achieve Stationarity

We can difference the data.
If the data contain a trend, we can fit some type of curve to the data and then model the residuals from that fit.
For non-constant variance, taking the logarithm or square root of the series may stabilize the variance.





Linear Trend Removed

https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc422.htm

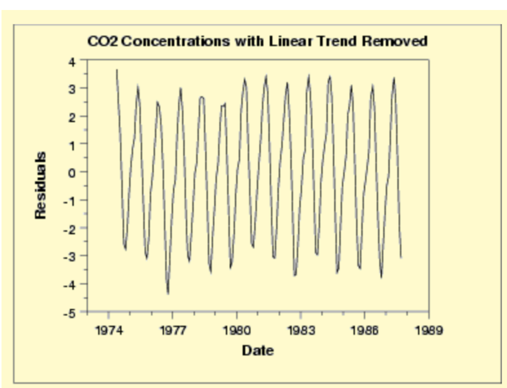
Seasonality

Stationarity

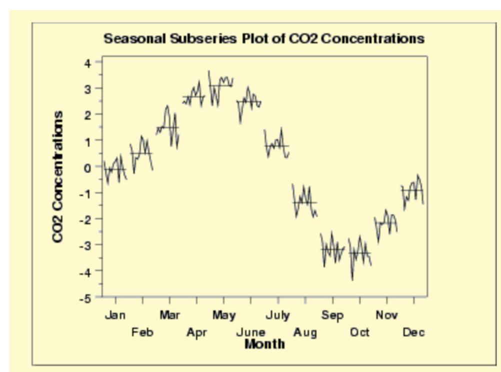
By seasonality, we mean periodic fluctuations. Seasonality is quite common in economic time series.
If seasonality is present, it must be incorporated into the time series model.

Detecting Seasonality

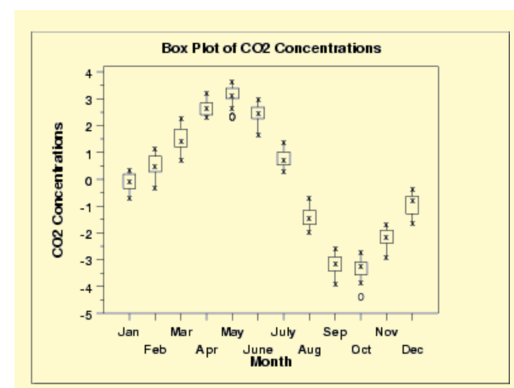
A run sequence plot will often show seasonality.
A seasonal subseries plot is a specialized technique for showing seasonality.
Multiple box plots can be used as an alternative to the seasonal subseries plot to detect seasonality.
The autocorrelation plot can help identify seasonality.



Run Sequence Plot

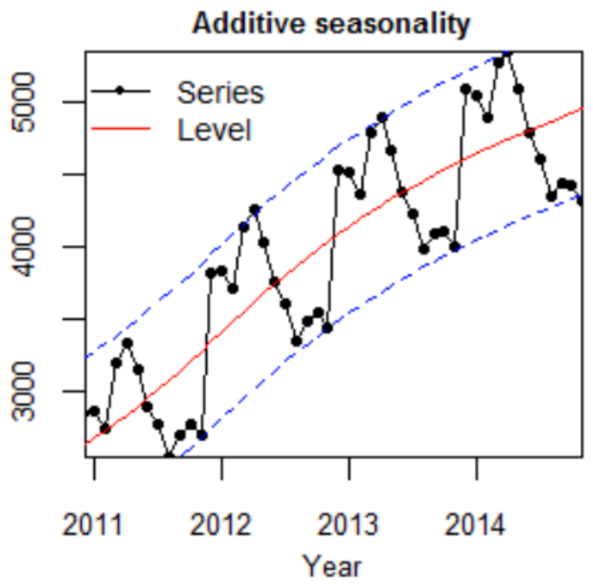


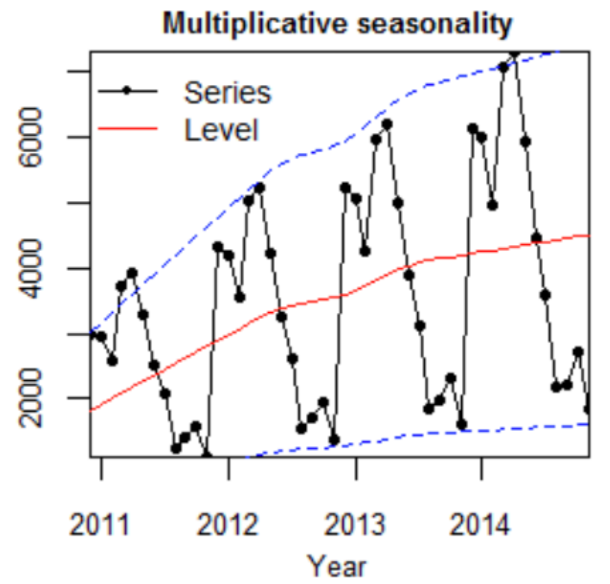
Seasonal Subseries Plot



Box Plot

Additive and multiplicative seasonality





In additive seasonality the amplitude of the seasonality remains constant, while in multiplicative changes as the level does, where the season becomes wider or smaller.

Level: fit a centered moving average (red line)
series: Add or multiply a factor to it to get the dotted blue lines for the two types of seasonality.

http://kourentzes.com/forecasting/2014/11/09/additive-and-multiplicative-seasonality/

Model Evaluation

Error

"mean squared error": The "error" = true amount spent minus the estimated amount.
The "error squared" is the error above, squared.
The "SSE" is the sum of the squared errors.
The "MSE" is the mean of the squared errors.
FPE (Final Prediction Error)

Models Comparison

Criteria for comparing quality of fit across multiple models:

- Akaike information criteria (AIC)
- Baysian information criteria (BIC)

These criteria are closely related and can be interpreted as an estimate of how much information would be lost if a given model is chosen. When comparing models, one wants to minimize AIC and BIC.

Model Evaluation

For model selected:
Examining ACF and PACF plots for model residuals. If model order parameters and structure are correctly specified, we would expect no significant autocorrelations present.
If the model is not correctly specified, that will usually be reflected in residuals in the form of trends, skeweness, or any other patterns not captured by the model.
Ideally, residuals should look like white noise, meaning they are normally distributed.
IF there are no significant autocorrelations present in the residuals. Then, the model was correctly specified.

Problem

Problem: Forecasting for a Bike rental demand.

Problem Description

This analysis attempts to generate a time series model based to predict daily rental demand of bicycles. The training dataset on the daily number of bicycles checkouts from a bike sharing service, is part of the UCI Machine Learning Repository. The goal is to predict the next 30 days of demand.

Tasks & approach

Identify possible trend or seasonality and predict future demand at a bike-sharing service using Arima and Holt-Winters models and generating a basic forecast. Although there are other machine learning models that could offer better predictions, Times series will be analyzed for forecasting.
Steps will be explained in the following slides.

Data fields

- Total of 731 observations of 19 features.

The core data set is related to the two-year historical log corresponding to years 2011 and 2012 from Capital Bikeshare system, Washington D.C.,

- R lbraries: ggplot2, forecast, tseries, TStools

Results

The data was smoothed weekly and the count of rentals with the date was extracted for time series analysis. The data presented trend and seasonality. The time series correspond to an additive seasonality. From the time series forecasting we can infer that the demand is also correlated to another additional features, because the model couldn't explained the rental behavior perfectly.

