- Introduction
- Time Series
- Time Series Models
- Questions?

Time Series Analysis: An Introduction with Python

Group 5

Hamed Zakeri, Sina Davari, Gurkaran Sran, Farshad Tajeddini
Samiramis Khazaei, Anbumanivel Mohan Suganthi, Dhilip Ilanseran

January 2023

Questions?

Time Series Models

Time Series

Introduction

Outlines

- **Introduction**
  - Definition of Time Series Data
  - Basic Concepts of Time Series Analysis (Stationarity, White Noise, Autocorrelation, etc.)
- **Time Series Characteristics**
  - Patterns (Trend, Seasonality, Cycle)
  - Decompositions
- **Time Series Models**
  - Mean Model
  - Linear Trend Model
  - Random Walk Model
  - Moving Average Models (ARMA & ARIMA, SARIMA, Exponential Smoothing)
  - Autoregressive Models (ARCH & GARCH)

Time Series Data

Questions?

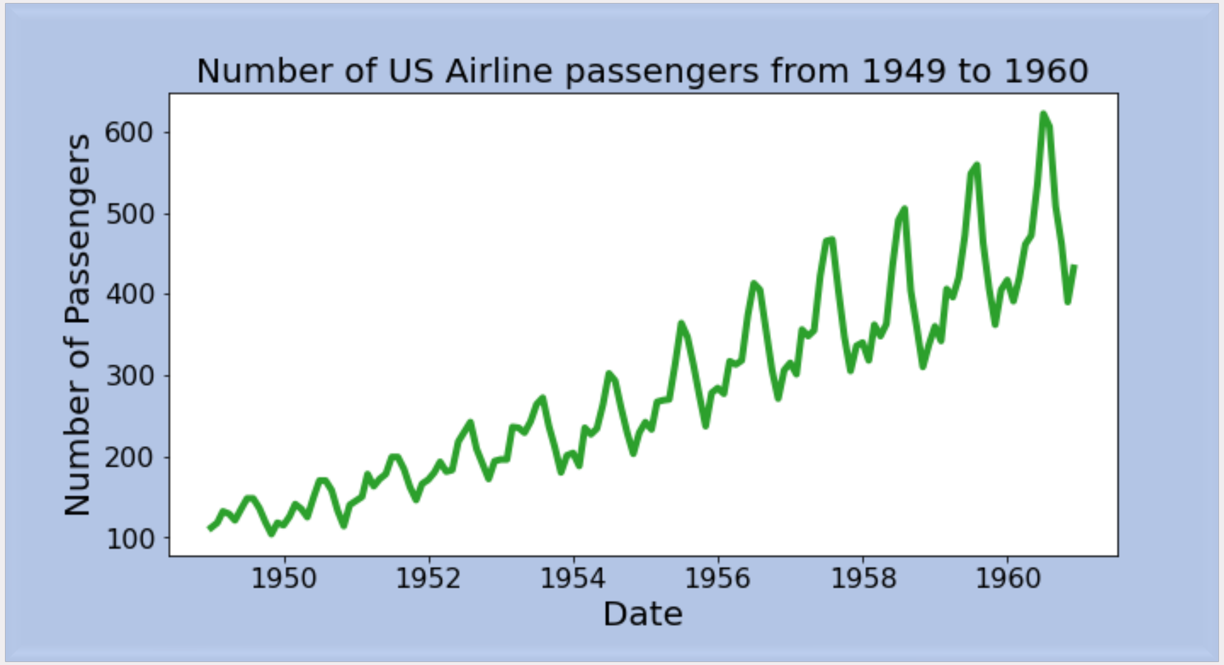
Time Series Models

Time Series

Introduction

What is "Time Series Data?"

Data recorded ***over time***, typically at regular time intervals (e.g., hourly, monthly, daily, etc.)



Types of Data

Country	time	# of flights	# of passengers
UK	2010	1456	728,000
Canada	2010	1356	678,000
Japan	2010	1267	633,500

Cross-sectional Data

Observations from multiple individuals at the same point in time or at different points in time (time does not play any significant role in the analysis).

Country	Time	# of flights	# of passengers
UK	2010	1456	728,000
UK	2015	5556	3,889,200
UK	2020	9567	9,567,000

Time series data

A sequence of data from a natural or social process (or a trial) observed over time.

Country	Time	# of flights	# of passengers
UK	2010	1456	728,000
UK	2015	5556	3,889,200
UK	2020	9567	9,567,000
Canada	2010	1356	678,000
Canada	2015	6045	4,231,500
Canada	2020	11567	11,567,000
Japan	2010	1267	633,500
Japan	2015	4879	3,415,300
Japan	2020	8796	8,796,000

Panel data

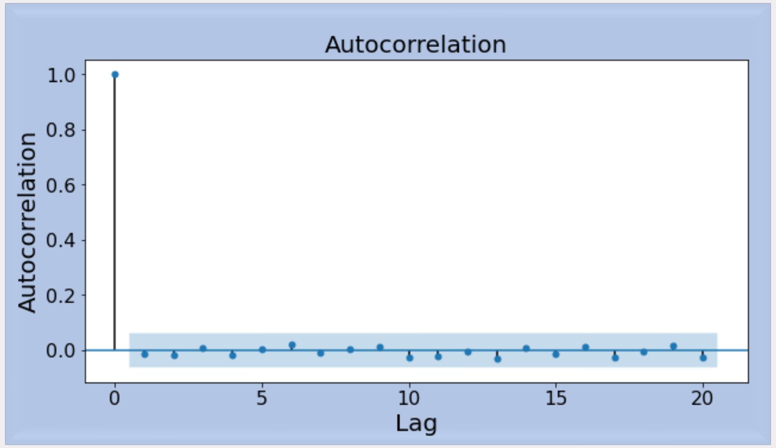
Multiple entities over multiple points in time. A panel data also known as longitudinal data.

Basic Concepts of Time Series Analysis

- **Autocorrelation**
  - Time series is linearly related to a previous version of itself.
  - Univariate time series consists of single observations recorded sequentially over equal time increments.
  - Autocorrelation measures the linear relationship between lagged values of a time series
  - $autocorrelation\ (x,lag) = corr\ (x_t,x_{t+lag})$

Series	Lagged Series
5	
10	5
15	10
20	15
25	20
⋮	⋮

- **Autocorrelation Function (ACF)**
  - Shows the entire autocorrelation function for different lags.
  - Any significant non-zero autocorrelations implies that the series can be forecast from the past.



Questions?

