Time Series

Intilaq Data Science Academy

Introduction to Time Series

Characteristics of Time Series

- **Order matters:** There is a dependency on time and changing the order could change the meaning of the data.
- **Sequential and equally spaced data points**
- **Each data point has at most one value**

→ Once the data is collected we will:

1. Identify the patterns represented by the sequence observations
2. Forecasting/ Predicting for future values

	A	B
1	Month	#Passengers
2	1949-01	112
3	1949-02	118
4	1949-03	132
5	1949-04	129
6	1949-05	121
7	1949-06	135
8	1949-07	148
9	1949-08	148

Characteristics of Time Series

- A stochastic process $\{y_t\}_{t=-\infty}^{\infty}$ is a collection of random variables or a process that develops in time according to probabilistic laws.

- The theory of stochastic processes gives us a formal way to look at time series variables.

- Time series is a realization or sample function from a certain stochastic process.

- A time series is a set of observations generated sequentially in time. Therefore, they are dependent to each other. This means that we do NOT have random sample.

Examples

- **Moving average process:** Let $\varepsilon_t \sim$ i.i.d.(0, 1), and $X_t = \varepsilon_t + 0.5\ \varepsilon_{t-1}$

- **Random Walk:** Let e1,e2,… be a sequence of i.i.d. rvs with 0 mean and variance σ_e^2 . The observed time series $\{Y_t, t = 1,2,\cdots,n\}$.

$$Y_t = e_1 + e_2 + \cdots + e_t$$

- Suppose that time series has the form: $Y_t = a + bt + e_t$

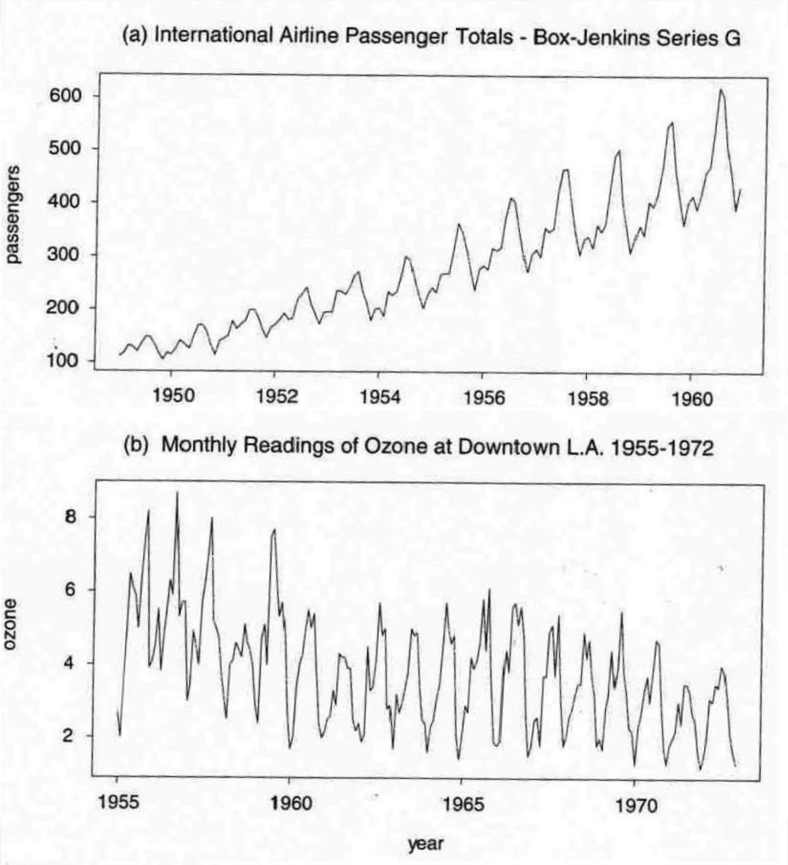
- Consider the time series with the form: $Y_t = (-1)^t e_t$

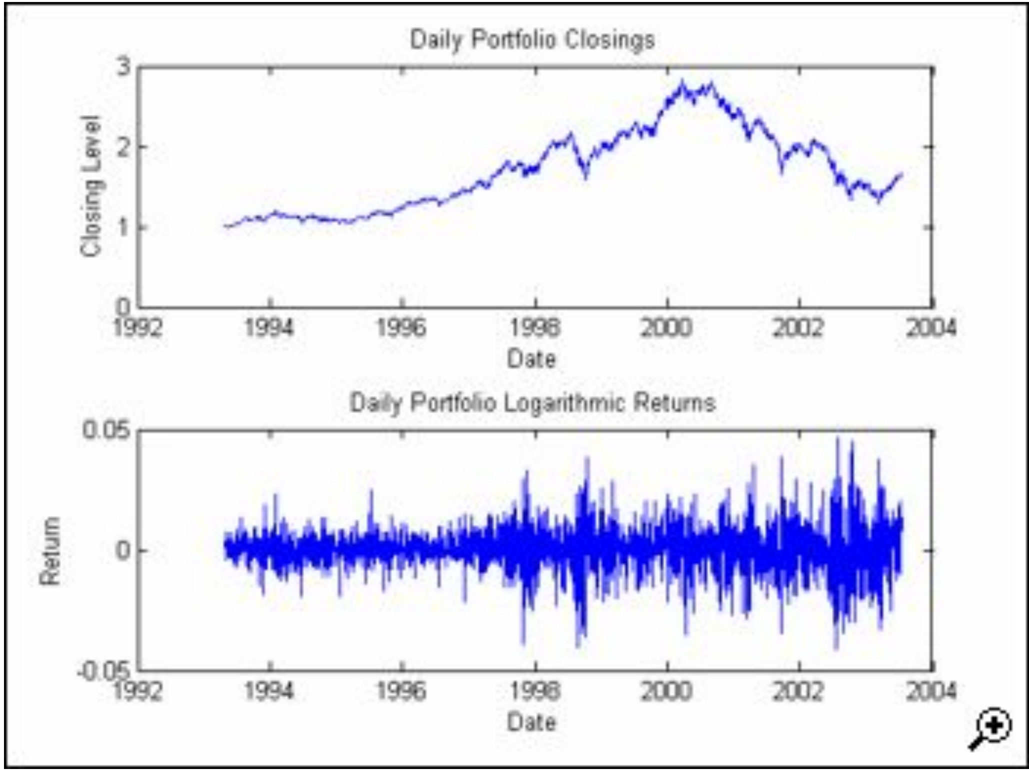
Domains of application

Data in business, economics, engineering, environment, medicine, earth sciences, and other areas of scientific investigations are often collected in the form of time series.

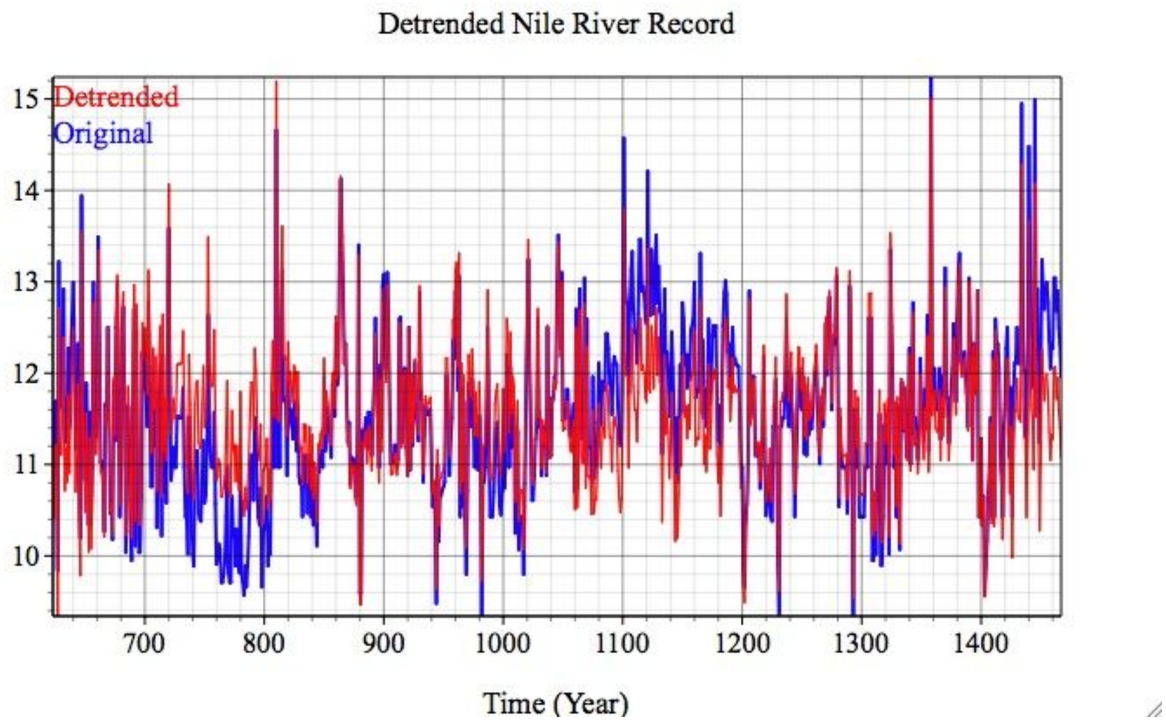
- Hourly temperature readings.
- Daily stock prices.
- Weekly traffic volume.
- Annual growth rate.
- Seasonal ice cream consumption.
- Electrical signals.

