ReDI School of Digital Integration

Time Series Forecasting

Lecture 8

David Nagy, Mohan Sukumar

We use tech to connect human potential and opportunity with dignity & humility

Learning Objectives

- Recap - Time Series Analysis
- Time Series Forecasting
- Standard Forecasting models
- AR, MA, ARIMA
- ETS

Time Series Analysis Techniques

Differencing

Differencing

A technique used in time series analysis to remove the dependence of the observations on time.

- Help stabilize the mean of the time series by removing the trend.
- Help remove the seasonal component.
- After differencing, the resulting time series is said to be stationary.

Differencing

The first-order difference - is the difference between the current observation and the previous observation.

The second-order difference - is the difference between the first-order difference and the previous first-order difference, and so on..

diff() method in pandas to perform differencing.
(the default value of periods=1 is used to compute the difference between consecutive values)

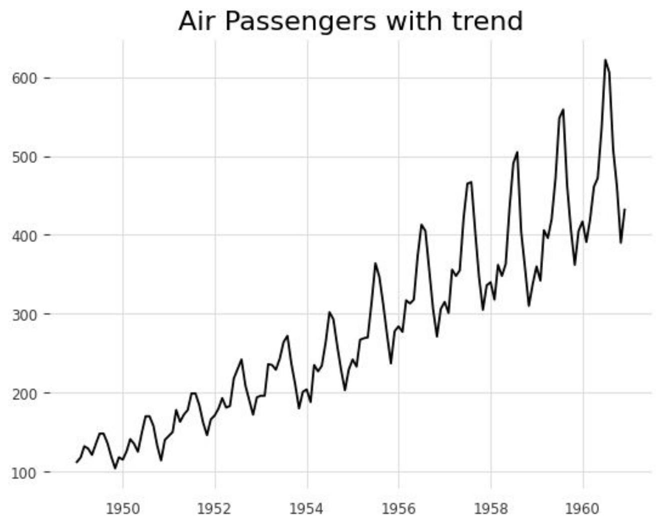
```python
import pandas as pd
import random

df = pd.DataFrame({"ts": random.sample(range(10, 30), 8)})

df["diff"] = df.diff()

df
```

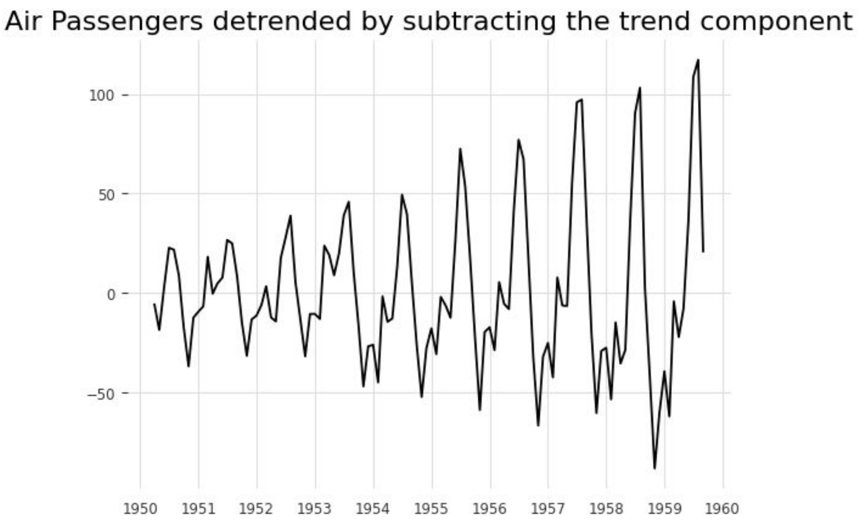
	ts	diff
0	24	NaN
1	13	-11.0
2	27	14.0
3	17	-10.0
4	10	-7.0
5	23	13.0
6	20	-3.0
7	26	6.0