Time Series Models

Time Series

Introduction

Basic Concepts of Time Series Analysis

- **Stationary**
  - A time series $\{X_t\}$ is strictly stationary if :

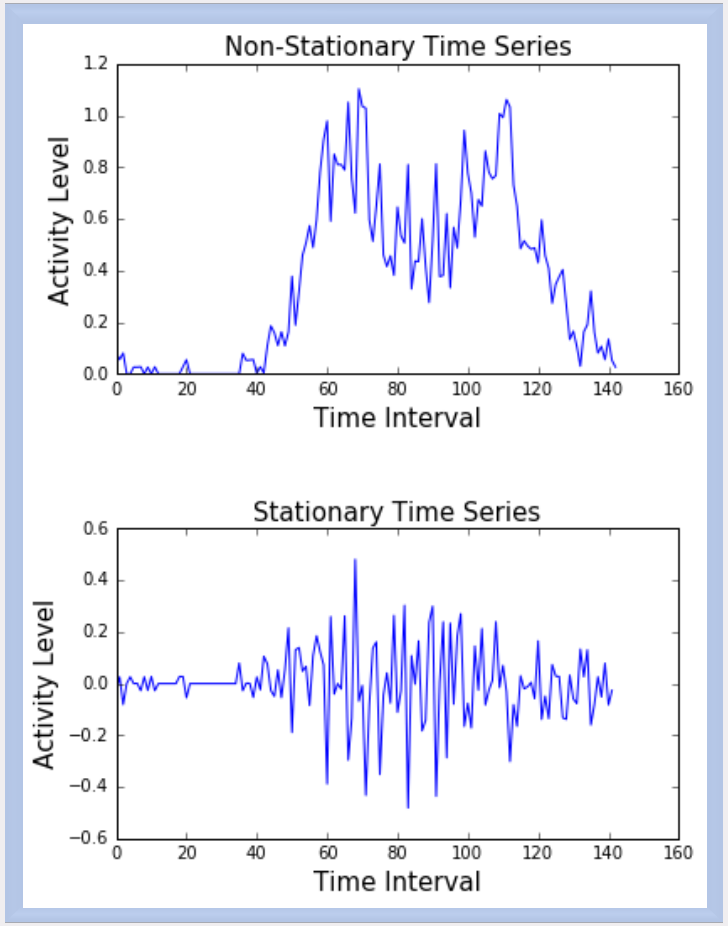
    $\{X_1,.., X_n\}$ and $\{X_{1+k},.., X_{n+k}\}$ possess the same joint **distribution** for any integer n ≥ 1 and any integer k.

- **Weak Stationary:**
  - mean stationary if :

    E(X_t) = μ is a constant.
  - Variance stationary if :

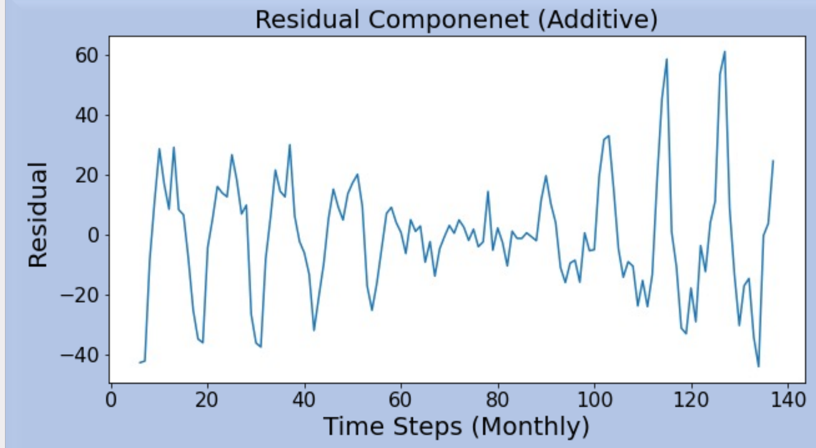
    Var(X_t) = E [$(x_t - \mu_t)^2$)] = σ^2 is a constant.

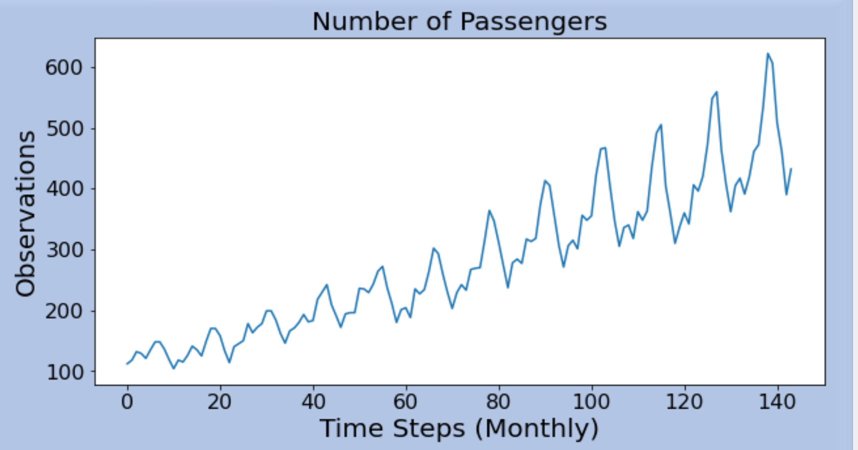
If a time series is both mean stationery and variance stationary, then its time series plot fluctuates around the mean line y = μ.



Basic Concepts of Time Series Analysis

- **Augmented Dicky-Fuller Test**
  - Null Hypothesis : Series in non-stationary






```python
from statsmodels.tsa.stattools import adfuller
ADF_stat = adfuller(additive_decomposition.resid.dropna().values, autolag='AIC')
print(f'\np-value: {ADF_stat[1]}')
print("Non-Stationary") if ADF_stat[1] > 0.05 else print("Stationary")
```

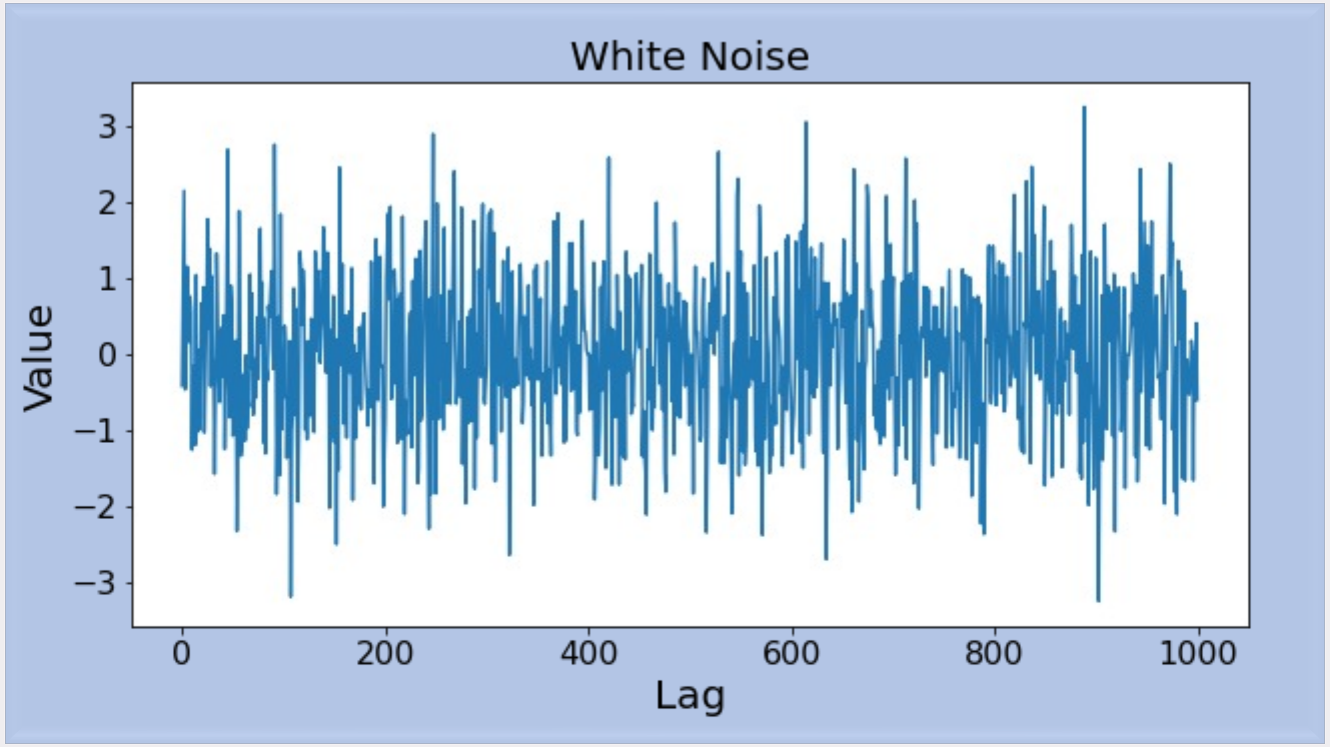
```
p-value: 5.516868902226193e-06
Stationary
```

```python
from statsmodels.tsa.stattools import adfuller
ADF_stat = adfuller(additive_decomposition.observed.dropna().values, autolag='AIC')
print(f'\np-value: {ADF_stat[1]}')
print("Non-Stationary") if ADF_stat[1] > 0.05 else print("Stationary")
```