Domains of application





Domains of application



Seasonality, Trend and Stationarity

Trend

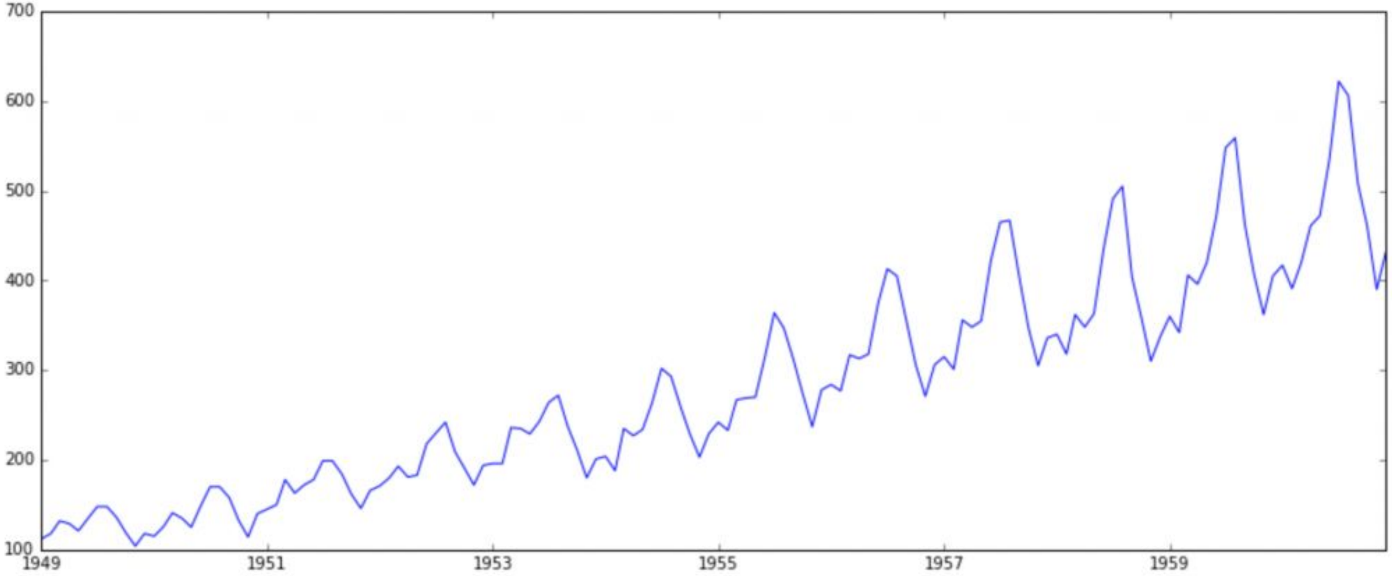
- A trend is a gradual shift or movement towards higher or lower values **over a long period of time.**
- Mathematically speaking, the expected value of a series is not constant over time:

$$\cancel{E[Y_t] = E[Y_{t+k}] \Rightarrow \mu_t = \mu_{t+k} = \mu, \forall t, k}$$

- When a trend pattern exhibits a general direction that is upward, where there are higher highs, and higher lows, we call it an uptrend.
- And vice versa, we call it a downtrend.

*When there’s no trend, we call a horizontal, or stationary trend.

Trend



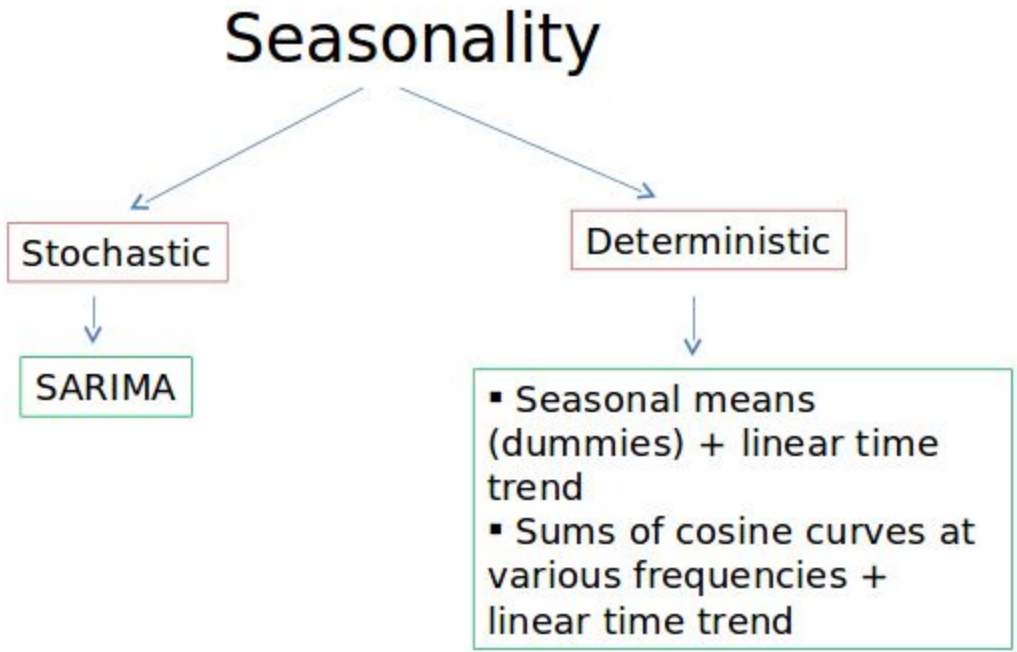
Seasonality

A TS that’s exhibiting a repeating pattern at fixed intervals of time (usually within a one year period) is said to have seasonal pattern or seasonality.
Examples:

- Seasonality of air conditioning cost (higher during summer, way lower during winter).
- A company selling coats will see jumps during winters.

“Business cycle" plays an important role in economics. In time series analysis, business cycle is typically represented by a seasonal (or periodic) model.
A smallest time period for this repetitive phenomenon is called a seasonal period, s.

Seasonality



Stationarity

Expected value of a series is constant over time, not a function of time

$$E[Y_t] = \mu, \forall t$$

The variance of a series is constant over time, homoscedastic.

$$Var[Y_t] = \sigma^2 < \infty, \forall t$$

$$Cov[Y_t, Y_{t-k}] = \gamma_k, \forall t$$

$$Corr[Y_t, Y_{t-k}] = \rho_k, \forall t$$

Not constant, not depend on time, depends on time interval, which we call "**lag**", k

→ We have constant mean and variance with covariance and correlation beings functions of the time difference only.

Why Stationarity?