Detrend and Deseasonalize

Detrending

It means to remove the trend component from the time series





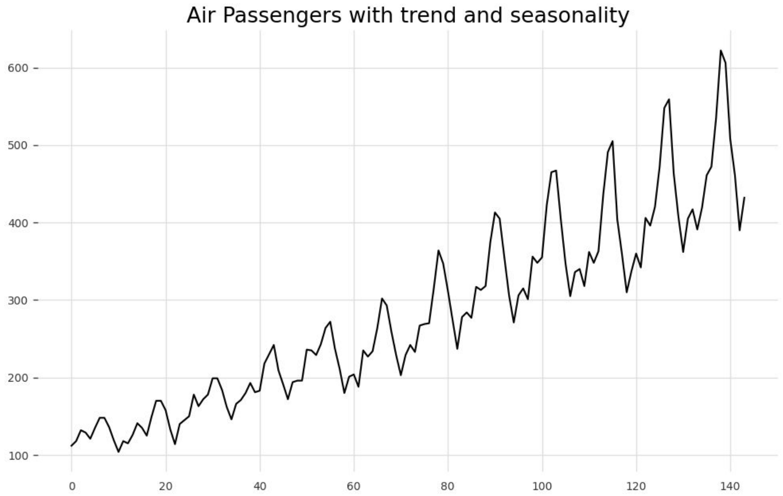
Detrending

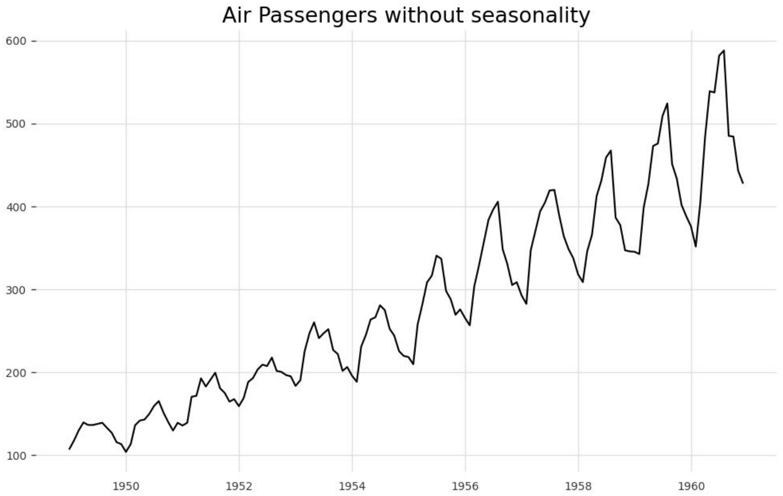
It means to remove the trend component from the time series

```python
# Using statmodels: Subtracting the Trend Component
from statsmodels.tsa.seasonal import seasonal_decompose
result_mul = seasonal_decompose(df['#Passengers'], model='multiplicative', period=30)
detrended = df['#Passengers'].values - result_mul.trend
plt.plot(detrended)
plt.title('Air Passengers detrended by subtracting the trend component', fontsize=16)
```