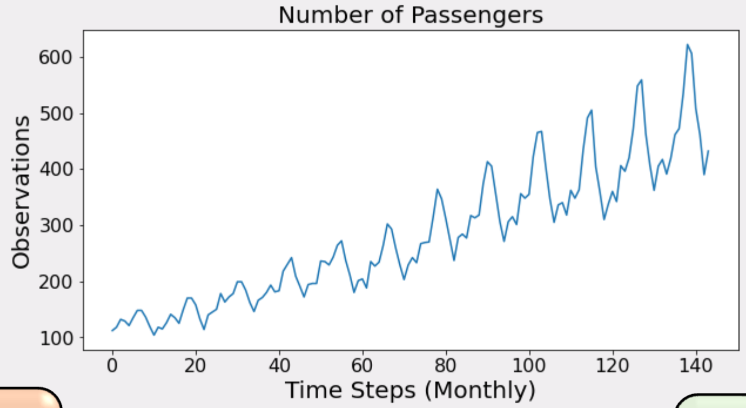
```
p-value: 0.991880243437641
Non-Stationary
```

Basic Concepts of Time Series Analysis

- **White Noise**
  - Constant mean
  - Constant variance
  - Zero autocorrelations at all lags

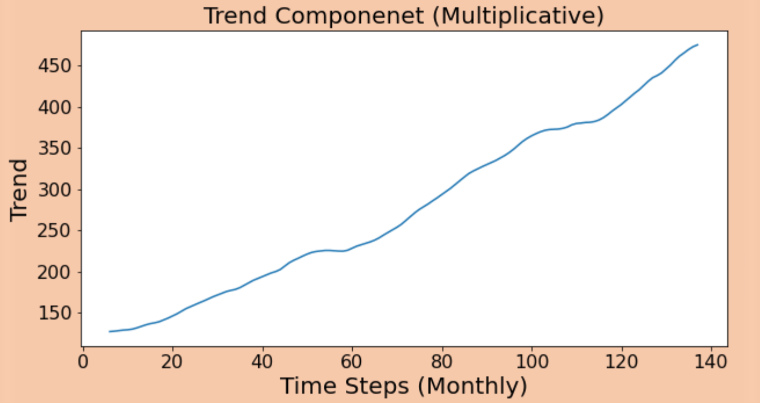


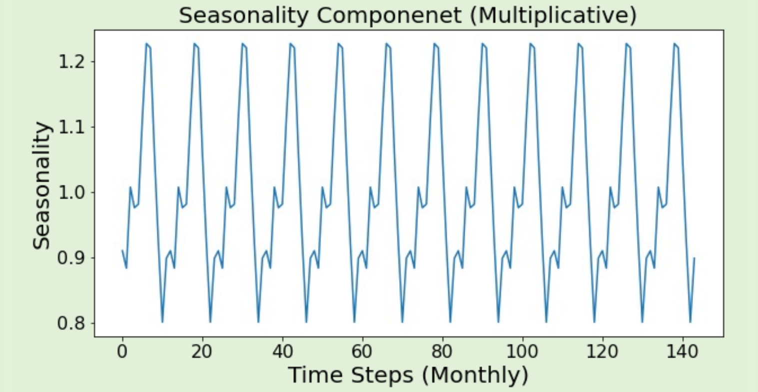
Time Series Patterns



Trend
long-term increase or decrease in the data.

Seasonality
affected by seasonal factors such as the time of the year, or the day of the week.



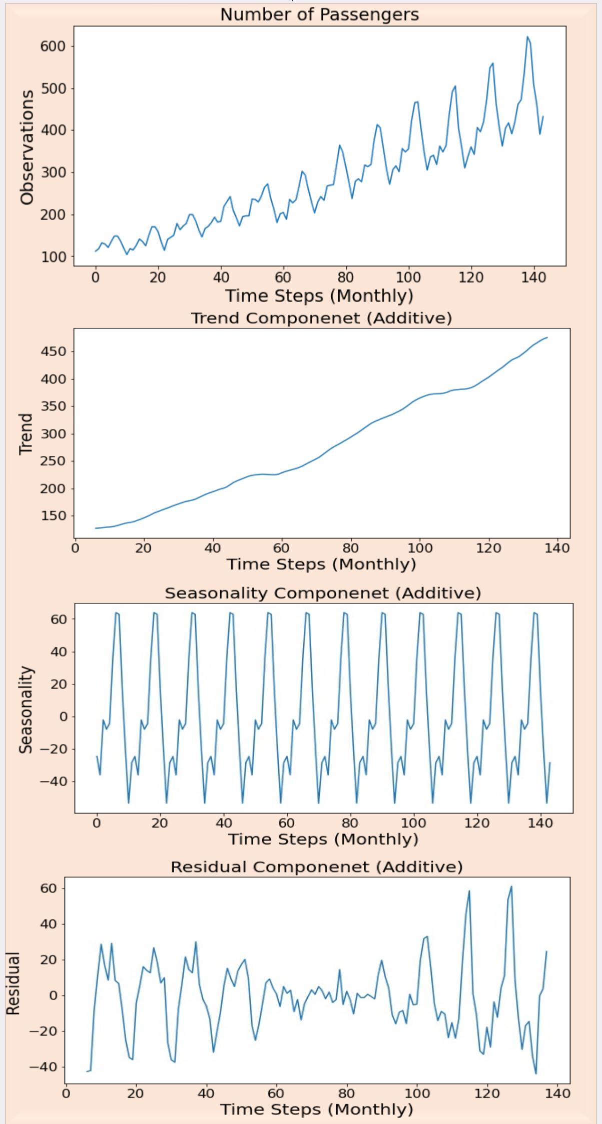


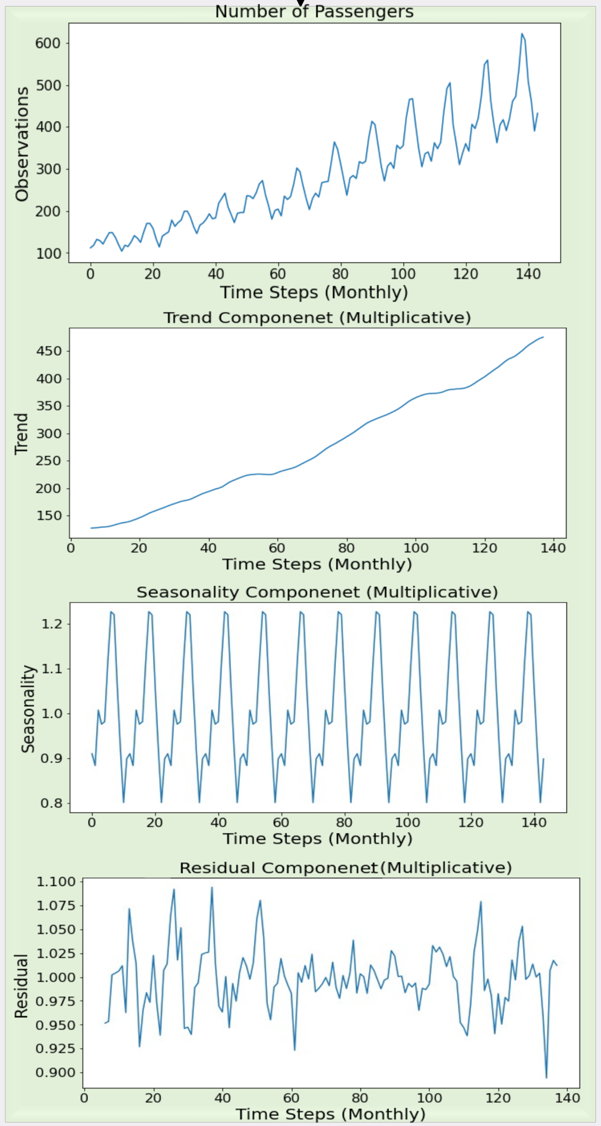
Cycle
Rise and fall pattern in the series that does not happen in fixed calendar-based intervals.