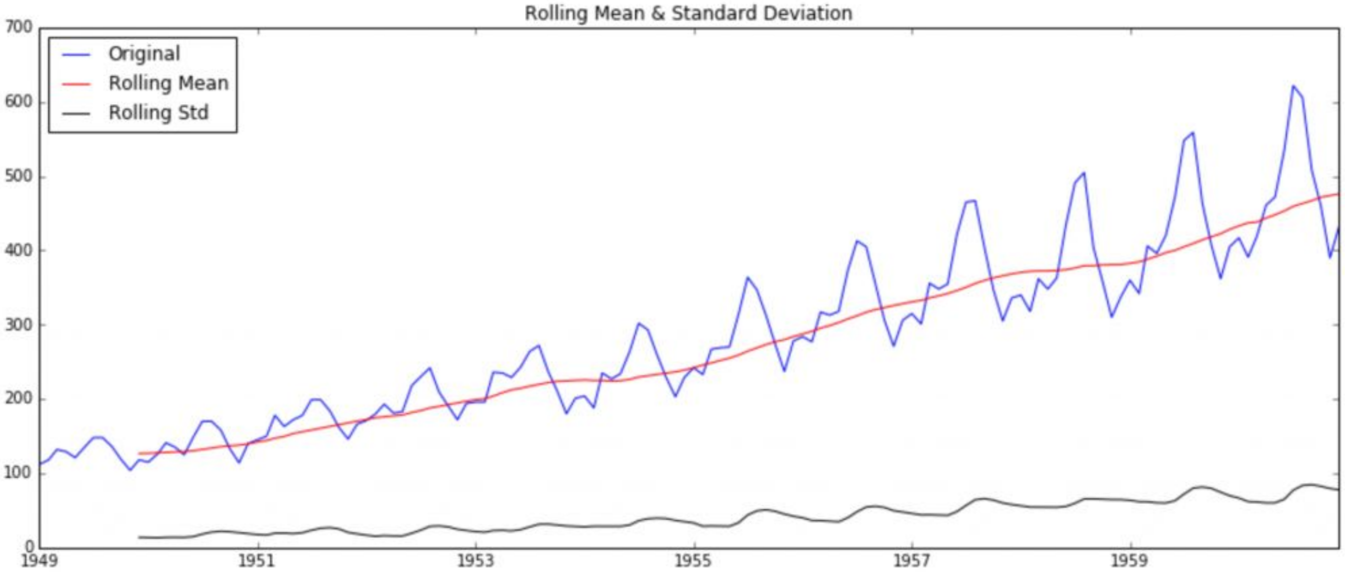
- Most of the TS models (ARMA or an ARIMA models) work on the assumption that the TS is stationary.
- Intuitively, we can see that if a TS has a particular behaviour over time, there is a very high probability that it will follow the same in the future
- The theories related to stationary series are more mature and easier to implement as compared to non-stationary series.

How to Check Stationarity of a TS?

1. **Plotting Rolling Statistics:** We can plot the moving average or moving variance and see if it varies with time. I.e. at any instant ‘t’, we’ll take the average/variance of the last year, last 12 months. **(This is more of a visual technique.)**

2. **Dickey-Fuller Test:** This is one of the statistical tests for checking stationarity. Here the null hypothesis is that the TS is non-stationary. The test results comprise of a Test Statistic and some Critical Values for difference confidence levels. If the ‘Test Statistic’ is less than the ‘Critical Value’, we can reject the null hypothesis and say that the series is stationary.

How to Check Stationarity of a TS?




```
Results of Dickey-Fuller Test:
Test Statistic                   0.815369
p-value                          0.991880
#Lags Used                      13.000000
Number of Observations Used    130.000000
Critical Value (5%)             -2.884042
Critical Value (1%)             -3.481682
Critical Value (10%)            -2.578770
dtype: float64
```

How to make a Time Series Stationary?

There are 2 major reasons behind non-stationaruty of a TS:

1. Trend: varying mean over time. For eg, in this case we saw that on average, the number of passengers was growing over time.

2. Seasonality – variations at specific time-frames. eg people might have a tendency to buy cars in a particular month because of pay increment or festivals.

Plan of attack

The underlying principle is to:

1. model or estimate the trend and seasonality in the series and remove those from the series to get a stationary series.

2. Statistical forecasting techniques can be implemented on this series.

3. The final step would be to convert the forecasted values into the original scale by applying trend and seasonality constraints back.

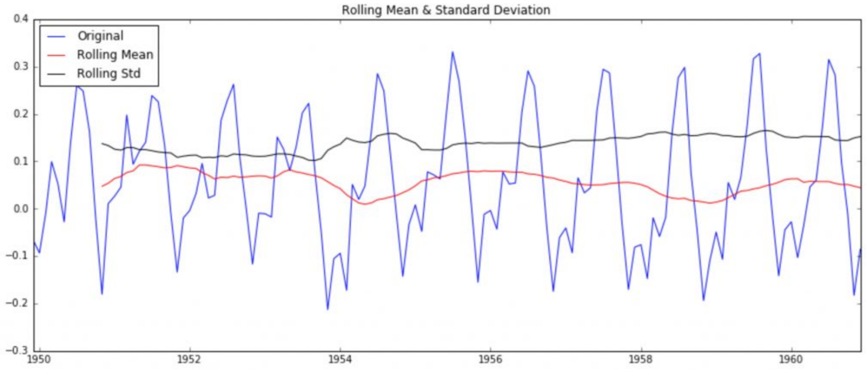
Estimating & Eliminating Trend and seasonality

There can be many ways of doing it and some of most commonly used are:

1. Apply **transformation** which penalize higher values more than smaller values. These can be taking a log, square root, cube root, etc.
2. **Aggregation:** taking average for a time period like monthly/weekly averages.
3. **Smoothing:** taking rolling averages [use **moving average**]
4. **Differencing:** taking the difference with a particular time lag
5. **Decomposition:** modeling both trend and seasonality and removing them from the model.

Moving average

In this approach, we take average of ‘k’ consecutive values depending on the frequency of time series. Here we can take the average over the past 1 year, i.e. last 12 values.

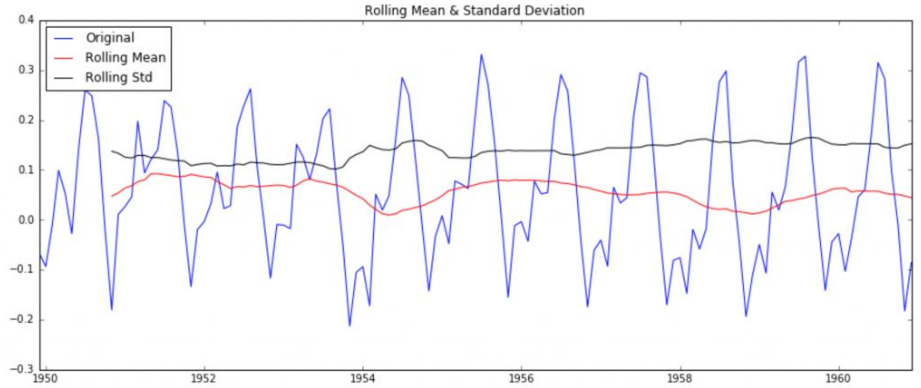



```
Results of Dickey-Fuller Test:
Test Statistic                 -3.162908
p-value                         0.022235
#Lags Used                     13.000000
Number of Observations Used   119.000000
Critical Value (5%)            -2.886151
Critical Value (1%)            -3.486535
Critical Value (10%)           -2.579896
dtype: float64
```

```
Month
1949-01-01          NaN
1949-02-01          NaN
1949-03-01          NaN
1949-04-01          NaN
1949-05-01          NaN
1949-06-01          NaN
1949-07-01          NaN
1949-08-01          NaN
1949-09-01          NaN
1949-10-01          NaN
1949-11-01          NaN
1949-12-01    -0.065494
Name: #Passengers, dtype: float64
```

Moving average

In this approach, we take average of ‘k’ consecutive values depending on the frequency of time series. Here we can take the average over the past 1 year, i.e. last 12 values.




```
Results of Dickey-Fuller Test:
Test Statistic                 -3.162908
p-value                         0.022235
#Lags Used                     13.000000
Number of Observations Used   119.000000
Critical Value (5%)            -2.886151
Critical Value (1%)            -3.486535
Critical Value (10%)           -2.579896
dtype: float64
```

```
Month
1949-01-01          NaN
1949-02-01          NaN
1949-03-01          NaN
1949-04-01          NaN
1949-05-01          NaN
1949-06-01          NaN
1949-07-01          NaN
1949-08-01          NaN
1949-09-01          NaN
1949-10-01          NaN
1949-11-01          NaN
1949-12-01    -0.065494
Name: #Passengers, dtype: float64
```