Features

dteday	date
season	1:springer, 2:summer, 3:fall, 4:winter
yr	year (0: 2011, 1:2012)
mnth	month (1 to 12)
holiday	weather day is holiday or not (extracted from http
weekday	day of the week
workingday	if day is neither weekend nor holiday is 1, otherwise is 0.
temp	Normalized temperature in Celsius. The values are divided to 41 (max)
atemp	Normalized feeling temperature in Celsius. The values are divided to 50 (max)
hum	Normalized humidity. The values are divided to 100 (max)
windspeed	Normalized wind speed. The values are divided to 67 (max)
casual	count of casual users
registered	count of registered users
cnt	count of total rental bikes including both casual and registered
weathersit	- 1: Clear, Few clouds, Partly cloudy, Partly cloudy
	- 2: Mist + Cloudy, Mist + Broken clouds, Mist + Few clouds, Mist
	- 3: Light Snow, Light Rain + Thunderstorm + Scattered clouds, Light Rain + Scattered clouds
	- 4: Heavy Rain + Ice Pallets + Thunderstorm + Mist, Snow + Fog

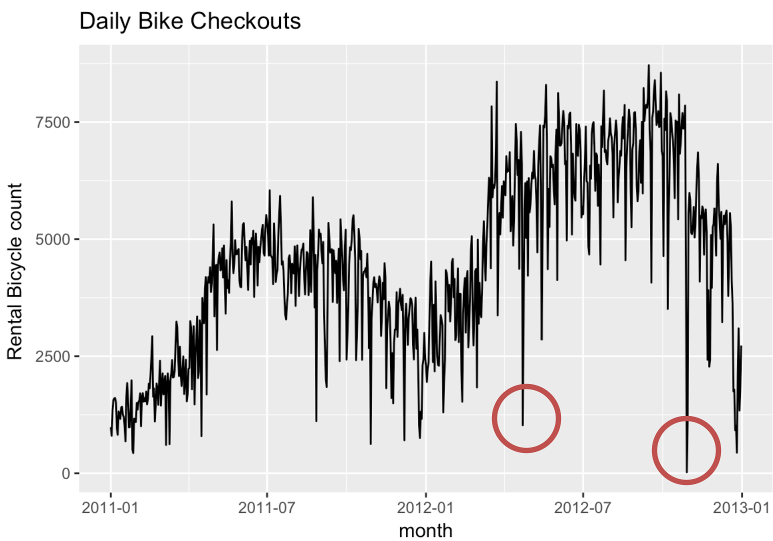
Forecasting Process Approach

Examine data	Plot the data and examine its patterns and irregularities Clean up any outliers or missing values if needed tsclean() is a convenient method for outlier removal and inputing missing values Take a logarithm of a series to help stabilize a strong growth trend
Decompose data	Does the series appear to have trends or seasonality? Use decompose() or stl() to examine and possibly remove components of the series
Stationarity Analysis	Is the series stationary? Use adf.test(), ACF, PACF plots to determine order of differencing needed
Choose model order	Choose order of the ARIMA by examining ACF and PACF plots Autocorrelations
Evaluate and Fit a model	Check residuals, which should haven no patterns and be normally distributed. If there are visible patterns or bias, plot ACF/PACF. Are any additional order parameters needed? Evaluate and iterate. Refit model if needed. Compare model errors and fit criteria such as AIC or BIC. Calculate forecast using the chosen model

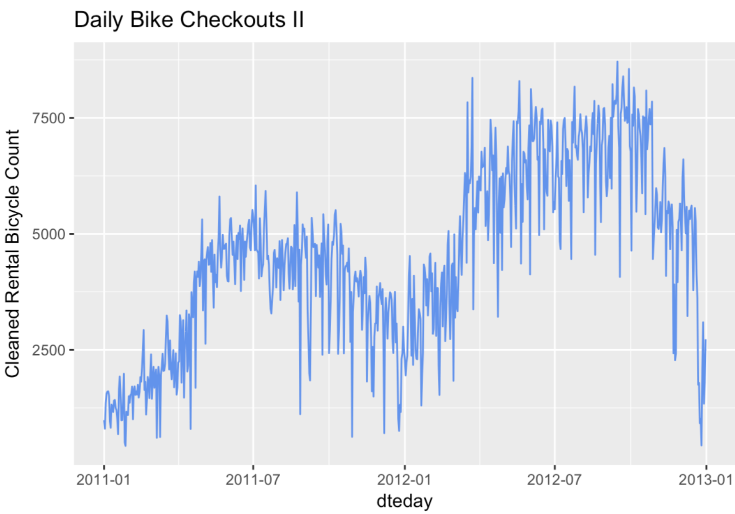
https://www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-data-science-tutorials

Daily Demand: Creating a time series object from with ts() and passing tsclean()

Plotting original data:



Plotting clean Time Series

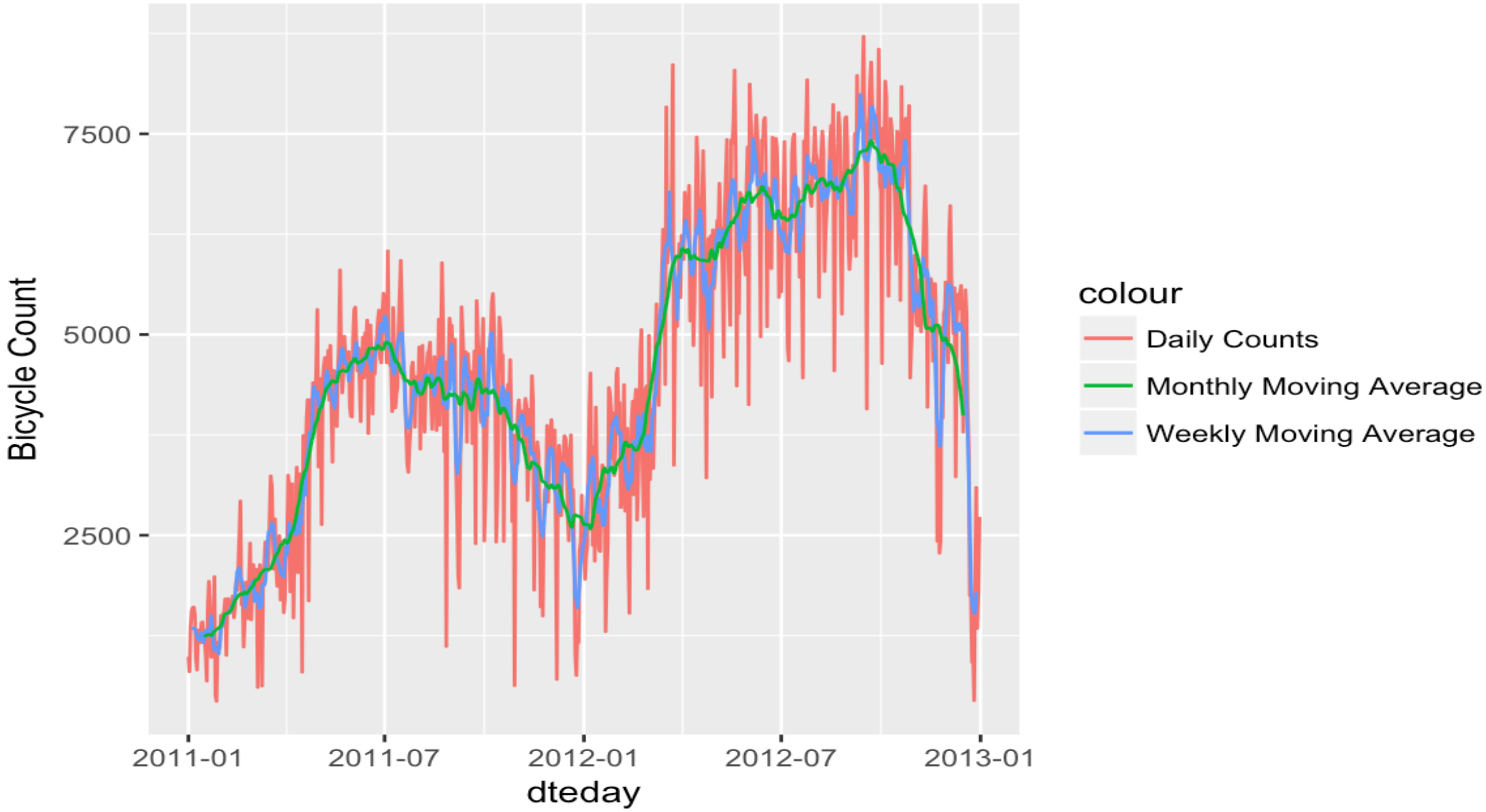


Possible Outliers:
In some cases, the number of bicycles checked out dropped below 100 on day and rose to over 4,000 the next day.

After cleaning:
Even after removing outliers, the daily data is still pretty volatile. Visually, we could a draw a line through the series tracing its bigger troughs and peaks while smoothing out noisy fluctuations (moving average).
The wider the window of the moving average, the smoother original series becomes.

