Time Series

Fundamentals

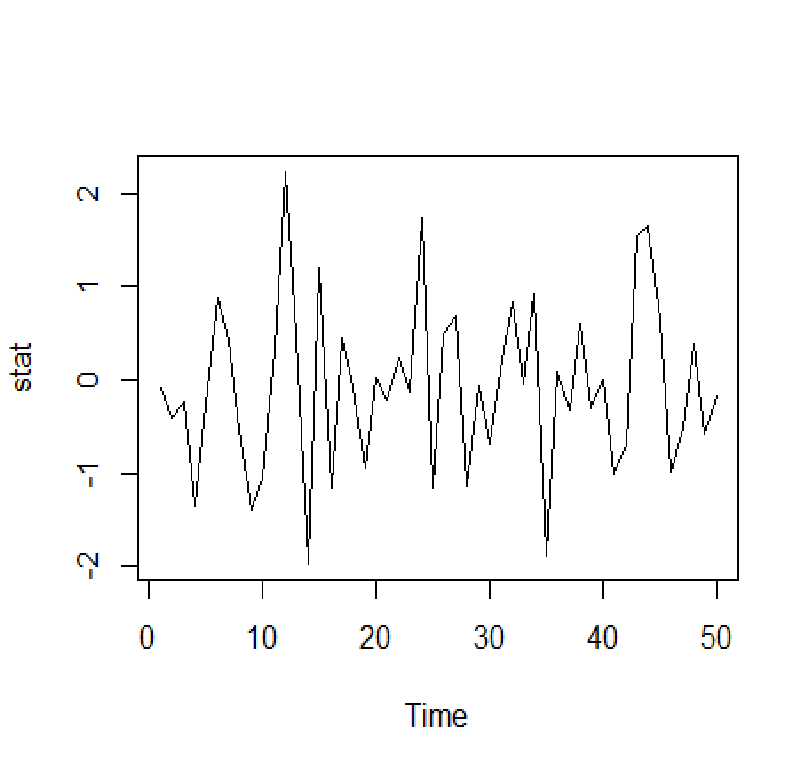
Sane's STATS Academy of Statistics

Stationary Process

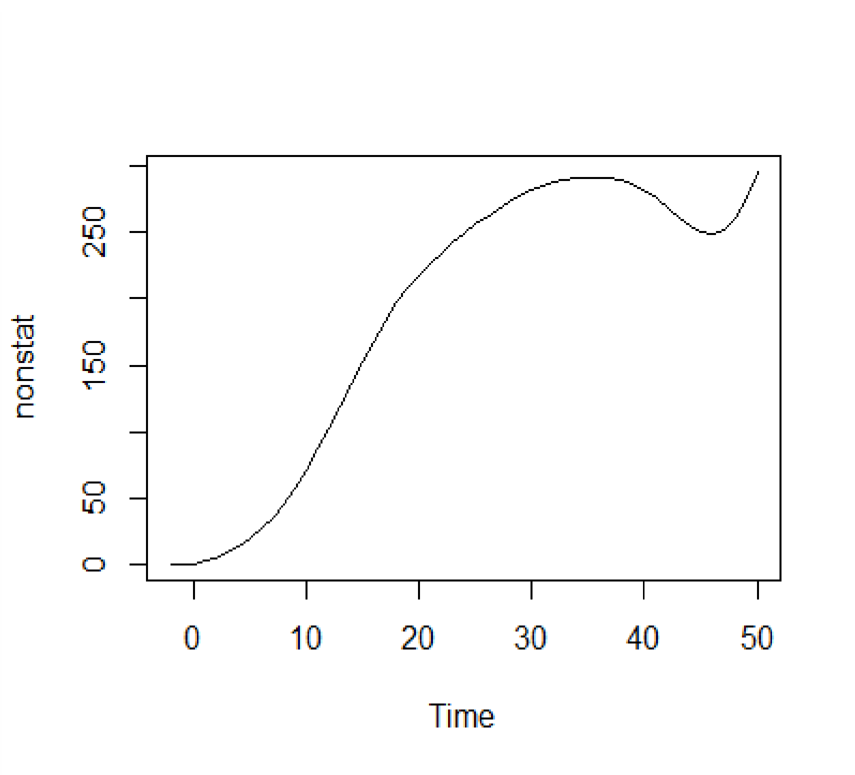
- Stationary process is that stochastic (probabilistic) process whose joint probability distribution does not change when shifted in time.
- In our context of time series, it is that time series whose mean and variance do not change over time.
- White Noise Model is the simplest example of Stationary series.
- For weak stationarity, covariance of y_t and y_s is constant for all $|t - s| = h$, for all h. e.g. $Cov(y_3, y_7) = Cov(y_{22}, y_{26})$