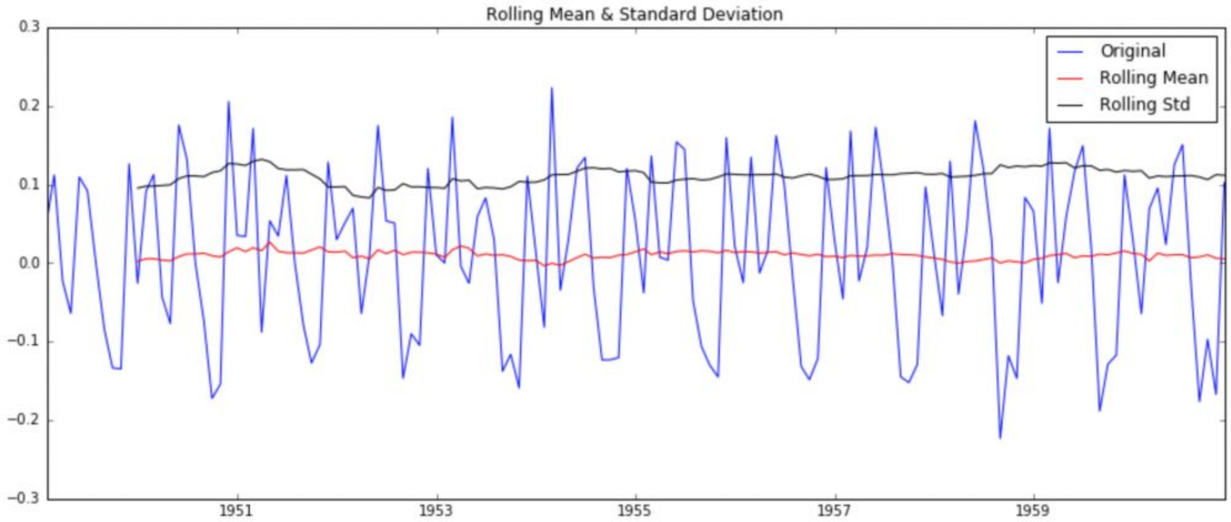
Differencing

One of the most common methods of dealing with both trend and seasonality is differencing.

In this technique, we take the difference of the observation at a particular instant with that at the previous instant. This mostly works well in improving stationarity. First order differencing can be done in Pandas as:

```
ts_log_diff = ts_log - ts_log.shift()
plt.plot(ts_log_diff)
```

Differencing

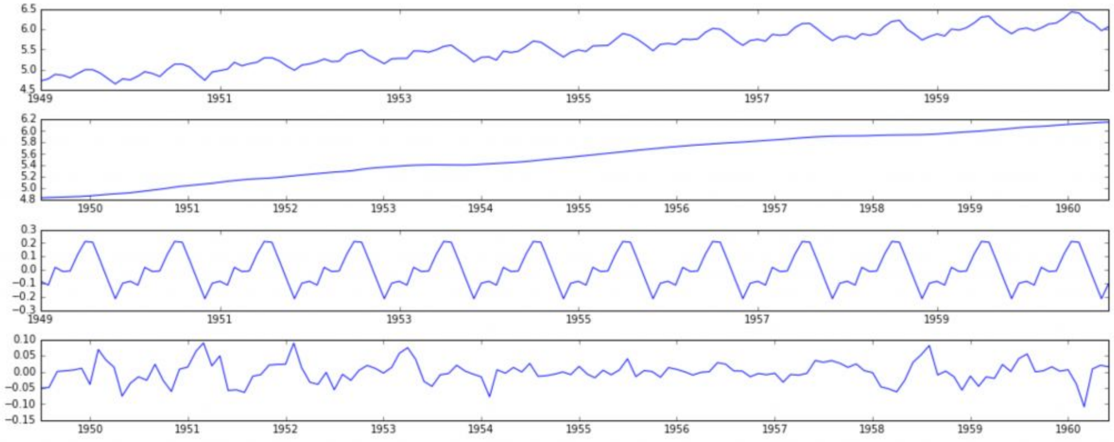



```
Results of Dickey-Fuller Test:
Test Statistic                 -2.717131
p-value                         0.071121
#Lags Used                     14.000000
Number of Observations Used   128.000000
Critical Value (5%)            -2.884398
Critical Value (1%)            -3.482501
Critical Value (10%)           -2.578960
dtype: float64
```

Decomposing

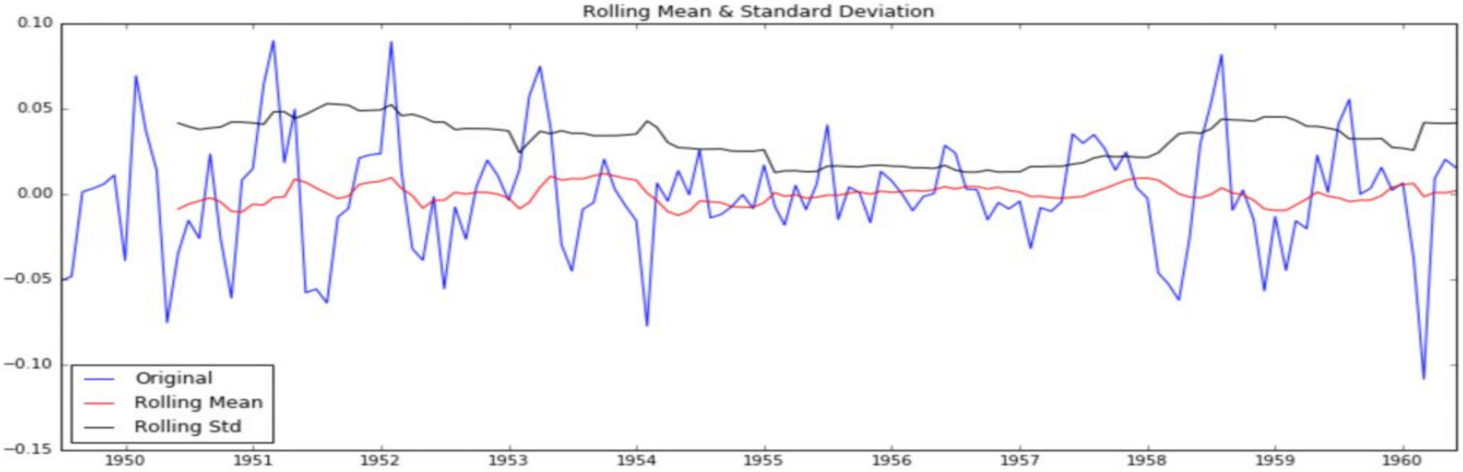
In this approach, both trend and seasonality are modeled separately and the remaining part of the series is returned:

Here we can see that the trend, seasonality are separated out from data and we can model the residuals.



Decomposing

Let’s check stationarity of residuals:




```
Results of Dickey-Fuller Test:
Test Statistic                -6.332387e+00
p-value                        2.885059e-08
#Lags Used                     9.000000e+00
Number of Observations Used    1.220000e+02
Critical Value (5%)           -2.885538e+00
Critical Value (1%)           -3.485122e+00
Critical Value (10%)          -2.579569e+00
dtype: float64
```

ARIMA Models

Auto Regressive AR model

The notation AR(p) indicates an autoregressive model of order p. The AR(p) model is defined as:

$$X_t = c + \sum_{i=1}^{p} \varphi_i X_{t-i} + \varepsilon_t$$

Where $\varphi_1, \ldots, \varphi_p$ are the parameters of the model, c is a constant, and ε_t is white noise.

→ The AR process is useful in describing situations in which the present value of a time series depends on its preceding values plus a random shock.

Auto Regressive AR model

Example AR(1):

$$X_t = c + \varphi X_{t-1} + \varepsilon_t$$

where ε_t is a white noise process with zero mean and constant variance.