Smoothing the time series

with weekly or monthly moving average (ma)

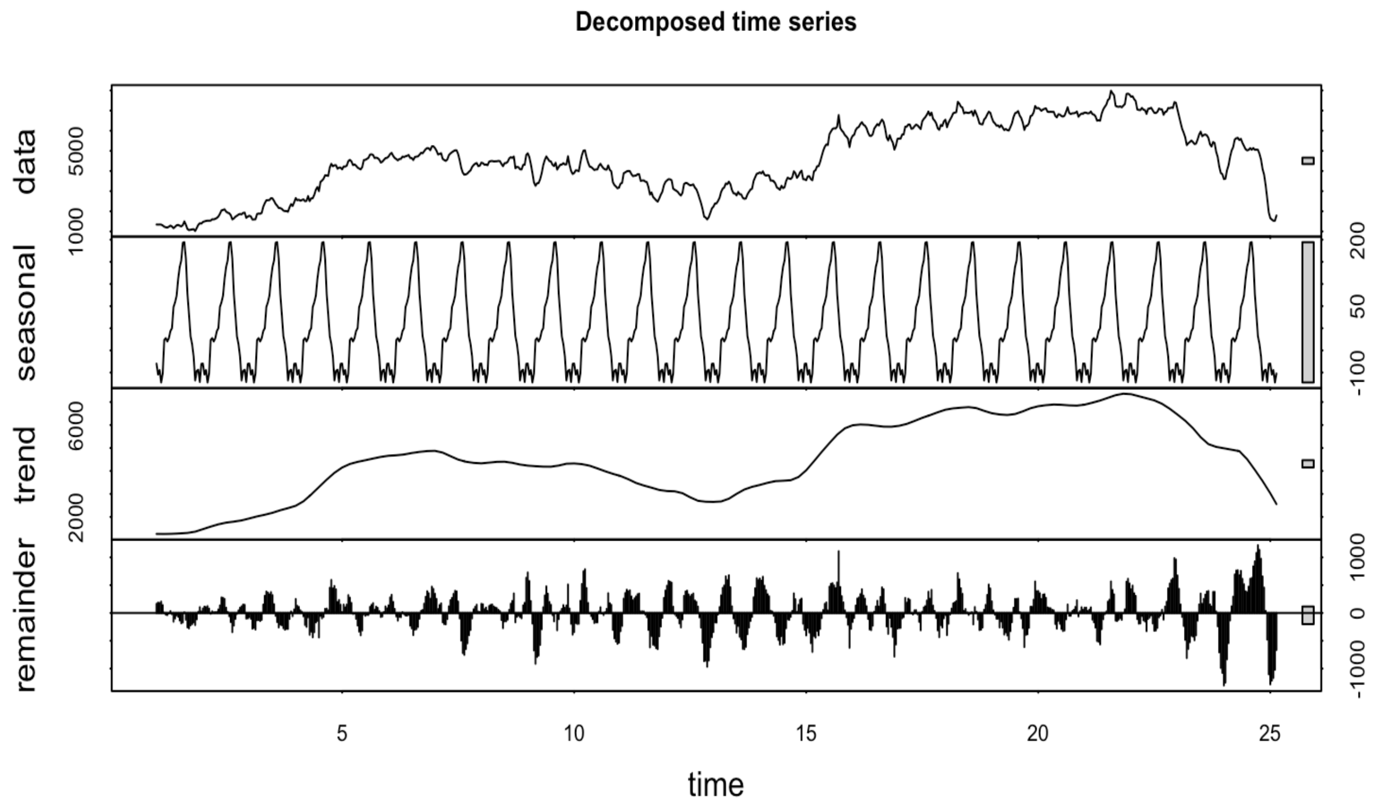


- Smoothing the series into something more stable and therefore predictable.
- The higher the window parameter, the more we capture the global trend, but miss on local trends

Conclusion: We will model the smoothed series of weekly moving average (blue line).

Decomposing Data

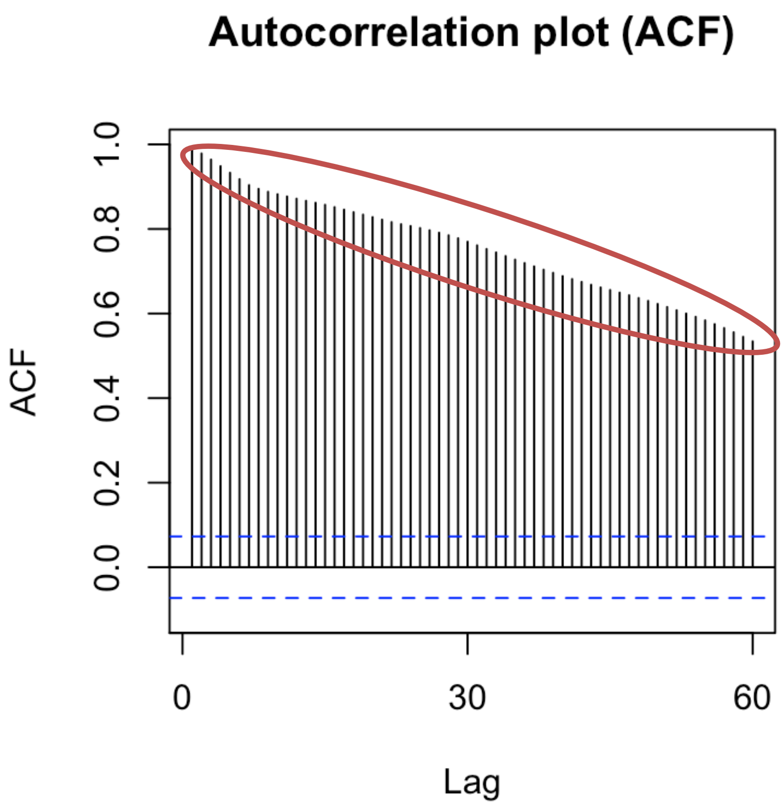
Components of seasonality, trend, and cycle.

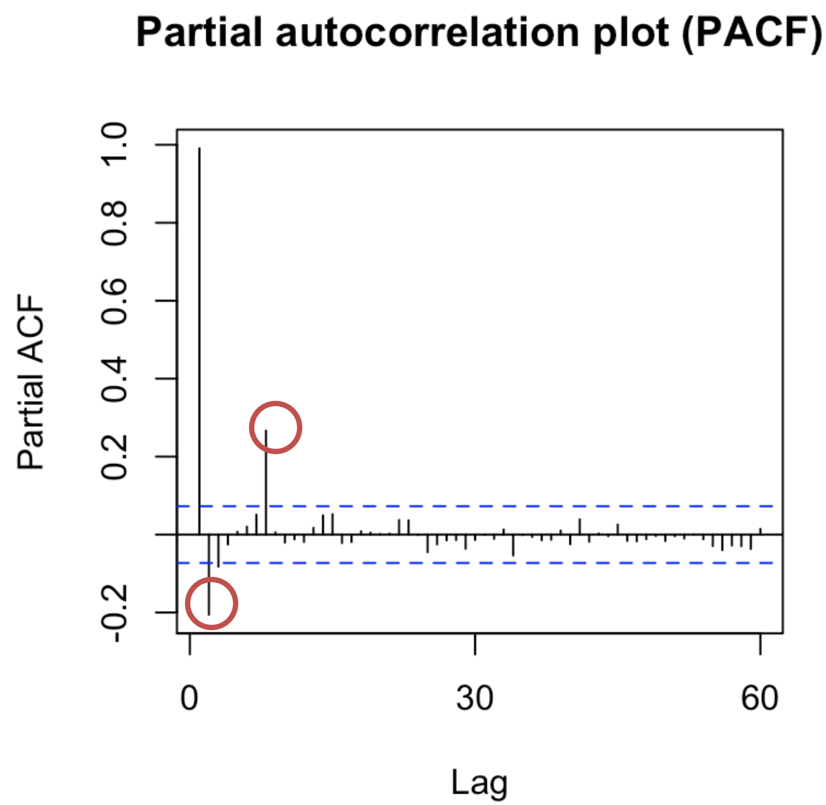


Calculating the seasonal component of the series using smoothing, and adjusting the original series by subtracting seasonality

Stationarity

A series is said to be stationary when its mean, variance, and autocovariance are time invariant.