Stationary and Non-Stationary

Stationary



Non-Stationary



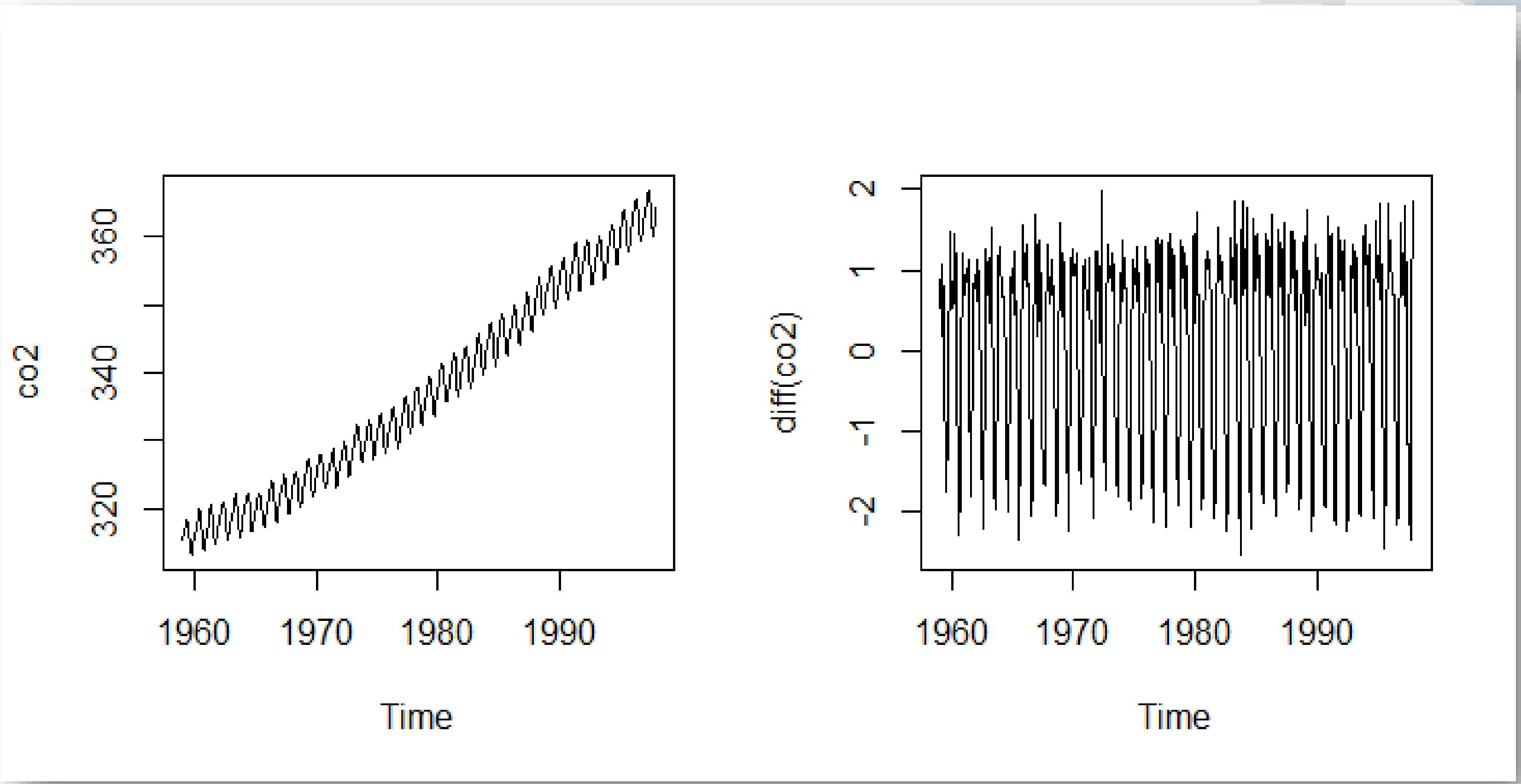
White Noise Model (WN Model)

- WN Model is a simple example of stationary process
- A weak White Noise has
  - A fixed constant mean
  - A fixed constant variance
  - No correlation of any time point value with any time point value

Random Walk (RW) Model

- RW Model is a simple example of non-stationary time series
- A random walk series has
  - No specific mean and variance
  - Strong dependence over time
- Changes or increments in RW series are white noise
- Random Walk Recursion: Today's value = Yesterday's Value + Noise
- In other words, $y_t = y_{t-1} + \in_t$, where $\in_t$ is white noise with mean zero
- RW Model has only one parameter i.e. variance of the white noise σ_ϵ^2

Example of RW Model



How to find Stationarity?