This process sometimes called as the Markov process because the distribution of Yt given Yt-1,Yt-2,… is exactly the same as the distribution of Yt given Yt-1.

Differencing term I(d)

- Process used to transform a time series into a stationary one.
- Takes the difference between the consecutive terms to the order d.

E.g:

I(1): $Y(t) = X(t) - X(t-1)$

I(2): $Y'(t) = Y(t) - Y(t-1)$

...

Moving average MA(q)

The notation MA(q) refers to the moving average model of order q:

$$X_t = \mu + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

where μ is the mean of the series, the θ_1, ..., θ_q are the parameters of the model. and the ε_t, ε_{t-1},..., ε_{t-q} are white noise error terms.

The moving average process is always stationary.

Moving average MA(q)

Errors are the average of this period’s random error and last period’s random error.

No memory of past levels.

The impact of shock to the series takes exactly 1-period to vanish for MA(1) process. In MA(2) process, the shock takes 2-periods and then fade away. In MA(1) process, the correlation would last only one period.

ARIMA(p,d,q)

Amount of periods to lag for in ARIMA calculations.

Given a time series of data Xt where t is an integer index and the Xt are real numbers, an ARMA(p’,q) model is given by:

$$X_t - \alpha_1 X_{t-1} - \cdots - \alpha_{p'} X_{t-p'} = \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q}$$

ARIMA(p,d,q)

Number of AR (Auto-Regressive) terms (p): AR terms are just lags of dependent variable. E.g. if p is 5, the predictors for x(t) will be x(t-1)….x(t-5).

Number of MA (Moving Average) terms (q): MA terms are lagged forecast errors in prediction equation. E.g. if q is 5, the predictors for x(t) will be e(t-1)….e(t-5) where e(i) is the difference between the moving average at ith instant and actual value.

Number of Differences (d): These are the number of nonseasonal differences.

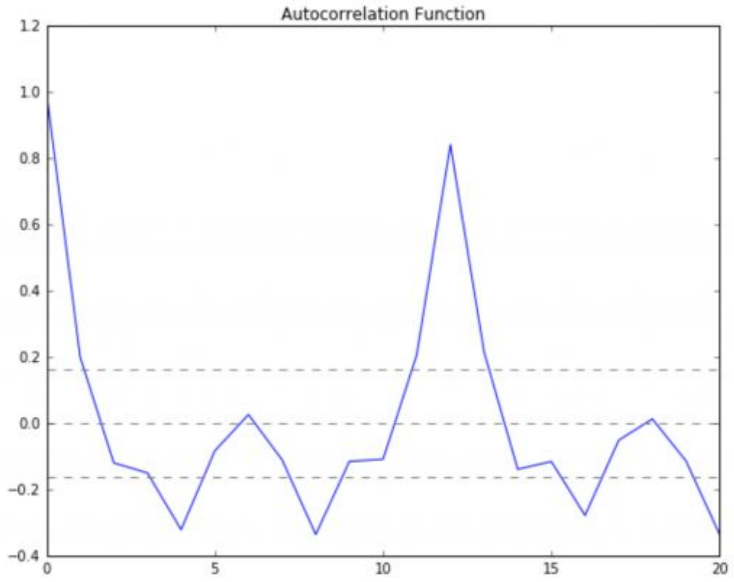
How to determine (p,d,q)?

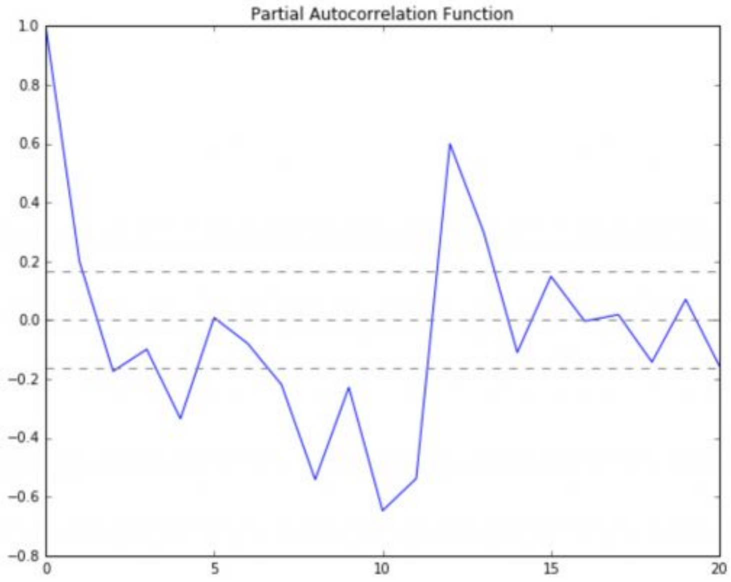
An importance concern here is how to determine the value of ‘p’ and ‘q’. We use two plots to determine these numbers:

Autocorrelation Function (ACF): It is a measure of the correlation between the the TS with a lagged version of itself. E.g. at lag 5, ACF would compare series at time instant ‘t1’…’t2’ with series at instant ‘t1-5’…’t2-5’ .

Partial Autocorrelation Function (PACF): This measures the correlation between the TS with a lagged version of itself but after eliminating the variations already explained by the intervening comparisons. Eg at lag 5, it will check the correlation but remove the effects already explained by lags 1 to 4.

How to determine (p,d,q)?





Confidence levels are given by $\pm 1.96/\sqrt{n}$ where n is the sample size.

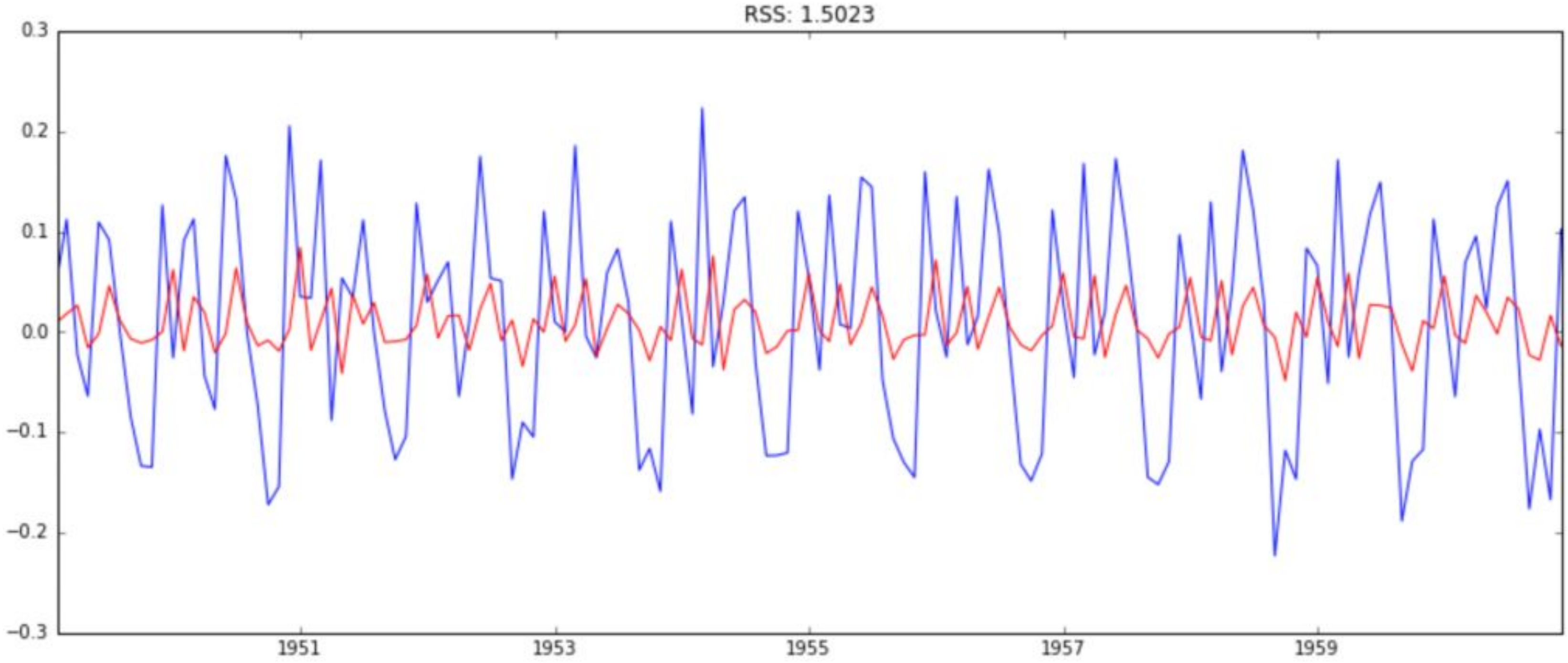
How to determine (p,d,q)?