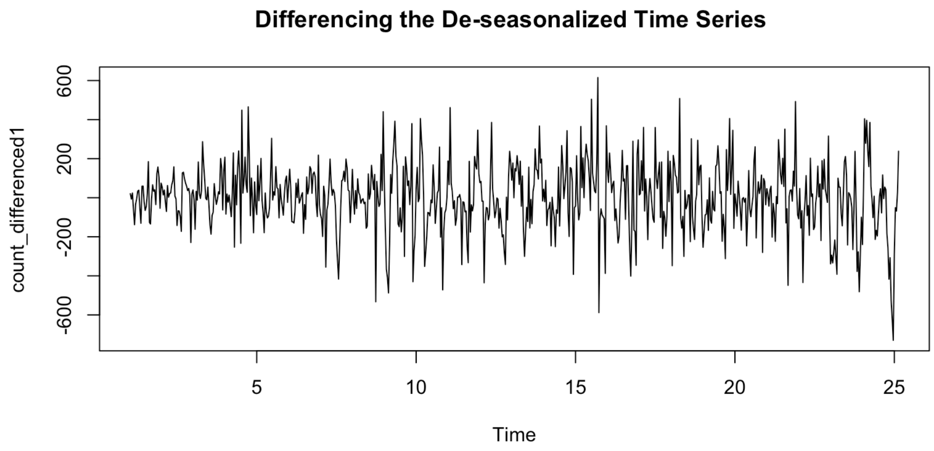




- ACF plot shows that the ts has significant autocorrelations with many lags.
- PACF plot only shows a spike at lags 1 and 7.
- Then the ACF autocorrelations could be due to carry-over correlation from the first or early lags

Differencing:

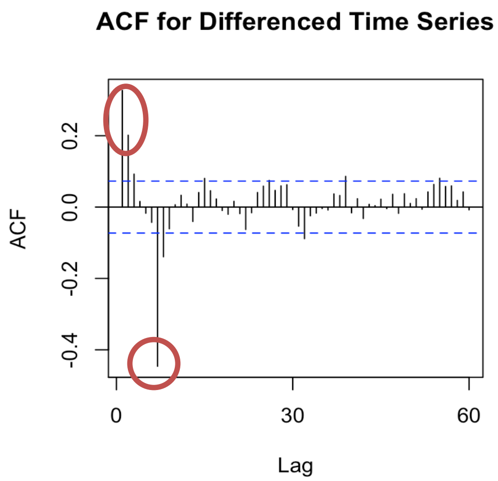
Starting with the order of *d = 1* and re-evaluate if further differencing is needed.

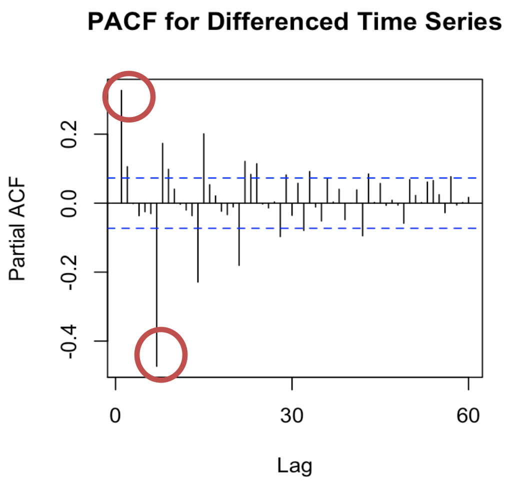


Now the data transformed looks stationary, there is not visible strong trend.

Differencing of order 1 terms is sufficient and should be included in the model.

Choosing p & q for the model:



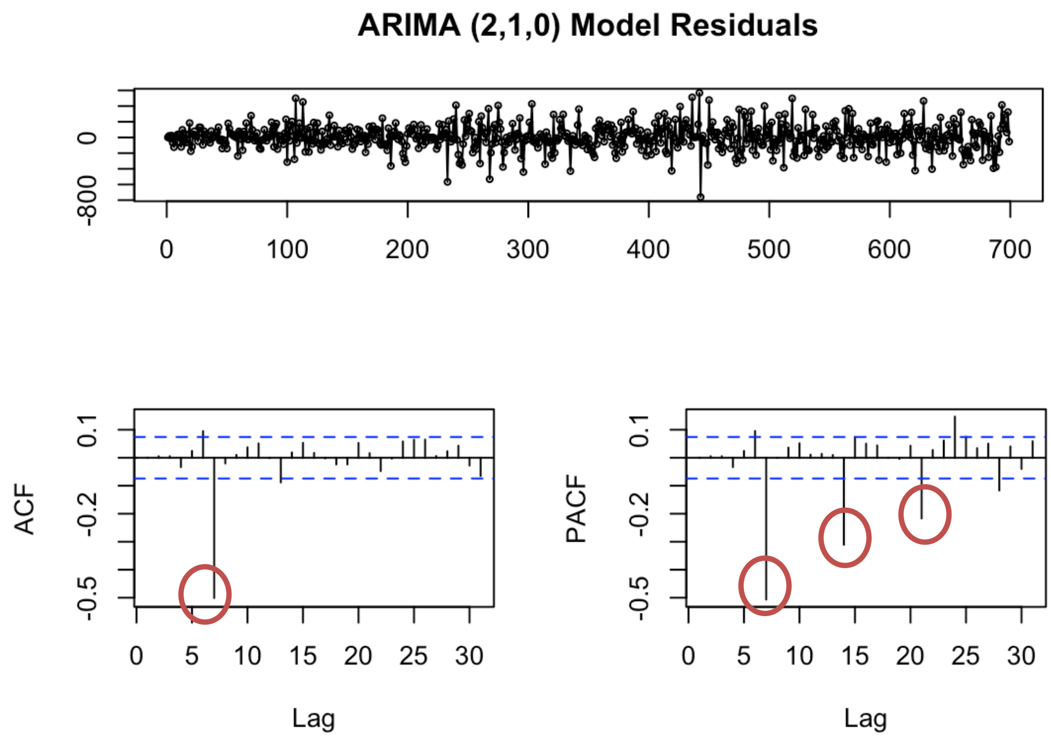


From ACF: There are significant auto correlations at lag 1, 2 and 7,8.
From PACF: Partial correlation plots show a significant spike at lag 1 and 7.
Conclusion: test models with AR or MA components of order 1, 2, or 7.

Fitting an ARIMA model:

Arima (p: AR order**, d:**degree of differencing**, q:**MA order**)**

AIC=9070.72

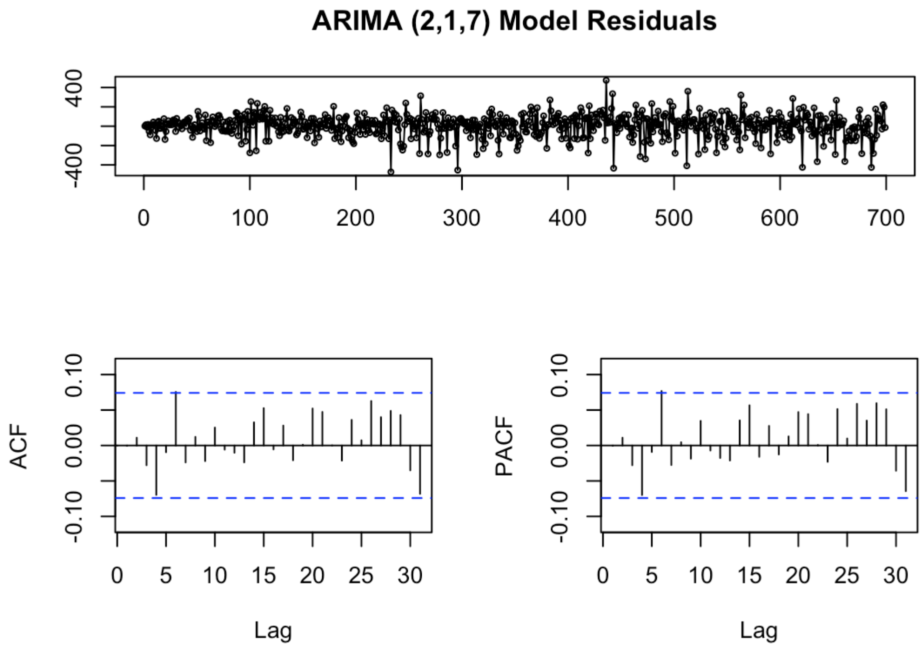


Result: There is a clear pattern present in ACF/PACF and model residuals plots repeating at lag 7.
This suggests that our model may be better off with a different specification, such as p = 7 or q = 7.