Time Series Decompositions

Three Components
1) Trend, 2) Seasonality, 3) Residuals

Additive Decomposition
Series = Trend + Seasonality + Residuals





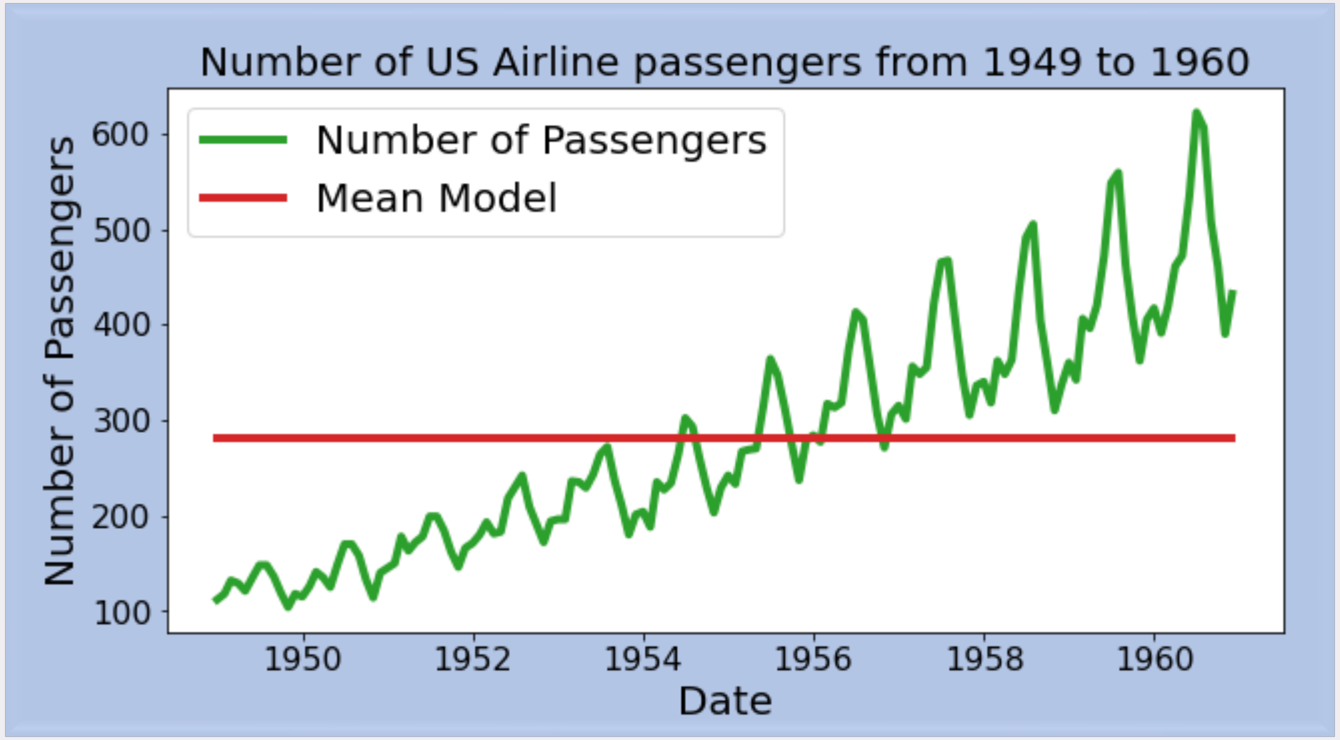
Multiplicative Decomposition
Series = Trend × Seasonality × Residuals