In this plot, the two dotted lines on either sides of 0 are the confidence intervals. These can be used to determine the ‘p’ and ‘q’ values as:

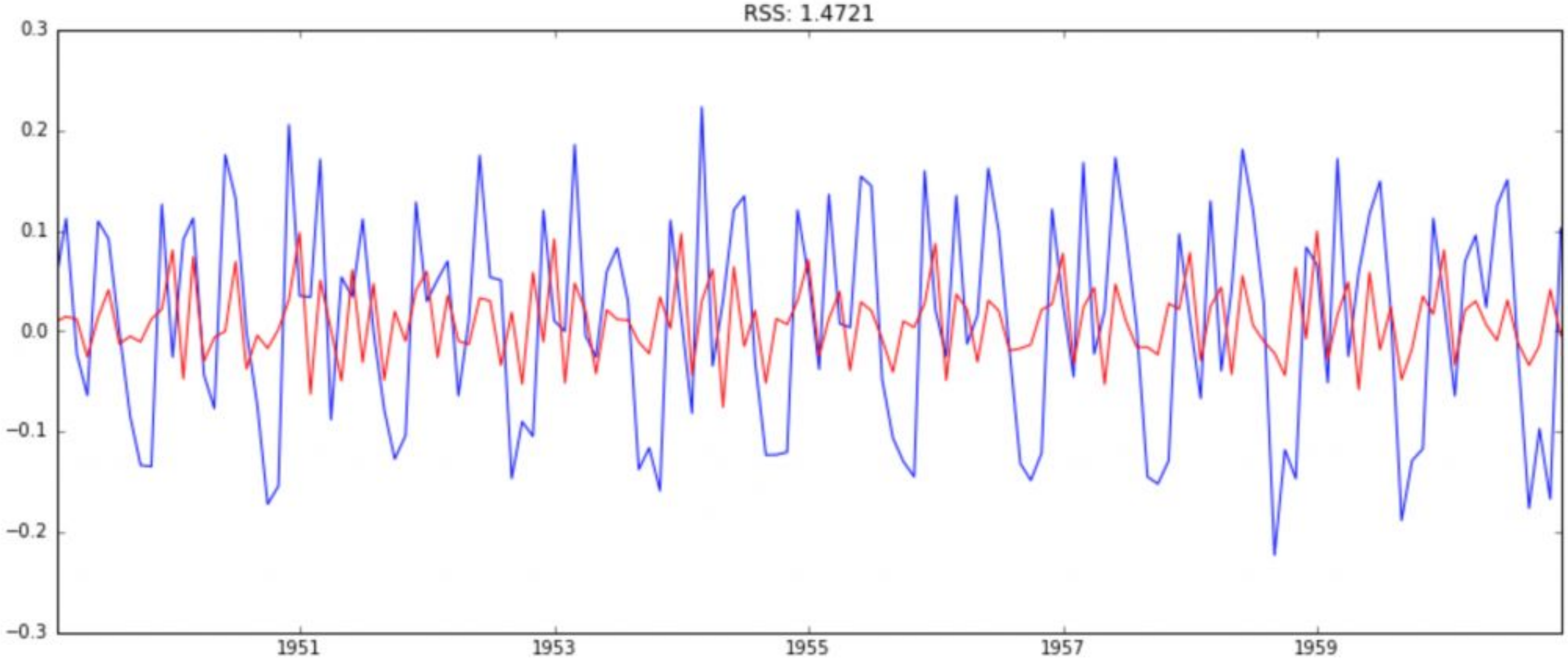
p : The lag value where the PACF chart crosses the upper confidence interval for the first time. If you notice closely, in this case p=2.

q : The lag value where the ACF chart crosses the upper confidence interval for the first time. If you notice closely, in this case q=2.

Evaluating the AR(2) model [ARIMA(2, 1, 0)]:

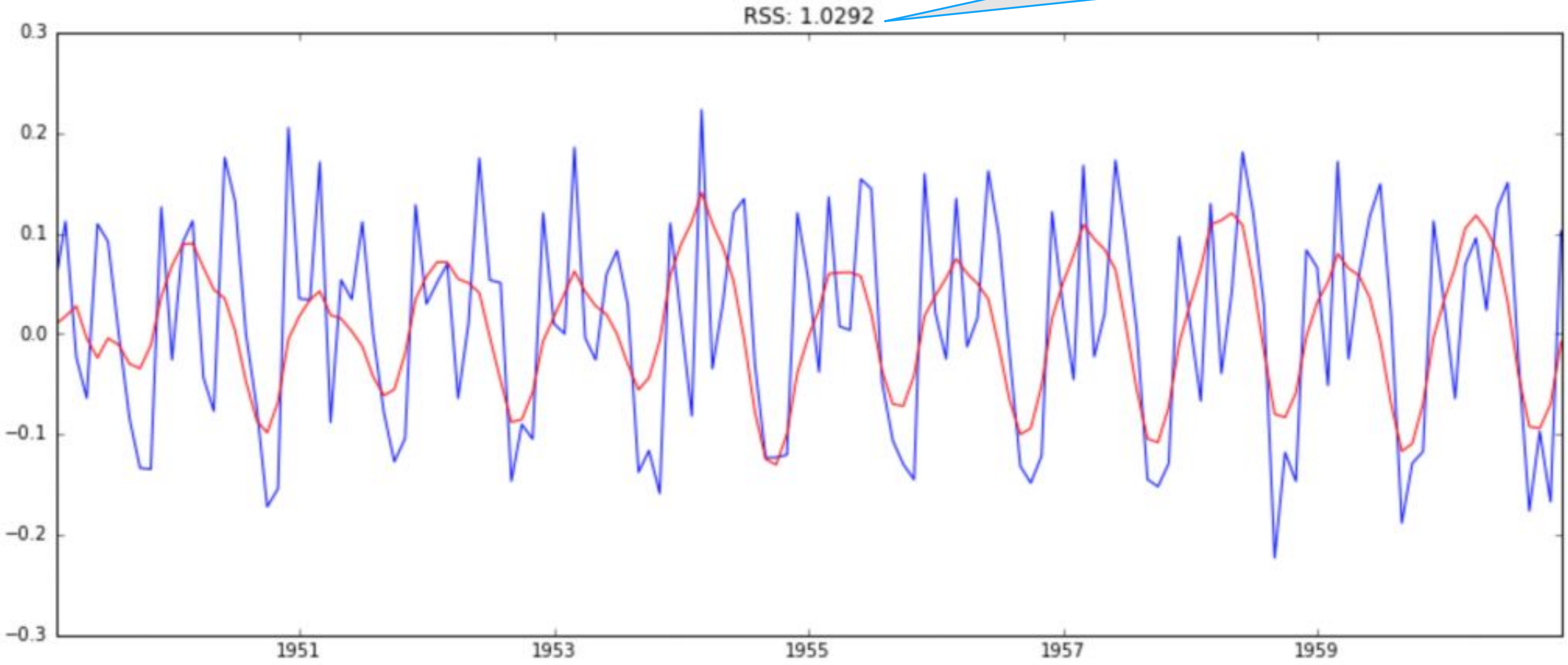


Evaluating the MA(2) model [ARIMA(0, 1, 2)]:



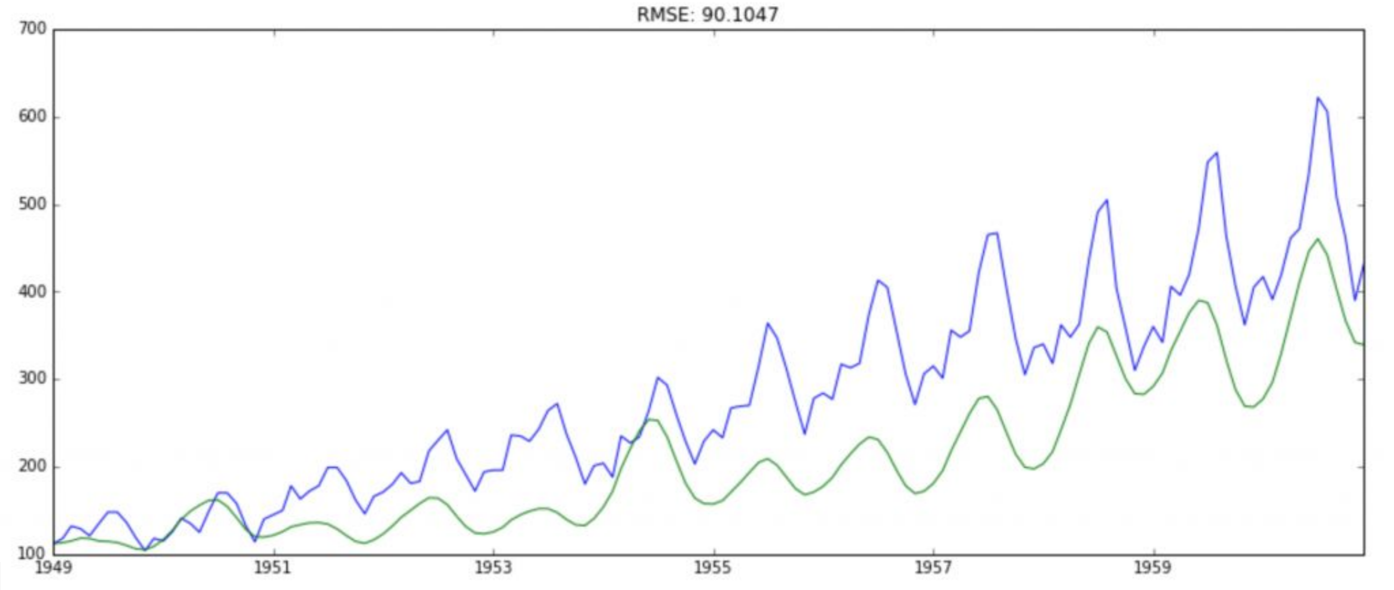
Evaluating combined model [ARIMA(2, 1, 2)]:

Here we can see that the AR and MA models have almost the same RSS but combined is significantly better.



Taking it back to the original scale

The way to convert the differencing to log scale is to add these differences consecutively to the base number:



Recommended Labs

Predict monthly count of riders for the Portland public transportation system: http://www.seanabu.com/2016/03/22/time-series-seasonal-ARIMA-model-in-python/

Airline passengers forecast example: https://www.analyticsvidhya.com/blog/2016/02/time-series-forecasting-codes-python/

Seasonal ARIMA models

Seasonal ARIMA(p, d, q)(P, D, Q)S

In a seasonal ARIMA model, seasonal AR and MA terms predict xt using data values and errors at times with lags that are multiples of S (the span of the seasonality).

With monthly data (and S = 12), a seasonal first order autoregressive model would use xt-12 to predict xt. For instance, if we were selling cooling fans we might predict this August's sales using last August's sales. (This relationship of predicting using last year's data would hold for any month of the year.)

Seasonal ARIMA(p, d, q)(P, D, Q)S

A seasonal second order autoregressive model would use x(t-12) and x(t-24) to predict xt. Here we would predict this August's values from the past two Augusts.

A seasonal first order MA(1) model (with S = 12) would use ε(t-12) as a predictor. A seasonal second order MA(2) model would use ε(t-12) and ε (t-24).

Model Identification

Model selection using criterion

- Besides ACF and PACF plots, we have also other tools for model identification.

- With messy real data, ACF and PACF plots become complicated and harder to interpret.

- Don’t forget to choose the best model with as few parameters as possible.

- It will be seen that many different models can fit to the same data so that we should choose the most appropriate (with less parameters) one and the information criteria will help us to decide this.

Model selection using criterion

The three well-known information criteria are:

- Akaike's information criterion (AIC) (Akaike, 1974)
- Schwarz's Bayesian Criterion (SBC) (Schwarz, 1978). Also known as Bayesian Information Criterion (BIC)
- Hannan-Quinn Criteria (HQIC) (Hannan&Quinn, 1979)

Akaike Information Criterion

Assume that a statistical model of M parameters is fitted to data:

$$AIC = -2\ln\left[\text{maximum likelihood}\right] + 2M.$$

For the ARMA model and n observations, the log-likelihood function:

$$\ln L = -\frac{n}{2}\ln 2\pi\sigma_a^2 - \frac{1}{2\sigma_a^2}\underbrace{S\left(\phi_p, \theta_q, \mu\right)}_{SS\,\mathrm{Residual}}$$

$$\left(\text{assuming } a_t \overset{i.i.d}{\sim} N\left(0, \sigma_a^2\right)\right).$$

Then, the AIC is: $AIC = n\ln\hat{\sigma}_a^2 + 2M$

Choose model (or the value of M) with minimum AIC.

Akaike Information Criterion

Lab to visit:
https://www.digitalocean.com/community/tutorials/a-guide-to-time-series-forecasting-with-arima-in-python-3

Exponential Smoothing

Simple Exponential Smoothing (SES)

- Suppressing short-run fluctuation by smoothing the series.
- Weighted averages of all previous values with more weights on recent values.
- Requires no trend, no seasonality.

For the observed time series, consider Y1, Y2, …, Yn. The equation for the model is:

$$S_t = \alpha Y_{t-1} + (1-\alpha) S_{t-1}$$

where : alpha is the smoothing parameter, 0< alpha <1.
Yt: the value of the observation at time t.
St: the value of the smoothed obs. at time t.

Simple Exponential Smoothing (SES)

Why Exponential?: For the observed time series Y1,Y2,…,Yn, Yn+1 can be expressed as a weighted sum of previous observations.

$$\hat{Y}_t(1) = c_0 Y_t + c_1 Y_{t-1} + c_2 Y_{t-2} + \cdots$$

where ci’s are the weights.

- Giving more weights to the recent observations, we can use the geometric weights (decreasing by a constant ratio for every unit increase in lag).