Fitting an ARIMA model:

Arima (p: AR order, d:degree of differencing, q:MA order)

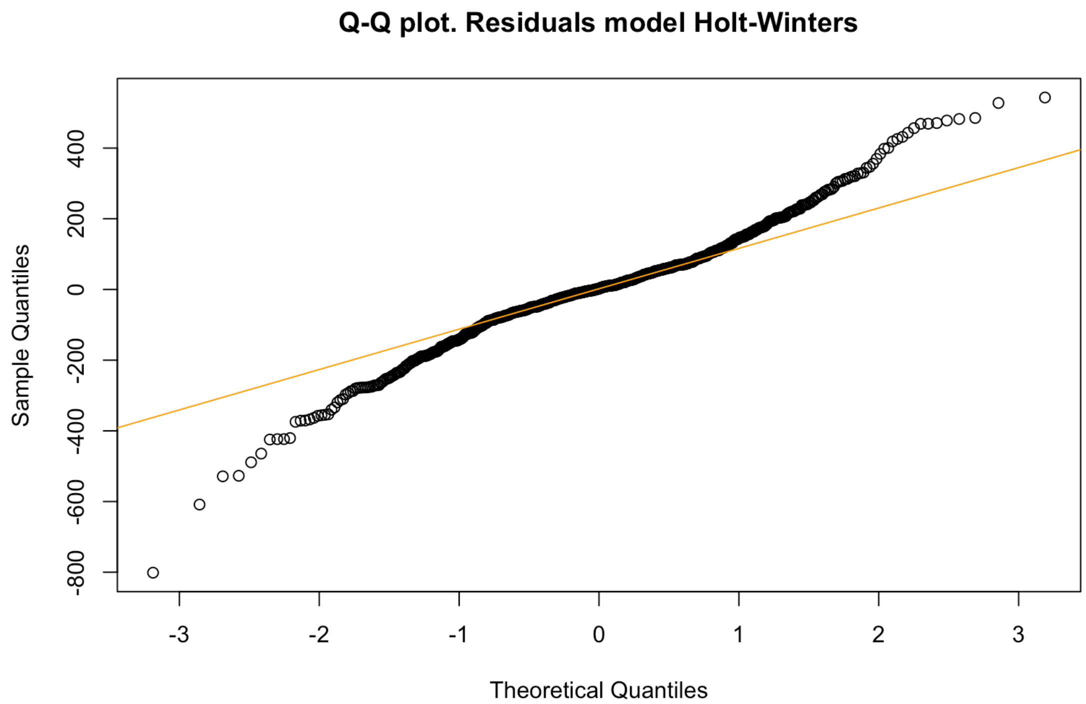
AIC= 8681.93 (Smaller)



The model uses an autoregressive term of second lag, incorporates differencing of degree 1, and a moving average model of order 7.

Result: There are no significant autocorrelations present in the residuals. The model was correctly specified.

Fitting a Triple exponential smoothing: Holt-Winters model:

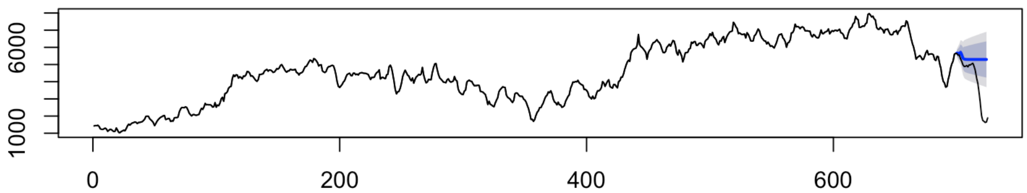


The model uses an autoregressive term of second lag, incorporates differencing of degree 1, and a moving average model of order 7.

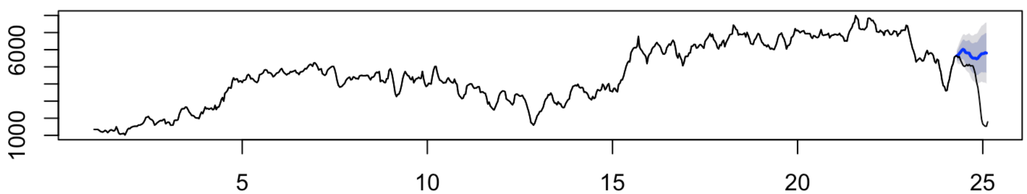
Result: There are no significant autocorrelations present in the residuals. The model was correctly specified.

Forecasting:

Forecasting using Arima Model



Forecasting using HoltWinters Model



Result: As we found the Forecasting plots and checking the residual plot and QQ plot, we can see that the residuals of the forecasting have a pattern, and are not normally distributed, which means the the models don’t fit the data so well.

In conclusion, daily bicycle demand is probably highly dependent on other factors, such weather, holidays, time of the day, etc. Therefore, other forecasting techniques, such as time series models that allow for inclusion of other predictors using methods such ARMAX or dynamic regression may help to explain the demand.

References

Lichman, M. (2013). UCI Machine Learning Repository [http://archive.ics.uci.edu/ml]. Irvine, CA: University of California, School of Information and Computer Science.

Dalinina, Ruslana (2017), Introduction to Forecasting with ARIMA in R. https://www.datascience.com/blog/introduction-to-forecasting-with-arima-in-r-learn-data-science-tutorials, march,2018.

NIST/SEMATECH e-Handbook of Statistical Methods, http://www.itl.nist.gov/div898/handbook/, march,2018.

Practical Time Series Forecasting with R, 2nd Edition, p. 29

Coghlan Avril. A Little Book of R For Time Series.2017. (https://media.readthedocs.org/pdf/a-little-book-of-r-for-time-series/latest/a-little-book-of-r-for-time-series.pdf)

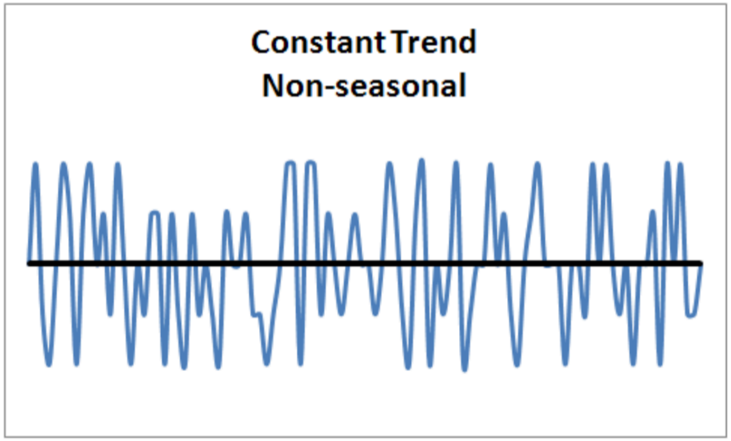
Hyndman Rob, George Athanasopoulos. 2013. Forecasting: Principles and Practice. https://www.otexts.org/book/fpp

Any questions?