- Dickey-Fuller Test can be used to test the stationarity of any time series
- Consider the expression of auto-regressive model

$$y_t = \beta y_{t-1} + \epsilon_t$$

Dickey–Fuller test checks whether the β in the expression above is 1 or less than 1

$H0$: $\beta = 1$ (the time series is non-stationary)

HA: $\beta < 1$ (the time series is stationary)

Dickey-Fuller Test in Python

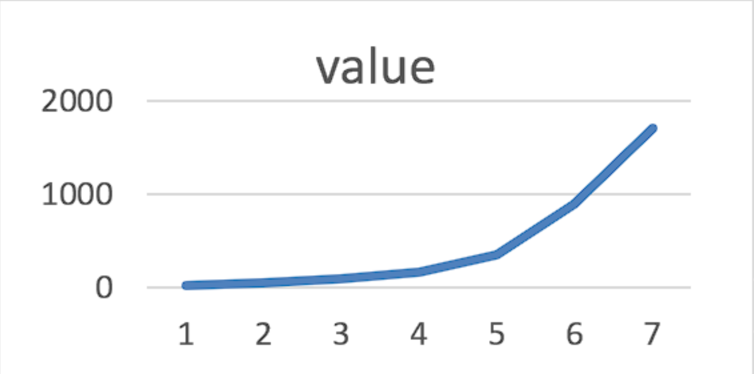
- *statsmodels.tsa.stattools.adfuller* is a Dicky–Fuller test and returns test statistics and p-value for the test of the null hypothesis.
- If the p-value is less than 0.05, the time series is stationary.

```
In [28]: result = adfuller(df['GasProd'], maxlag=10)
    ...: print("P-Value =", result[1])
    ...: if result[1] < 0.05:
    ...:     print("Time Series is Stationary")
    ...: else:
    ...:     print("Time Series is not Stationary")
P-Value = 0.9981674130928889
Time Series is not Stationary
```

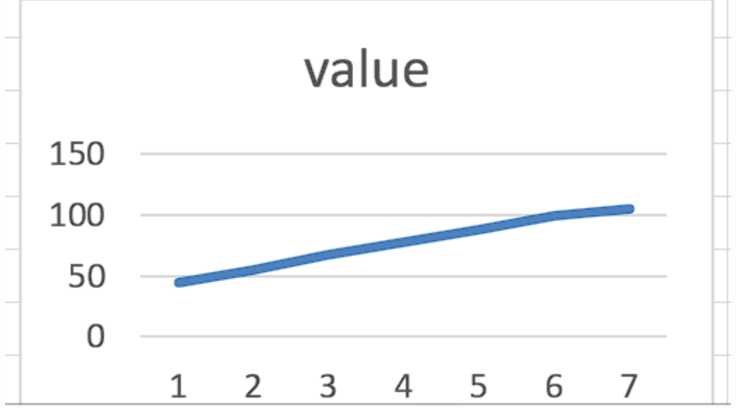
What can be done for stationarity?

- We can difference the time series

value	diff 1st	2nd	3rd
23			
44	21		
89	45	24	
157	68	23	-1
350	193	125	102
890	540	347	222
1706	816	276	-71



value	diff 1st
45	
55	10
67	12
78	11
88	10
100	12
105	5



Autocorrelation

What is Autocorrelation?