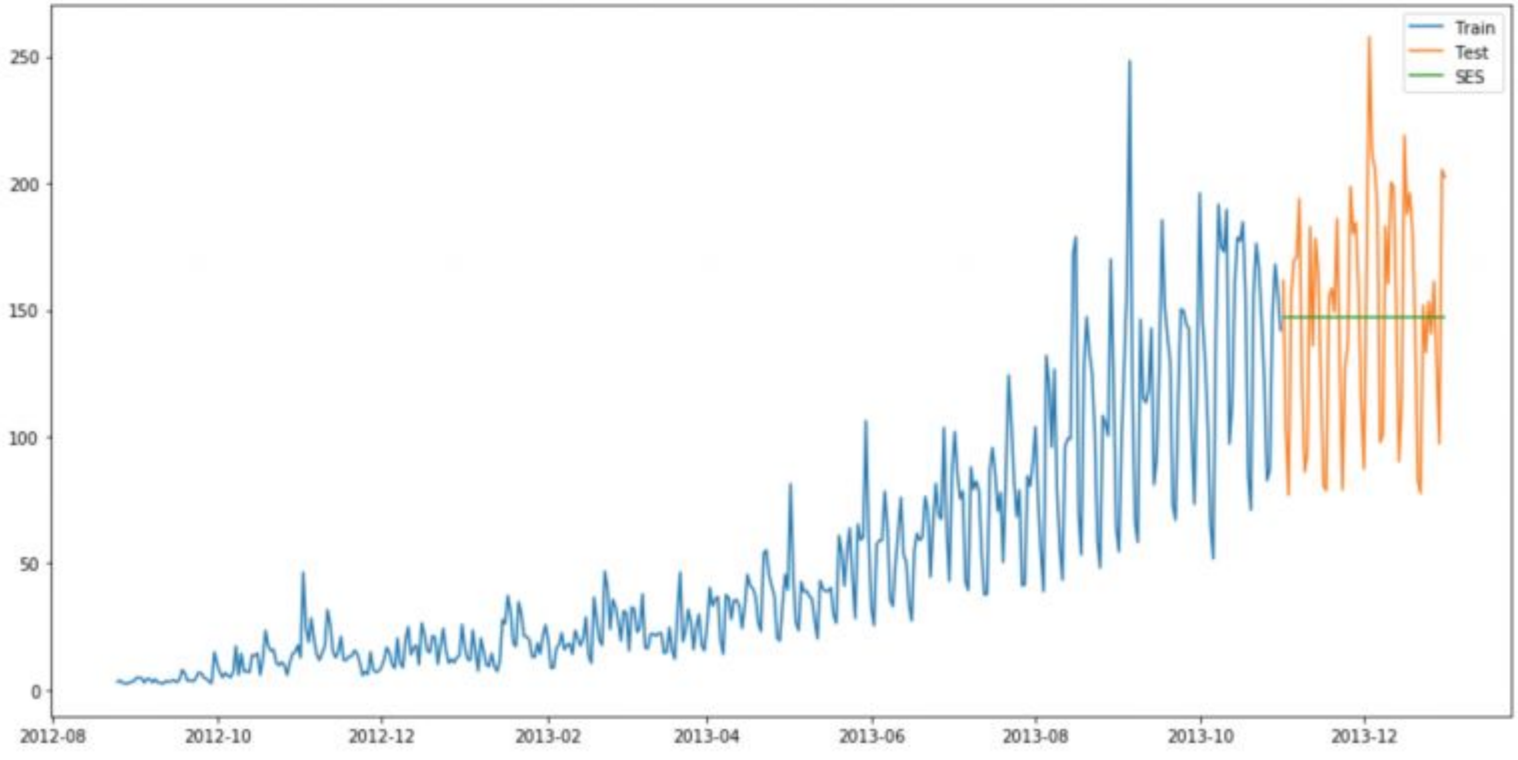
$$\Rightarrow c_i = \alpha(1-\alpha)^i ; i = 0,1,\ldots;0 \le \alpha \le 1.$$

Remarks on Alpha (smoothing parameter).

- Choose alpha between 0 and 1.
- If it is = 1, it becomes a naive model; if alpha is close to 1, more weights are put on recent values. The model fully utilizes forecast errors.
- If it is close to 0, distant values are given weights comparable to recent values. Choose close to 0 when there are big random variations in the data.
- Alpha is often selected as to minimize the MSE:

$$e_t = Y_t - S_t \Rightarrow min \sum_{t=1}^{n} e_t^2 \Rightarrow \alpha$$

Simple Exponential Smoothing (SES)



Holt's Linear Trend method (or DES)

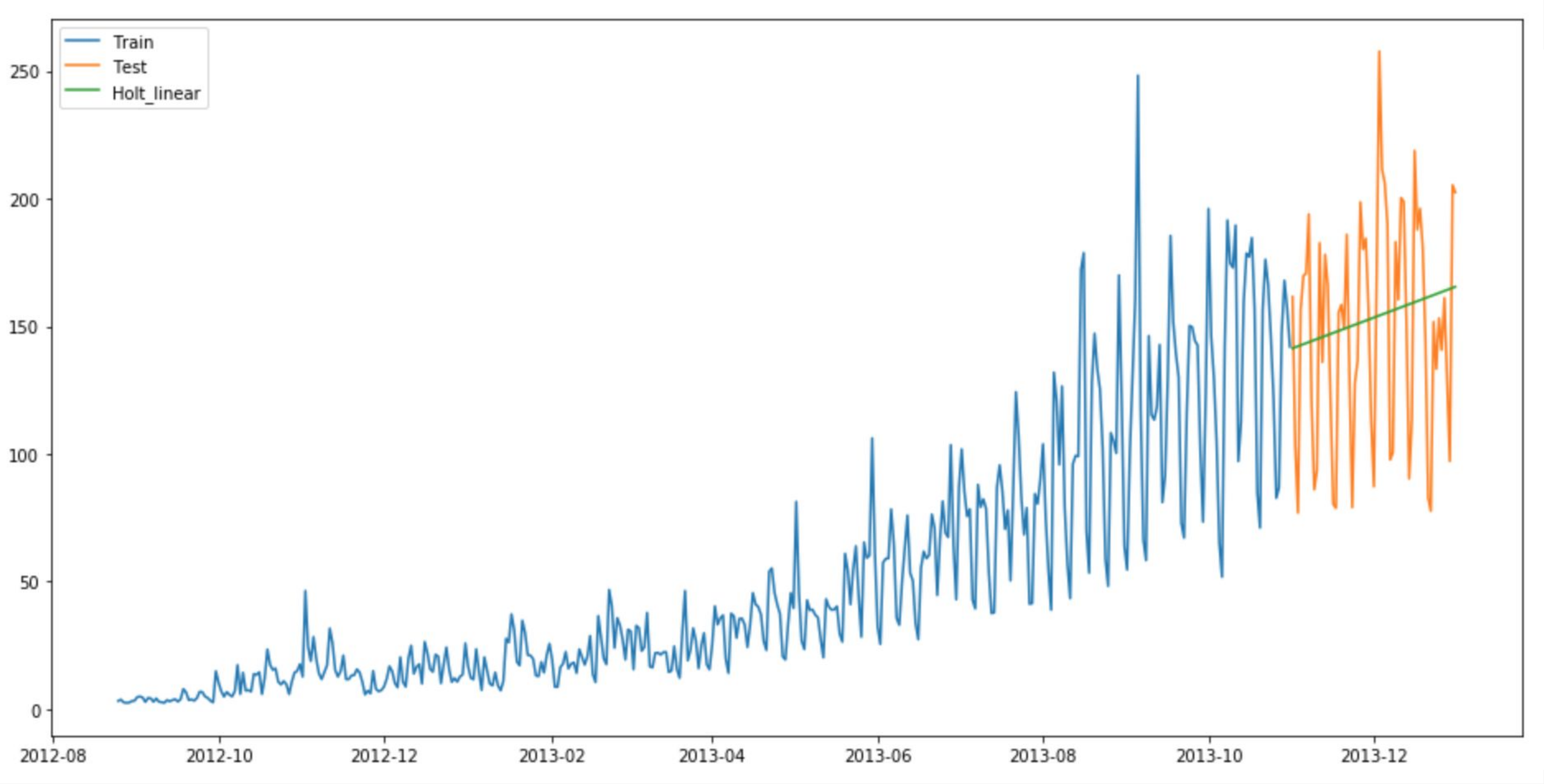
- Also called double exponential smoothing

Forecast equation: $\hat{y}_{t+h|t} = \ell_t + h\, b_t$

where l(t) is the level defined by: $\ell_t = \alpha y_t + (1-\alpha)(\ell_{t-1}+b_{t-1})$
And b(t) is the trend defined by: $b_t = \beta*(\ell_t - \ell_{t-1})+(1-\beta)b_{t-1}$

- Alpha and Beta values are taken by minimizing MSE as in SES, and $0 \leq \alpha \leq 1$, $0 \leq \beta \leq 1$.
- Holt extended simple exponential smoothing to allow forecasting of data with a trend.
- Consists in exponential smoothing applied to both level(the average value in the series) and trend.

Holt’s Linear Trend method (or DES)



Holt-Winters Method (or TES)

- Also called triple exponential smoothing

Forecast equation: $\text{forecast } F_{t+k} = L_t + kb_t + S_{t+k-s},$

where L(t) is the level defined by: $L_t = \alpha(y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1});$
And b(t) is the trend defined by: $b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1},$
And S(t) is the Seasonality defined by: $S_t = \gamma(y_t - L_t) + (1-\gamma)S_{t-s}$

- Alpha, Beta and Gamma values are taken by minimizing MSE as in SES, and $0 \le \alpha \le 1$, $0 \le \beta \le 1$ and $0 \le \gamma \le 1$.

- Holt extended simple exponential smoothing to allow forecasting of data with a trend and a seasonality effect.

Holt-Winters Method (or TES)