Time Series Models

Let's model

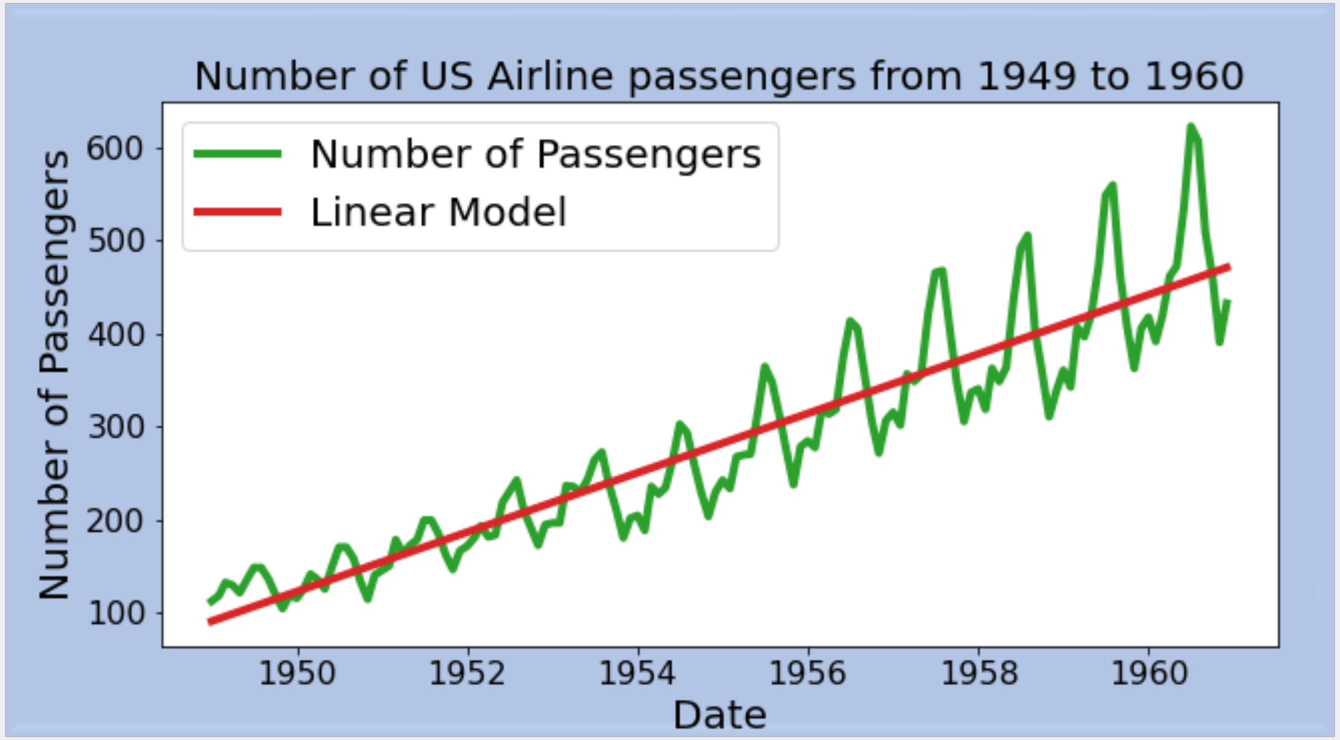
Mean Model Example

- **Mean Model:**
  - $Y(t) = \bar{Y}$



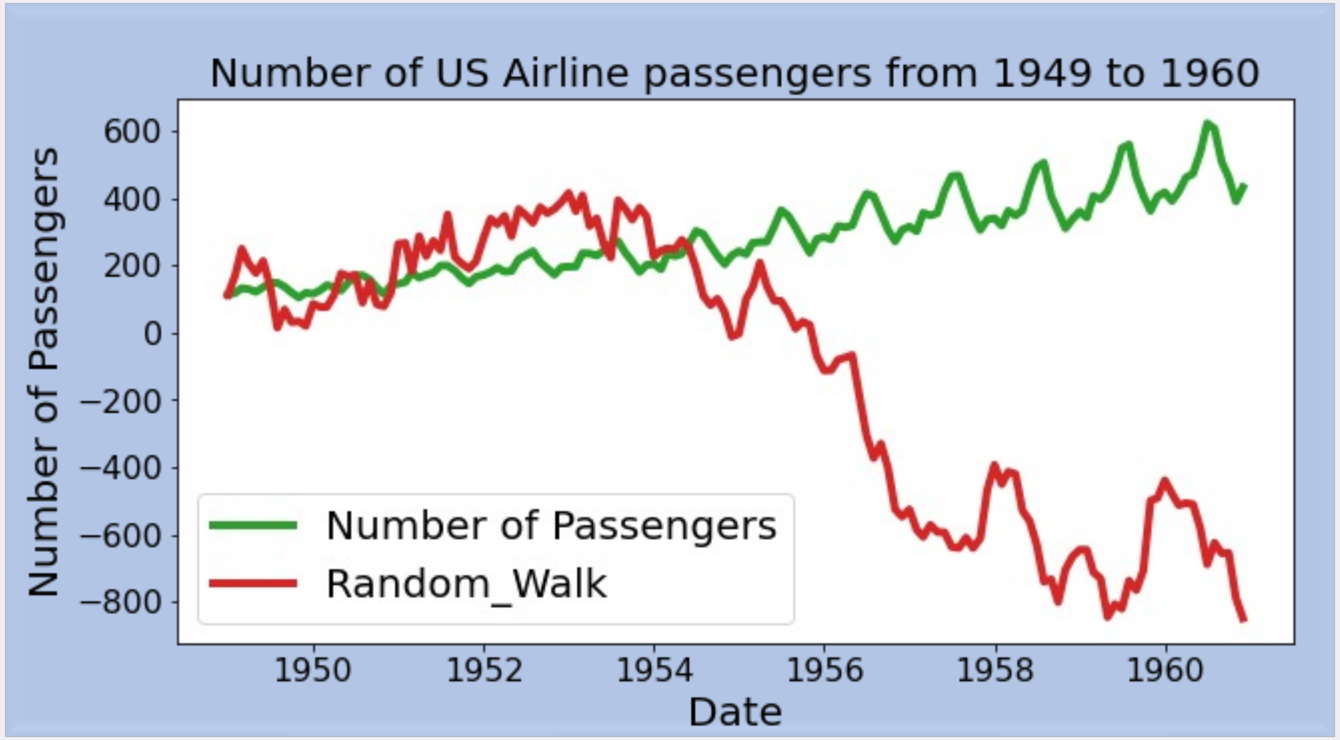
Linear Trend Model Example

- **Linear Trend Model :** $Y(t) = \beta_0 + \beta_1 X(t)$



Random Walk Model Example

- **Random Walk Model :** $Y(t) = Y(t-1) + \epsilon \quad , \quad \epsilon \sim N(0, \sigma^2)$



Questions?

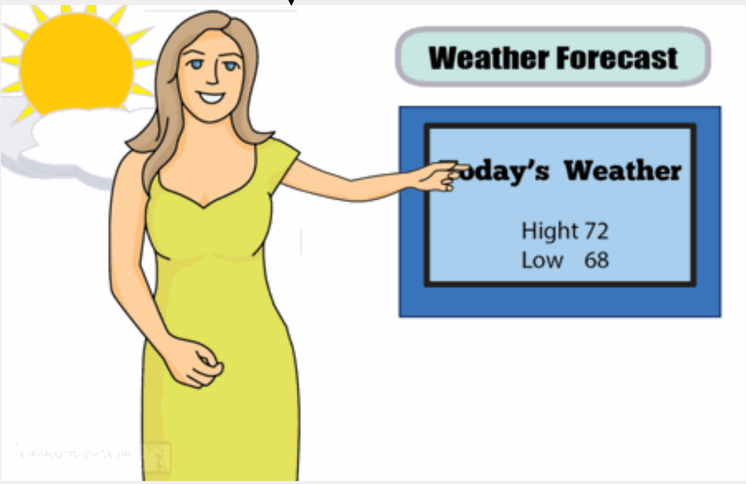
Time Series Models

Time Series

Introduction

Time Series Models

Preparing the data.
(Make it stationary)



Questions?

Time Series Models

Time Series

Introduction

MA Model

Moving Average (MA) models can be used to both forecast and smooth out time series data. The types of models include:

- Simple Moving Average models
- Exponentially Weighted Moving Average modes
- Autoregressive Moving Average models
- ARMA
- ARIMA
- SARIMA