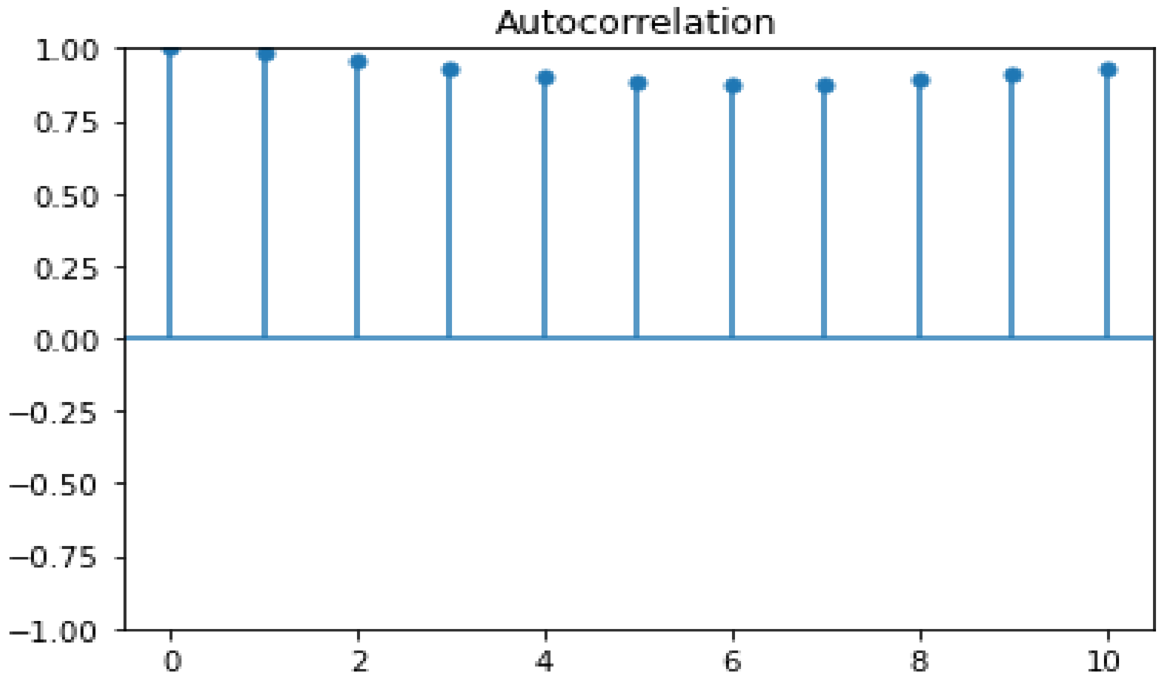
- Autocorrelation is correlation between the elements of a series and others from the same series separated from them by a given interval.
- Lag 1 Autocorrelation: Correlation of today's value with yesterday's value
- Lag 2 Autocorrelation: Correlation between today's and day before yesterday's values
- Lag k Autocorrelation: Correlation between Day 1 with Day k values

Auto-correlation(ACF) Formula

- Let y_t be value of time series at time t. The ACF between the series y_t and y_{t-h}correlation can be expressed as

$$\frac{Covariance(y_t, y_{t-h})}{Variance(y_t)}$$

Calculating acf



- We observe that, the correlation goes on decreasing with the increase in the lag. This is not the case with every time series.

Autoregressive Models

AR Process

Autoregressive Model

- In this model, we consider that today's observation is regressed on yesterday's observation or any of the previous day's observation.
- Model:

$$Today's\ Value = Constant + Slope * Yesterday's\ Value + Noise$$

- Software may use mean centered version of this model as

$$(Today's\ Value - Mean) = Slope * (Yesterday's\ Value - Mean) + Noise$$

- By notations, $y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t$, where ϵ_t is a white noise with mean 0 with variance σ_ϵ^2 and ϕ and μ are the slope and mean respectively

AR Process

$$y_t - \mu = \phi(y_{t-1} - \mu) + \epsilon_t$$

- If slope $\phi = 0$ then $y_t = \mu + \epsilon_t$ and y_t will be white noise with mean μ and variance σ_ϵ^2
- If slope $\phi \neq 0$ then the process of $\{y_t\}$ is autocorrelated
- Large value of Ø implies greater dependency of current values with previous values
- Negative value of Ø implies oscillatory time series
- If $\mu = 0$ and slope $\phi = 1$, then $y_t = y_{t-1} + \epsilon_t$, which is a random walk process

Simple Moving Average Model

MA Process

Simple Moving Average Model

- Simple MA model:

$$T\text{oday}'\text{s Value} = \text{Mean} + \text{Noise} + \text{Slope} * (\text{Yesterday}'\text{s Noise})$$

- In mathematical notations,

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

Where

μ: Mean of the series

θ: Slope

ϵ_t: Error or Noise at time t which has mean 0 and some variance σ_ϵ^2

- At $\theta = 0$, the model will be a white noise with mean μ and variance σ_ϵ^2

Simple Moving Average Model

$$y_t = \mu + \epsilon_t + \theta\epsilon_{t-1}$$

- If θ is non-zero then y_t depends on both ϵ_t and ϵ_{t-1} and the process is auto correlated
- Larger values of θ imply greater autocorrelation
- Negative values of θ imply oscillatory time series

Questions?