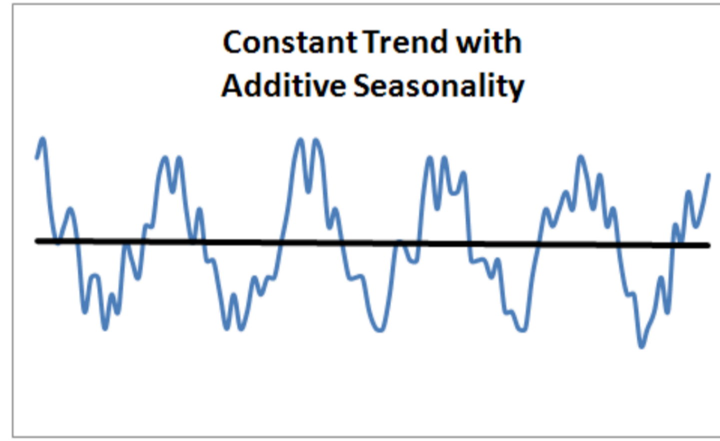
THANK YOU

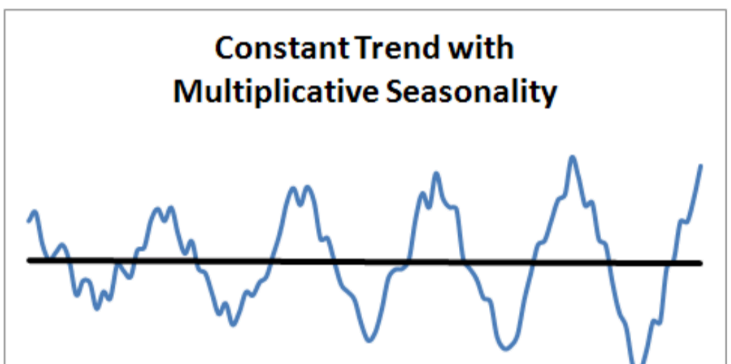
Appendix. Techniques Background

Common time series trends

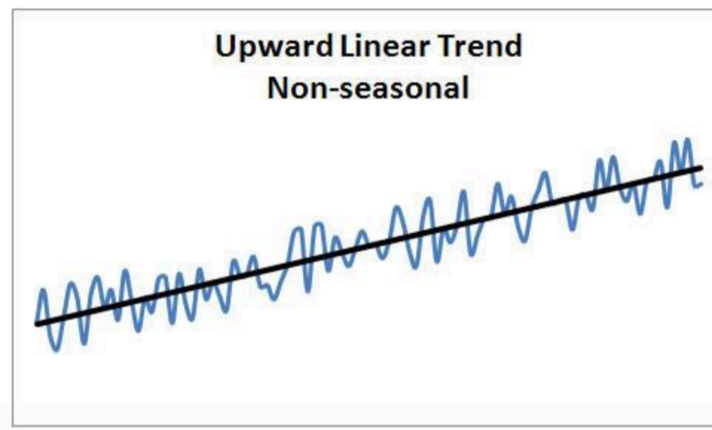


just junk

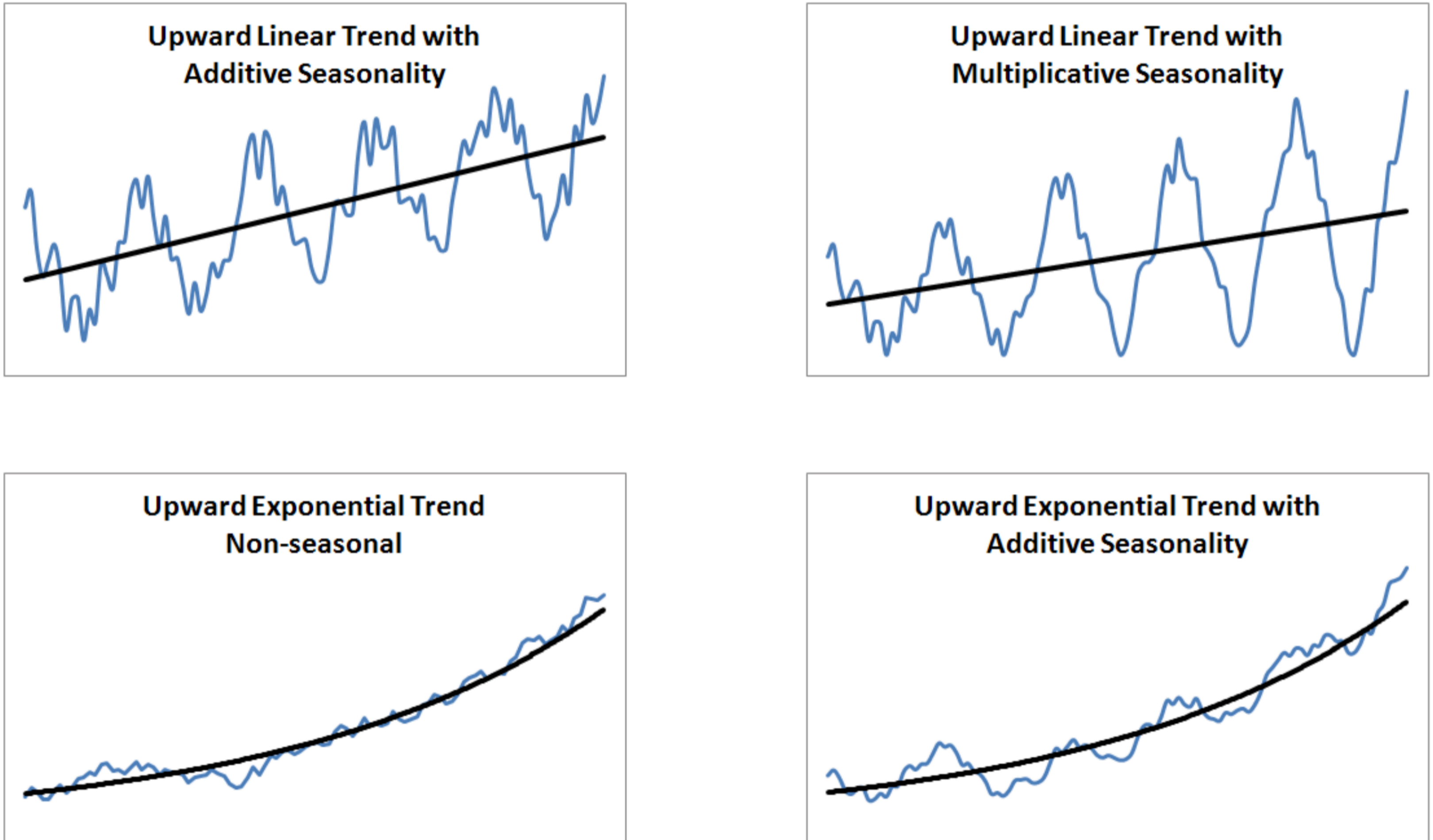


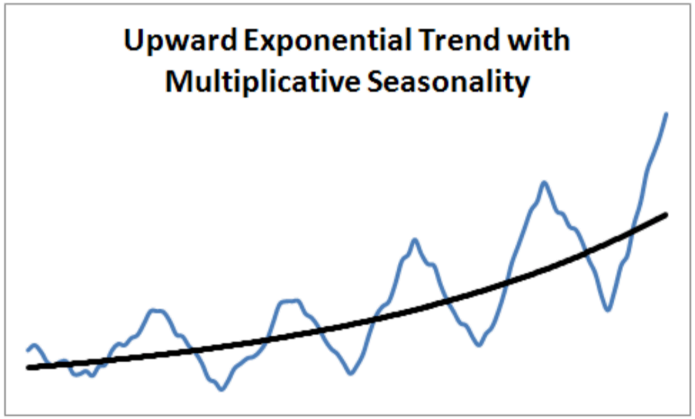


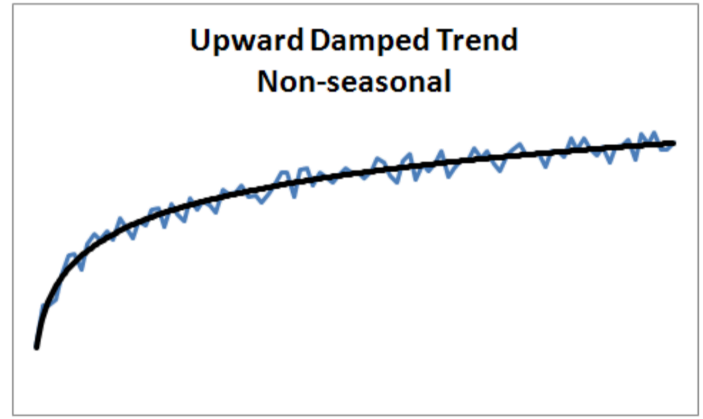
its growing

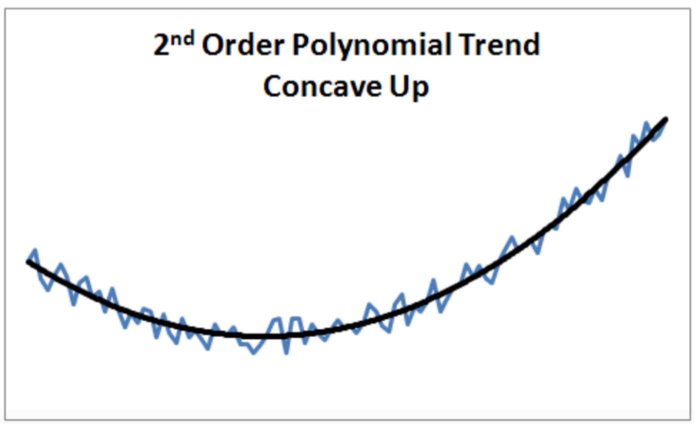


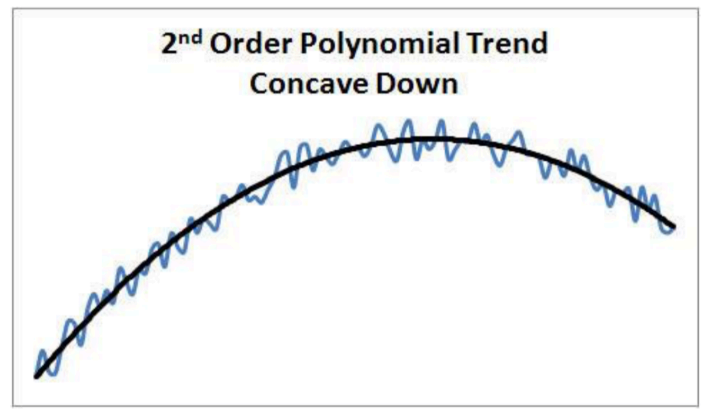
Images from: Practical Time Series Forecasting with R, 2nd Edition, p. 29


Upward Linear Trend with Additive Seasonality
Upward Linear Trend with Multiplicative Seasonality
Upward Exponential Trend Non-seasonal
Upward Exponential Trend with Additive Seasonality


Upward Exponential Trend with
Multiplicative Seasonality


Upward Damped Trend
Non-seasonal


2nd Order Polynomial Trend
Concave Up

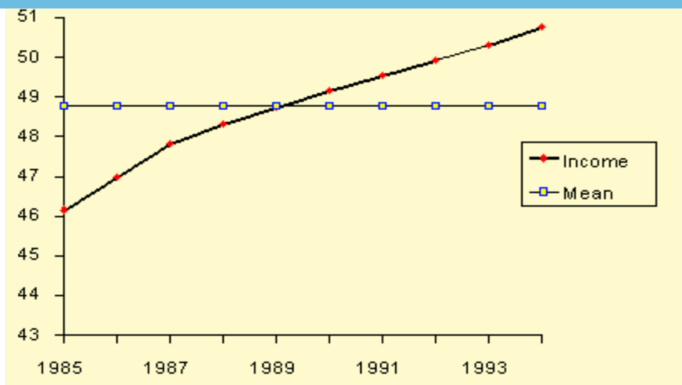

2nd Order Polynomial Trend
Concave Down

Smoothing Techniques

(Reduce the effect due to random variation)

Simple Averaging Method

The "simple" average or mean of all past observations is only a useful estimate for forecasting when there are no trends.



Weight Averaging Method

Another way of computing the average is by adding each value divided by the number of values by not weighting all past observations equally.

$$\bar{x} = \frac{1}{n}\sum_{i=1}^{n} x_i = \left(\frac{1}{n}\right)x_1 + \left(\frac{1}{n}\right)x_2 + \ldots + \left(\frac{1}{n}\right)x_n .$$

The $\left(\frac{1}{n}\right)$ are the weights and, of course, they sum to 1.

Single Moving Average (MA)

Compute the mean of successive smaller sets of numbers of past data. Let us set M, the size of the "smaller set"

$$M_t = \frac{X_t + X_{t-1} + \cdots + X_{t-N+1}}{N} .$$

M= 3

Supplier	$	MA
1	9	
2	8	
3	9	8.667
4	12	9.667
5	9	10.000
6	12	11.000

Centered Moving Average

Smooth the smoothed values

Interim Steps

Period	Value	MA	Centered
1	9		
1.5			
2	8		
2.5		9.5	
3	9		9.5
3.5		9.5	
4	12		10.0
4.5		10.5	
5	9		10.750
5.5		11.0	
6	12		

M= 4