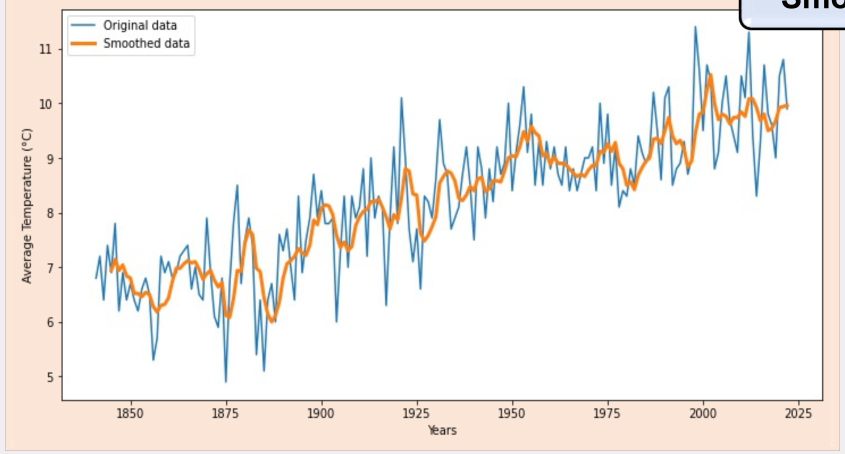
MA Models

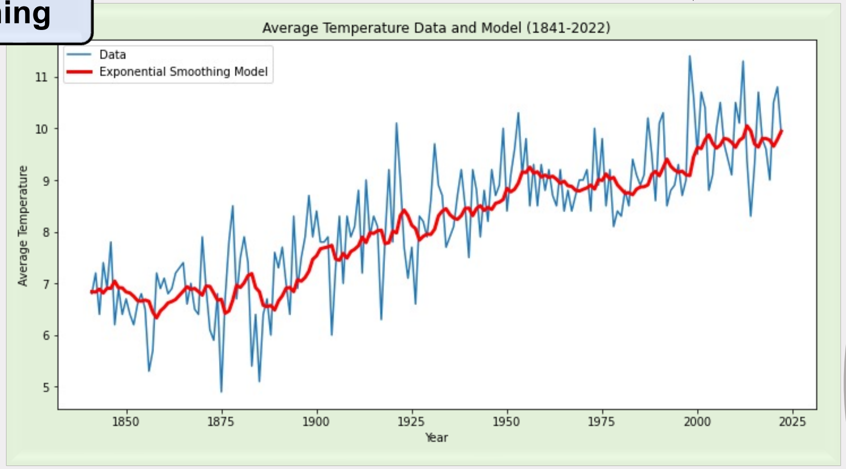
MA models

Simple

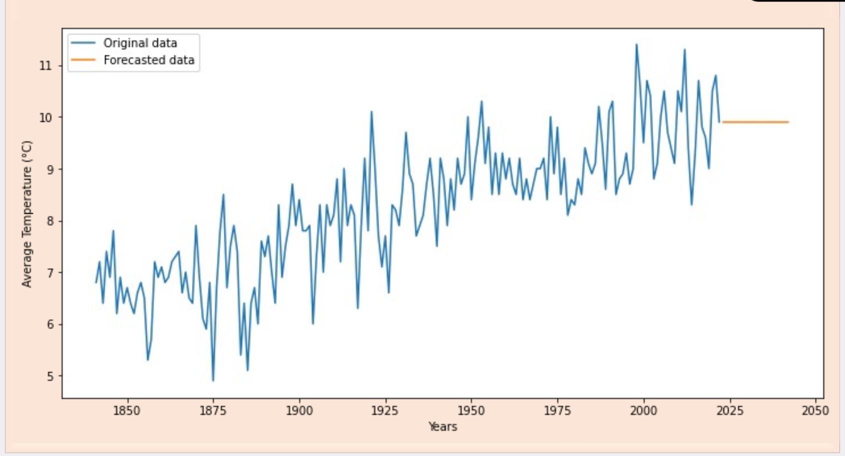
Exponential

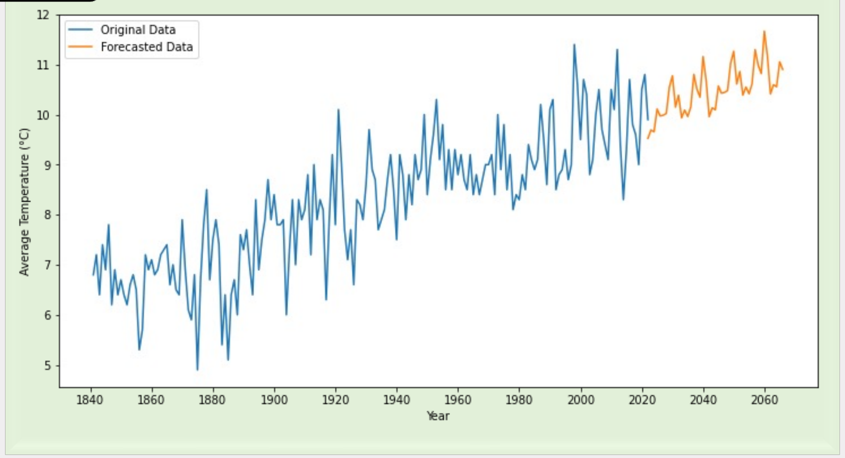
Smoothing





Forecasting





Error & Differencing

- **Error**
  - Example: Let's say you are forecasting sales for the next month. Your last five predictions were
  - In a moving average model, these errors (-10, 5, -10, -2, -3) would be used as predictors to forecast sales for the next month.

Year	1	2	3	4	5
Actual	90	100	115	112	118
Forecasting	100	95	105	110	115
Error	-10	5	-10	-2	-3

- **Differencing**
  - Differencing makes the time series stationary, meaning that its statistical properties (mean, variance, etc.) do not change over time.
  - To calculate the first difference of this time series, we subtract each value from the value in the previous period
  - The first difference time series data has a mean of close to zero, which indicates that it is stationary. If the original time series is not stationary, we can continue differencing until it becomes stationary.

Year	1	2	3	4	5
Actual	100	110	105	95	100
1st Difference	-	10	-5	-10	5

Questions?

Time Series Models

Time Series

Introduction

AR Models

- **Autoregressive (AR) Models**

Factors Affecting Variable — Not Considered, unlike regression

Time series $\{X_t\}$ must be stationary to use this model.

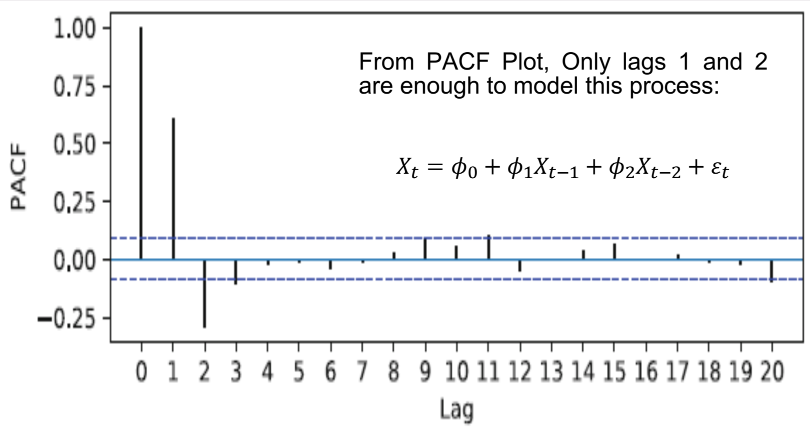
Past Values $\Rightarrow$ Act as Predictors

white noise

$$X_t = \phi_0 + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \ldots + \phi_p X_{p-2} + \varepsilon_t$$

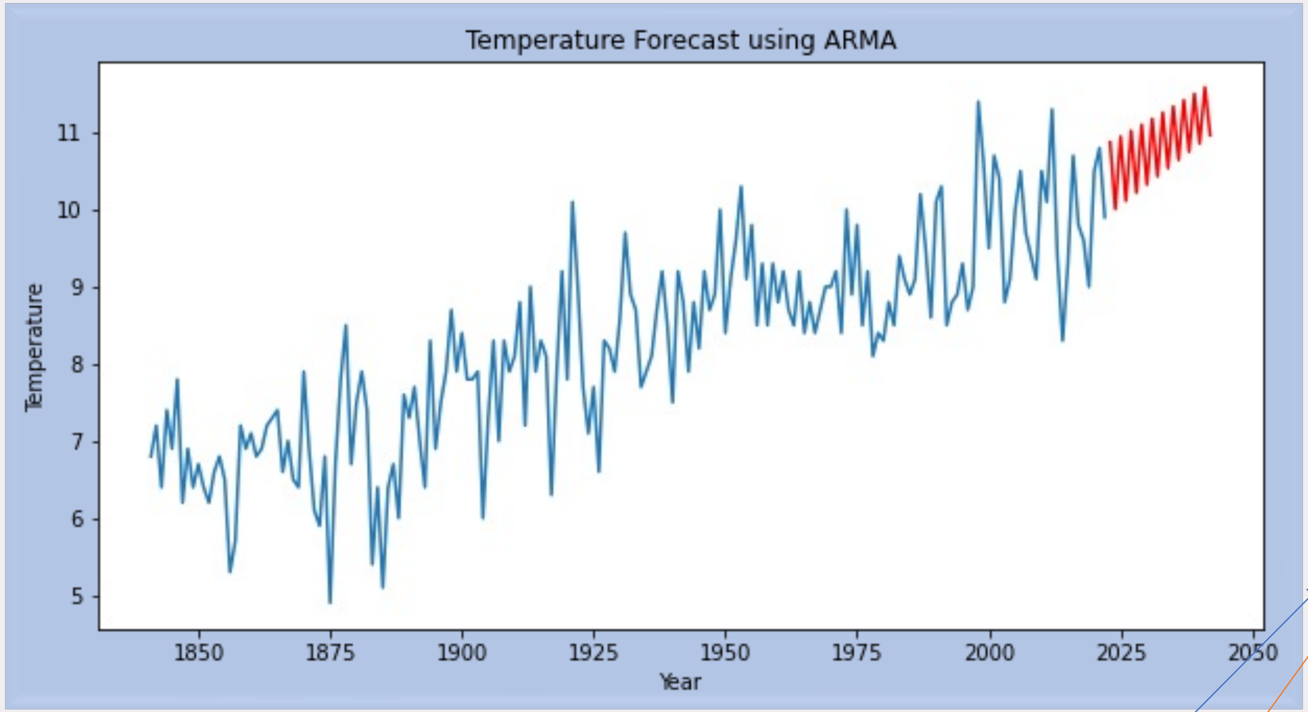
- Order, p of AR model can be judged by a Partial Auto Correlation Function (PACF), to get best possible prediction using the simplest possible model.
- PACF gives us a quantitative measure of the direct effect of a given lag on the quantity being predicted, eliminating the effects of intermediate lag terms.
- Autoregression works well but can be used along with moving average models to better forecast the data.