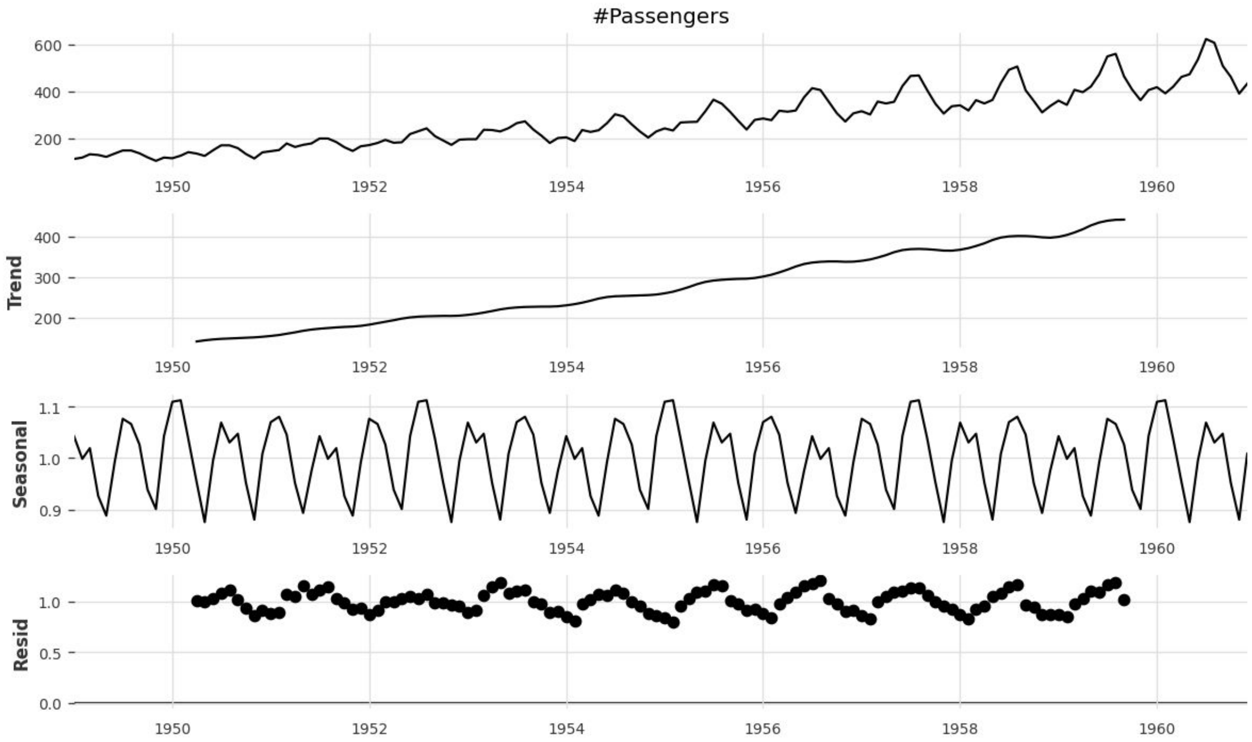
Detrending

It means to remove the seasonal component from the time series



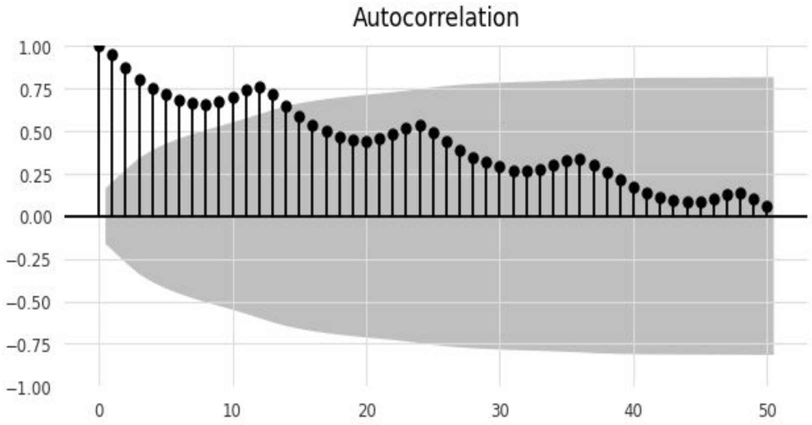


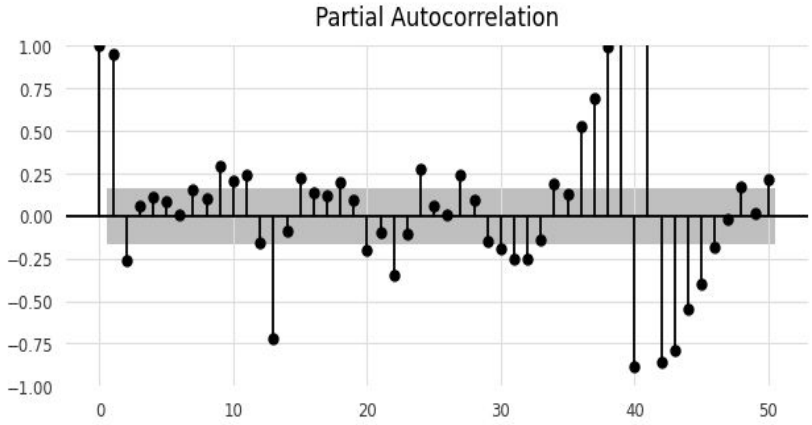
Complete decomposition



Autocorrelation

Autocorrelation and Partial Autocorrelation



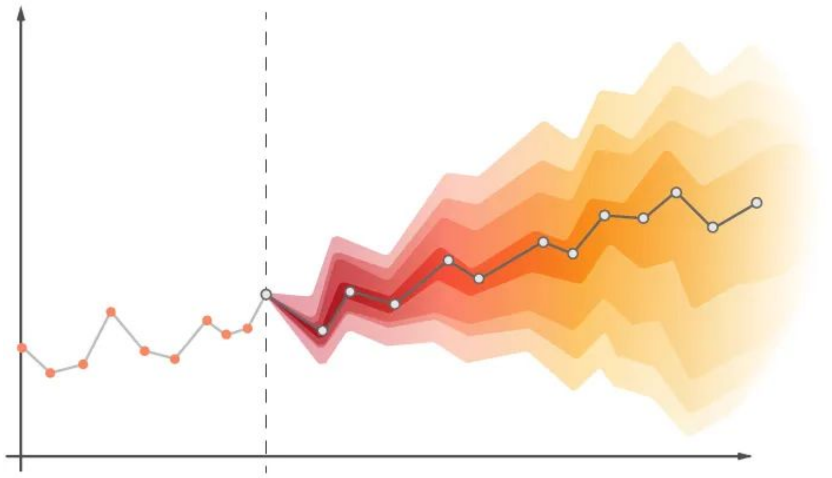


Break

Time Series Forecasting

What is forecasting

- Future predictions based on the time series data analysis.
- Understand the time series characteristics like trend, seasonality etc
- Identify the best method to make the time series stationary
- Reverse transformation of data is possible



Training and testing

- **training set**—a subset to train a model.
- **test set**—a subset to test the trained model.

Make sure that your test set meets the following two conditions:

- Is large enough to yield statistically meaningful results.
- Is representative of the data set as a whole. In other words, don't pick a test set with different characteristics than the training set

Standard Forecasting Models

Auto Regression

AR - Auto Regression

Uses past values(lags) of the forecast variable to predict future values.

$$\mathbf{y_t = c + \Sigma(\varphi_i * y_{t-i}) + \varepsilon_t}$$

where:

$\mathbf{y_t}$ is the value of the variable at time t

c is a constant term

$\boldsymbol{\varphi_i}$ is a numeric constant by which we multiply the lagged variable $\mathbf{y_{t-i}}$

$\boldsymbol{\varepsilon_t}$ is the error term at time t

Moving Average

MA - Moving Average

Help smooth out short-term fluctuations and highlight longer-term trends

Simple Moving Average (SMA) - average of a fixed number of data points over a specified time period.

Weighted Moving Average (WMA) - Assigns a weight to each data point based on its position in the time period. Help reduce the influence of older data points.

Exponential Moving Average (EMA) - Assigns exponentially decreasing weights to the data points.
Recent data points have a higher weight than older ones, but all data points contribute to the calculation.

ARIMA

ARIMA - the order

Represented by three parameters: p, d, and q.

- p - number of autoregressive (AR) terms
- d - number of differences (d)
- q - number of moving average (MA) terms