Period	Value	Centered MA
1	9	
2	8	
3	9	9.5
4	12	10.0
5	9	10.75
6	12	
7	11	

https://www.itl.nist.gov/div898/handbook/pmc/section4/pmc422.htm

Exponential Smoothing (ES)

ES assigns exponentially decreasing weights as the observation get older.
In other words, recent observations are given relatively more weight in forecasting than the older observations.

In ES, however, there are one or more smoothing parameters to be determined (or estimated) and these choices determine the weights assigned to the observations.

For any time period t, the smoothed value S_t is found by computing

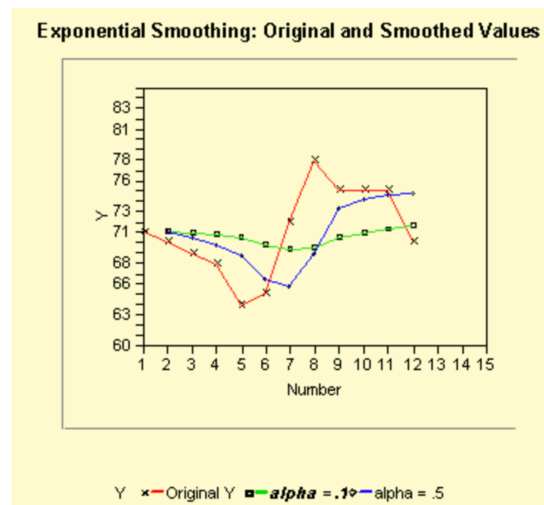
$$S_t = \alpha y_{t-1} + (1-\alpha)S_{t-1} \qquad 0 < \alpha \le 1 \qquad t \ge 3\,.$$

This is the *basic equation of exponential smoothing* and the constant or parameter α is called the *smoothing constant*.

$$\begin{aligned} S_t &= \alpha y_{t-1} + (1-\alpha)\left[\alpha y_{t-2} + (1-\alpha)S_{t-2}\right] \\ &= \alpha y_{t-1} + \alpha(1-\alpha)y_{t-2} + (1-\alpha)^2 S_{t-2}\,. \end{aligned}$$

$$S_t = \alpha \sum_{i=1}^{t-2} (1-\alpha)^{i-1} y_{t-i} + (1-\alpha)^{t-2} S_2\,, \quad t \ge 2\,.$$

This smoothing scheme begins by setting S2 to y1, where Si stands for smoothed observation or EWMA, and y stands for the original observation.



The speed at which the older responses are dampened (smoothed) is a function of the value of α. When α is close to 1, dampening is quick and when α is close to 0, dampening is slow.
Best value for α so the value which results in the smallest MSE

Forecasting with Exponential Smoothing

Single Exponential Smoothing

The new forecast is the old one plus an adjustment for the error that occurred in the last forecast.

The forecasting formula is the basic equation

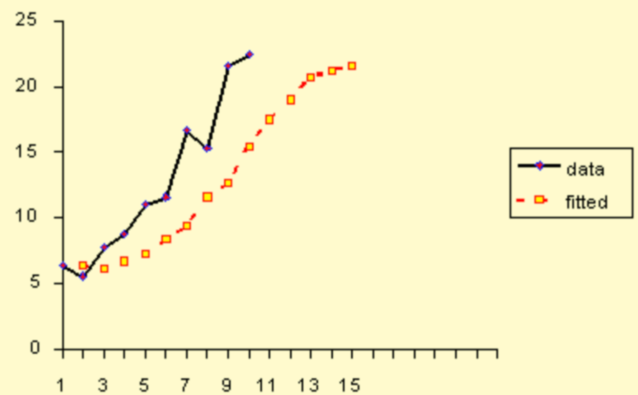
$$S_{t+1} = \alpha y_t + (1-\alpha)S_t, \quad 0 < \alpha \leq 1, \quad t > 0.$$

This can be written as:

$$S_{t+1} = S_t + \alpha \epsilon_t ,$$

where ϵ_t is the forecast error (actual - forecast) for period t.