For an AR(1) model:
when $\varphi_1 = 0, X_t$ is equivalent to white noise;
when $\varphi_1 = 1, \varphi_0 = 0, X_t$ is equivalent to a random walk;
when $\varphi_1 = 1, \varphi_0 \neq 0, X_t$ is equivalent to a random walk with drift;
when $\varphi_1 < 0, X_t$ tends to oscillate around the mean.



ARMA Models

- **Autoregressive Moving Average (ARMA)**
  - ARMA is the combination of the AR and MA models. It predicts the future values based on both the previous values and errors. Thus, ARMA has better performance than AR and MA models alone.

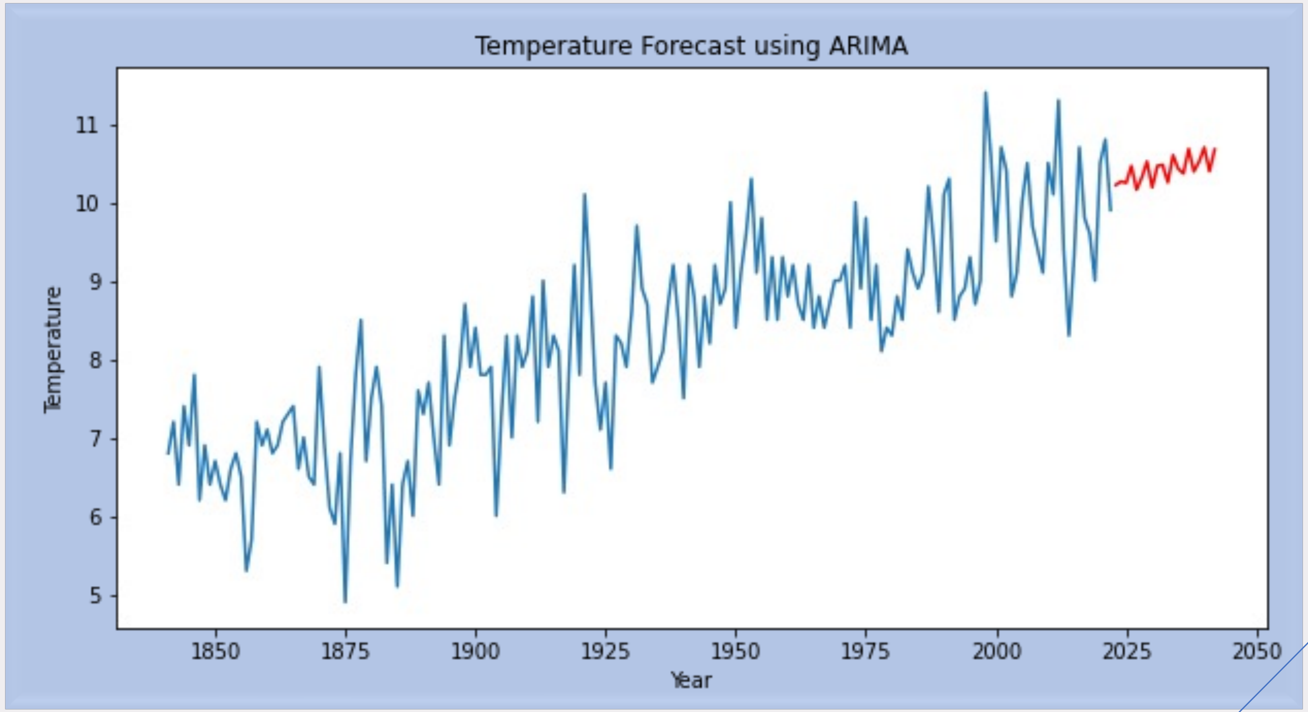


AR terms
MA terms

```
model  = ARMA(df['Temperature'], order=(3, 1))
result = model.fit()
```

ARIMA Models

- **Autoregressive Integrated Moving Average (ARIMA)**
  - Similar to ARMA models. The difference between ARMA and ARIMA is the integration part. The integrated I stands for the number of times differencing is needed to make the times series stationary.