ARIMA(1,1,1) model : $y_t - y_{t-1} = c + \varphi_1 * (y_{t-1} - y_{t-2}) + \theta_1 * (e_{t-1})$

Where:

$\mathbf{y_t}$ is the value of the time series at time t
$\mathbf{y_{t-1}}$ is the value of the time series at time t-1
$\mathbf{e_{t-1}}$ is the error term at time t-1
$\boldsymbol{\varphi_1}$ and $\boldsymbol{\theta_1}$ are the autoregressive and moving average coefficients respectively

Representation of ARIMA of order (p,d,q) in python

```
from statsmodels.tsa.arima.model  import ARIMA

model  = ARIMA(y,  order=(2,1,2))
result = model.fit()

print(result.params)
```

"params" is a vector or array that contains the estimated values of the AR, MA, and differencing parameters of an ARIMA model, which can be used to make forecasts of future values.

Arima model limitations

- Real-world time series exhibit nonlinear patterns and dependencies that cannot be captured by linear models.

- Stationarity assumption.

- Memory-based models that only use a finite number of past values of the time series to make predictions

ARIMA

Combines the advantages of Autoregressive (AR), Moving Average (MA), and differencing techniques(I).

- Confirm the stationarity with ADF test.
- By examining the ACF and PACF plots, identify the order of the AR and MA components, as well as the level of differencing required to make the time series stationary.
- Fit the Model.
- Validate the Model.
- Forecast values.
- Evaluate the Model.

SARIMA

SARIMA

Includes the seasonal component.

Notation for a SARIMA model is (p, d, q)(P, D, Q)s

where:
p: the order of the autoregressive (AR) term
d: the order of differencing required to make the time series stationary
q: the order of the moving average (MA) term

P: the order of the seasonal autoregressive (SAR) term
D: the order of seasonal differencing required to make the time series stationary
Q: the order of the seasonal moving average (SMA) term
s: the length of the seasonal cycle (e.g., 12 for monthly data with annual seasonality)

SARIMA Model approach

To fit a SARIMA model, the same approach as ARIMA can be used.

- Values of p, d, and q are selected based on the ACF and PACF plots
- Values of P, D, Q, and s are selected based on the seasonal ACF and PACF plots.
- The residuals are checked for stationarity and autocorrelation.

WE DID IT!