The resulting graph looks like:



Double Exponential Smoothing

Introduction of a second equation with a second constant, γ, which must be chosen in conjunction with α. The second smoothing equation then updates the trend, which is expressed as the difference between the last two values.

$$S_t = \alpha y_t + (1-\alpha)(S_{t-1} + b_{t-1}) \quad 0 \leq \alpha \leq 1$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \quad 0 \leq \gamma \leq 1$$

S_1 is in general set to y_1. Here are three suggestions for b_1.

$$b_1 = y_2 - y_1$$

$$b_1 = \frac{1}{3}[(y_2 - y_1) + (y_3 - y_2) + (y_4 - y_3)]$$

$$b_1 = \frac{y_n - y_1}{n-1}$$

Forecasting with Double Exponential Smoothing(LASP)

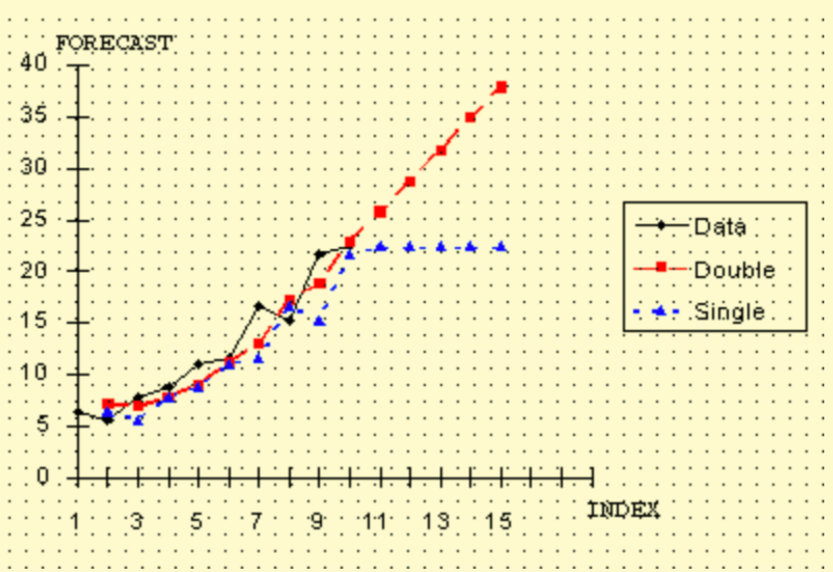
Forecasting formula

The one-period-ahead forecast is given by:

$$F_{t+1} = S_t + b_t \, .$$

The m-periods-ahead forecast is given by:

$$F_{t+m} = S_t + mb_t \, .$$



Triple Exponential Smoothing

Holt-Winters" (HW) method.

To handle seasonality, we have to add a third parameter.

The basic equations for their method are given by:

$$S_t = \alpha \frac{y_t}{I_{t-L}} + (1-\alpha)(S_{t-1} + b_{t-1}) \qquad \text{OVERALL SMOOTHING}$$

$$b_t = \gamma(S_t - S_{t-1}) + (1-\gamma)b_{t-1} \qquad \text{TREND SMOOTHING}$$

$$I_t = \beta \frac{y_t}{S_t} + (1-\beta)I_{t-L} \qquad \text{SEASONAL SMOOTHING}$$

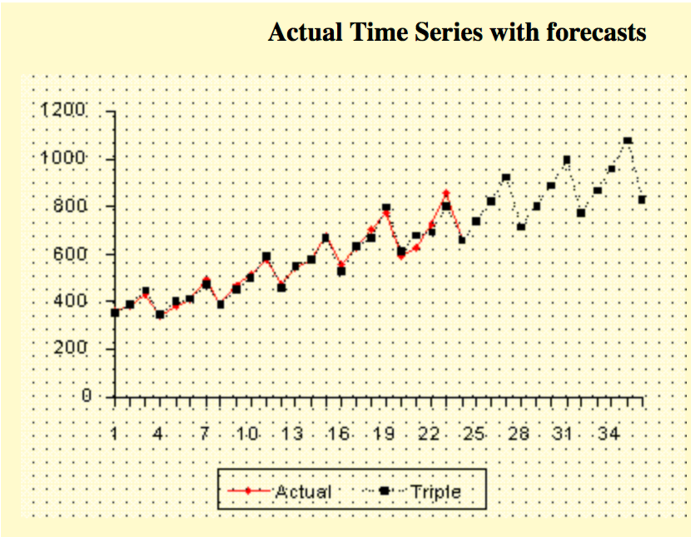
$$F_{t+m} = (S_t + mb_t)I_{t-L+m} \qquad \text{FORECAST ,}$$

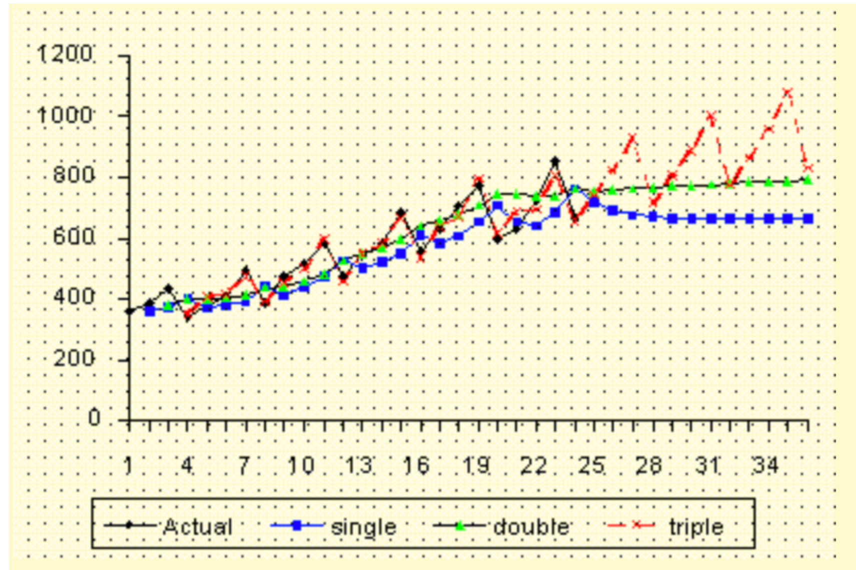
where

- y is the observation
- S is the smoothed observation
- b is the trend factor
- I is the seasonal index
- F is the forecast at m periods ahead
- t is an index denoting a time period

Comparison of Exponential Smooting methods

Data shows trend and seasonality





Comparison of MSEs

MSE	α demand	γ trend	β seasonality
6906	0.4694		
5054	0.1086	1.0000	
936	1.0000		1.0000
520	0.7556	0.0000	0.9837

Holt-Winters Method

HW has 3 updating equations, each with a constant that ranges from 0 to 1. The equations are intended to give more weight to recent observations and less weights to observations further in the past. These weights are geometrically decreasing by a constant ratio.

Arima

ARIMA

ARIMA stands for auto-regressive integrated moving average and is specified by these three order parameters: *(p, d, q)*. The process of fitting an ARIMA model is sometimes referred to as the Box-Jenkins method.

The Box-Jenkins ARIMA model is a combination of the AR and MA models differencing non-stationary series one or more times to achieve stationarity.

The auto-regressive parameter p specifies the number of lags used in the model. For example, AR(2) or, equivalently, ARIMA(2,0, 0).

The d represents the degree of differencing in the integrated (I(d)) component. Differencing a series involves simply subtracting its current and previous values d times.

An auto regressive (AR(p)) component is referring to the use of past values in the regression equation for the series Y.

$$Y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + e_t$$

A moving average (MA(q)) component represents the error of the model as a combination of previous error terms e_t.

The order q determines the number of terms to include in the model:

$$Y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q} + e_t$$

Differencing, autoregressive, and moving average components make up a non-seasonal ARIMA model, where y_d is Y differenced d times and c is a constant:

$$Y_t = c + \phi_1 y_{d\ t-1} + \phi_p y_{d\ t-p} + \dots + \theta_1 e_{t-1} + \theta_q e_{t-q} + e_t$$