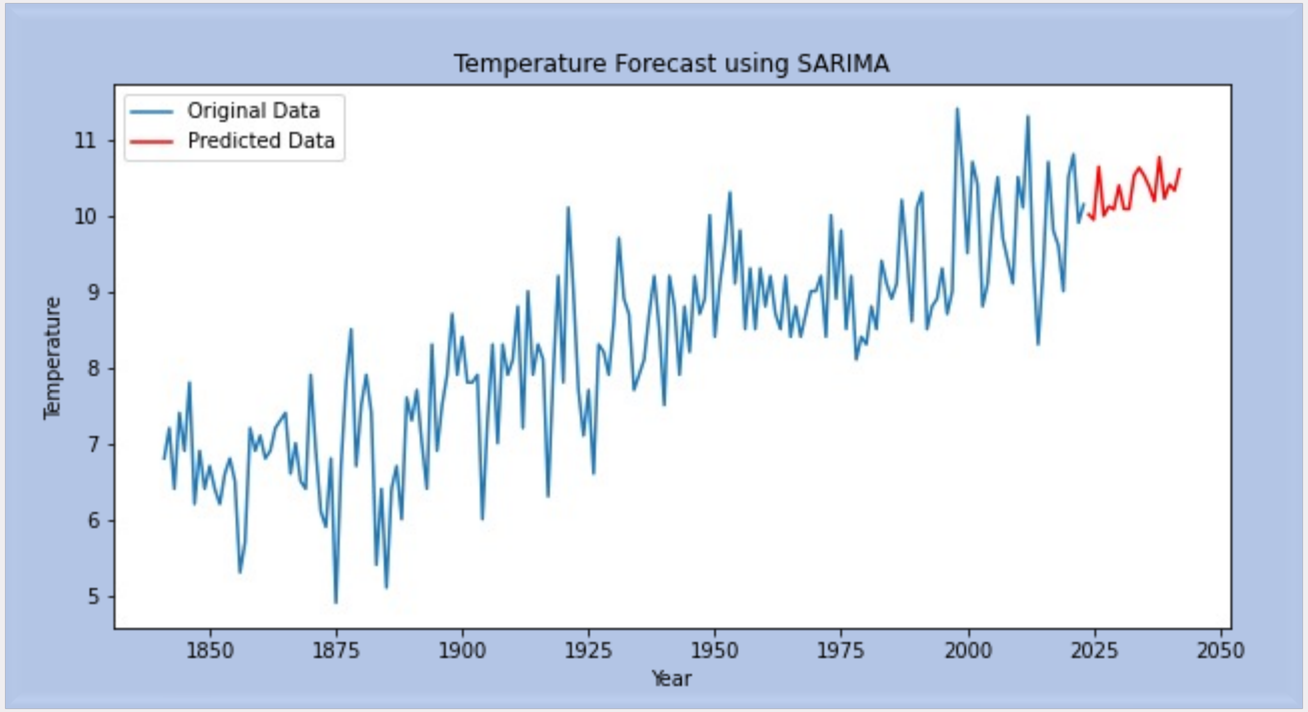


AR terms
Order of differencing
MA terms

```
model = ARIMA(df['Temperature'], order=(5, 2, 3))
result = model.fit()
```

SARIMA Models

- **Seasonal Autoregressive Integrated Moving Average (SARIMA)**
  - SARIMA is an extension of ARIMA that explicitly supports univariate time series data with a seasonal component.

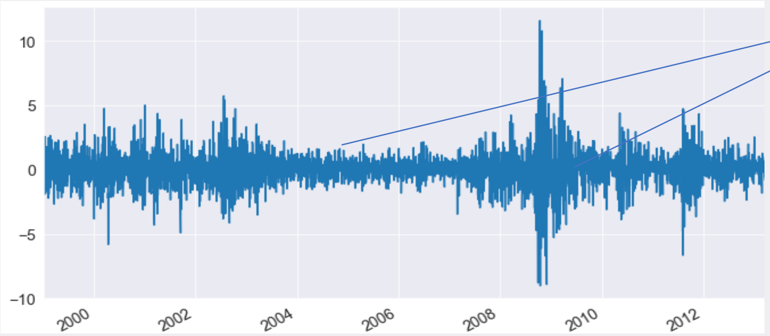


#seasons (e.g. 12 for monthly data, 4 for quarterly data)

```
model = SARIMAX(df['Temperature'], order=(2, 1, 2), seasonal_order=(1, 1, 1, 12))
```

ARCH Model

- **Auto Regressive Conditional Heteroskedastic (ARCH)**
  - This method models the volatility (variance) of the current residual as a function of the variance of past residuals.
  - Fit a best possible time series model to a dataset, and plot the residuals i.e. the differences between the predictions and the actual historical values.
  - Time series $\{X_t\}$ must be stationary other than conditional change in variance.



Volatility in the model tends to have a pattern

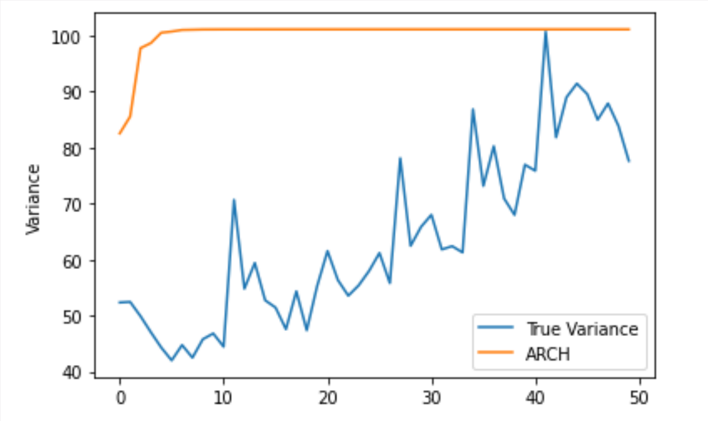
What if we model this volatility and modify the error term in our model?

- Obtain squares of error terms and regress them:

$\epsilon_t = \sigma_t z_t$

$$\sigma^2{}_t = \alpha_0 + \alpha_1 \epsilon^2{}_{t-1} + \ldots + \alpha_q \epsilon^2{}_{t-q} = \alpha_0 + \sum_{i=1}^{q} \alpha_i \epsilon^2{}_{i-i}$$

Where $\alpha_0 > 0, \alpha_i \geq 0, i > 0$



Variance Predicted by ARCH(2) model on a dataset.

GARCH Model

- **Generalized Auto Regressive Conditional Heteroskedastic (GARCH)**
  - This model is an extension of ARCH model in which moving average model is assumed for error variance
  - Thus, GARCH incorporates the ARMA model into ARCH, which only consists of AR.
  - A GARCH $(\boldsymbol{p}, \boldsymbol{q})$ model can be described as:

ARCH Terms

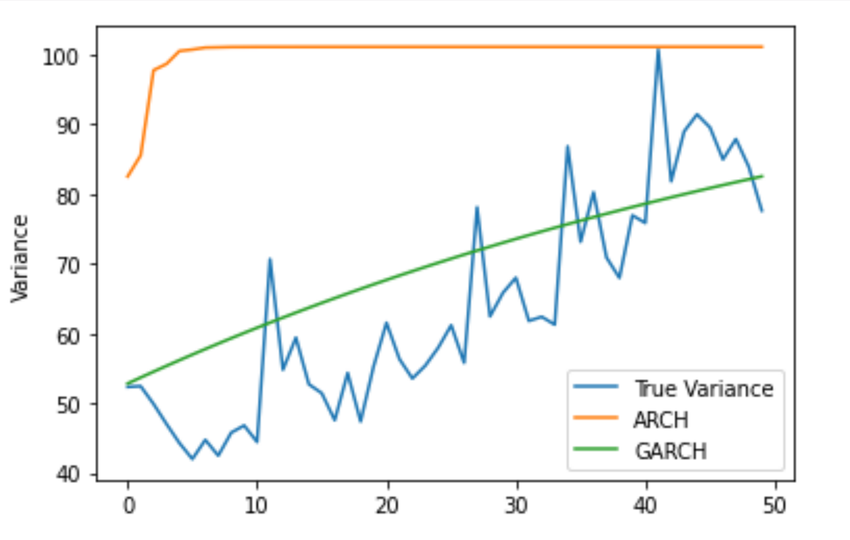
GARCH Terms

$$\epsilon_t = \sigma_t z_t$$

$$\sigma^2{}_t = \omega + \boxed{\alpha_1 \epsilon^2{}_{t-1} + \ \dots + \alpha_q \epsilon^2{}_{q-1}} + \boxed{\beta_1 \sigma^2{}_{t-1} + \ \dots + \beta_p \sigma^2{}_{p-1}}$$

$$\sigma^2{}_t = \omega + \sum_{i=1}^{q} \alpha_i \epsilon^2{}_{i-1} + \sum_{i=1}^{p} \beta_i \sigma^2{}_{i-1}$$

q is number of lag squared residual errors
p is the order of auto-regression for error variance



Variance Predicted by ARCH(2) model vs GARCH(1,2) model on a dataset

References:

[1] "Time Series Analysis in Python – A Comprehensive Guide with Examples," Machine Learning Plus, [Online]. Available: https://www.machinelearningplus.com/time-series/time-series-analysis-python/.

[2] "ARIMA Model – Complete Guide to Time Series Forecasting in Python," Machine Learning Plus, [Online]. Available: https://www.machinelearningplus.com/time-series/arima-model-time-series-forecasting-python/.

Thank You For Your Attention!