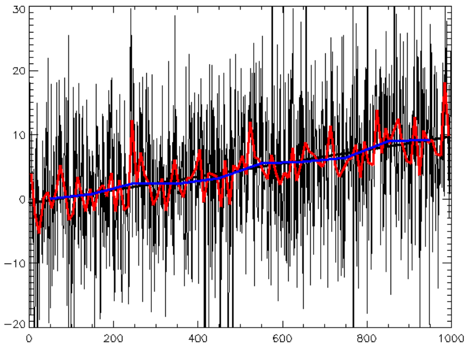
Week 3
Time Series Analysis

Content

- Introduction to Time Series Data
- Time Series Data Analysis
- AR and MA models
- ARIMA and SARIMA Models
- Time series forecasting with python

Introduction to Time Series Data

- A time series is a collection of data y_t (t=1,2,…,T), with the interval between y_t and yt+1 being fixed and constant.
- We can think of time series as being generated by a stochastic process, or the data generating process (DGP).
- A time series (sample) is a particular realization of the DGP (population).
- Time series analysis is the estimation of difference equations containing stochastic (error) terms (Enders 2010).
- Examples:
  - Price of a stock over successive days
  - Sizes of video frames
  - Sizes of packets over network
  - Sizes of queries to a database system
  - Number of active virtual machines in a cloud



Introduction to Time Series Data

- Types of time series data
  - Single time series
    - U.S. presidential approval, monthly (1978:1-2004:7)
    - Number of militarized disputes in the world annually (1816-2001)
    - Changes in the monthly Dow Jones stock market value (1978:1-2001:1)
  - Pooled time series
    - Dyad-year analyses of interstate conflict
    - State-year analyses of welfare policies
    - Country-year analyses of economic growth
- Goal: Develop models of such series for resource allocation and improving user experience.

Time Series Data Analysis

- Models Autoregressive Models
  - Predict the variable as a linear regression of the immediate past value:

$$\hat{x}_t = a_0 + a_1 x_{t-1}$$

  - Here, $\hat{x}_t$ is the best estimate of xt given the past history

$$\{x_0, x_1, \ldots, x_{t-1}\}$$

  - Even though we know the complete past history, we assume that x_t can be predicted based on just xt-1.
  - Auto-Regressive = Regression on Self
  - Error: $e_t = x_t - \hat{x}_t = x_t - a_0 - a_1 x_{t-1}$
  - Model: $x_t = a_0 + a_1 x_{t-1} + e_t$
  - Best a_0 and $a_1 \Rightarrow$ minimize the sum of square of errors

Time Series Data Analysis

- Example

  - The number of disk access for 50 database queries were measured to be: 73, 67, 83, 53, 78, 88, 57, 1, 29, 14, 80, 77, 19, 14, 41, 55, 74, 98, 84, 88, 78, 15, 66, 99, 80, 75, 124, 103, 57, 49, 70, 112, 107, 123, 79, 92, 89, 116, 71, 68, 59, 84, 39, 33, 71, 83, 77, 37, 27, 30.

  - For this data:

$$\sum_{t=2}^{50} x_t = 3313 \quad \sum_{t=2}^{50} x_{t-1} = 3356$$

$$\sum_{t=2}^{50} x_t x_{t-1} = 248147 \quad \sum_{t=2}^{50} x_{t-1}^2 = 272102 \quad n = 49$$

$$\begin{aligned} a_0 &= \frac{\sum x_t \sum x_{t-1}^2 - \sum x_{t-1} \sum x_t x_{t-1}}{n \sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2} \\ &= \frac{3313 \times 272102 - 3356 \times 248147}{49 \times 272102 - 3356^2} = 33.181 \end{aligned}$$

Time Series Data Analysis

- Example (Cont)

  - The number of disk access for 50 database queries were measured to be: 73, 67, 83, 53, 78, 88, 57, 1, 29, 14, 80, 77, 19, 14, 41, 55, 74, 98, 84, 88, 78, 15, 66, 99, 80, 75, 124, 103, 57, 49, 70, 112, 107, 123, 79, 92, 89, 116, 71, 68, 59, 84, 39, 33, 71, 83, 77, 37, 27, 30.

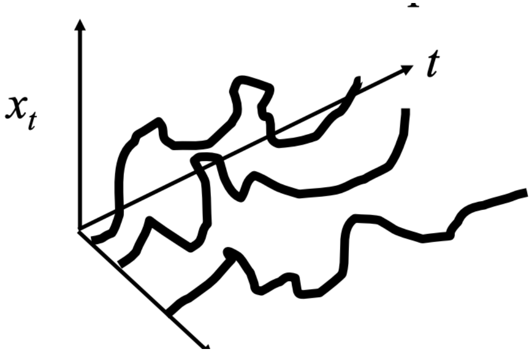
  - For this data:

$$
\begin{aligned}
a_1 &= \frac{n \sum x_t x_{t-1} - \sum x_t \sum x_{t-1}}{n \sum x_{t-1}^2 - \left(\sum x_{t-1}\right)^2} \\
&= \frac{49 \times 248147 - 3313 \times 3356}{49 \times 272102 - 3356^2} = 0.503
\end{aligned}
$$

  - SSE = 32995.57

Time Series Data Analysis

- **Stationary Process Stationary Process**
  - Each realization of a random process will be different:



  - x is function of the realization i (space) and time t: x(i, t)
  - We can study the distribution of xt in space.
  - Each xt has a distribution, e.g., Normal $f(x_t) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x_t-\mu)^2}{2\sigma^2}}$
  - If this same distribution (normal) with the same parameters μ, σ applies to xt+1, xt+2, …, we say xt is stationary.

Time Series Data Analysis

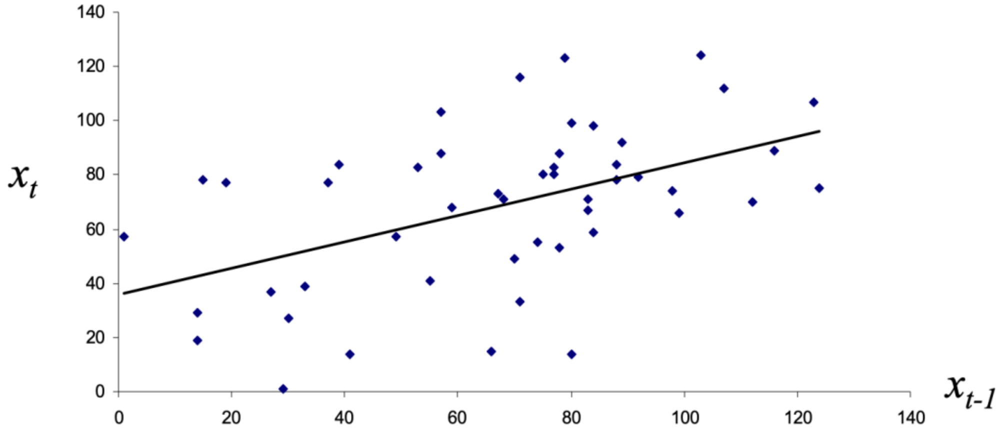
- **Stationary Process Stationary Process (Cont)**
  - Stationary = Standing in time
  - Distribution does not change with time.
  - Similarly, the joint distribution of xt and xt-k depends only on k not on t.
  - The joint distribution of xt, xt-1, …, xt-k depends only on k not on t.

Time Series Data Analysis

- **Assumptions**
  - Linear relationship between successive values
  - Normal Independent identically distributed errors:
    - Normal errors
    - Independent errors
  - Additive errors
  - x_t is a Stationary process

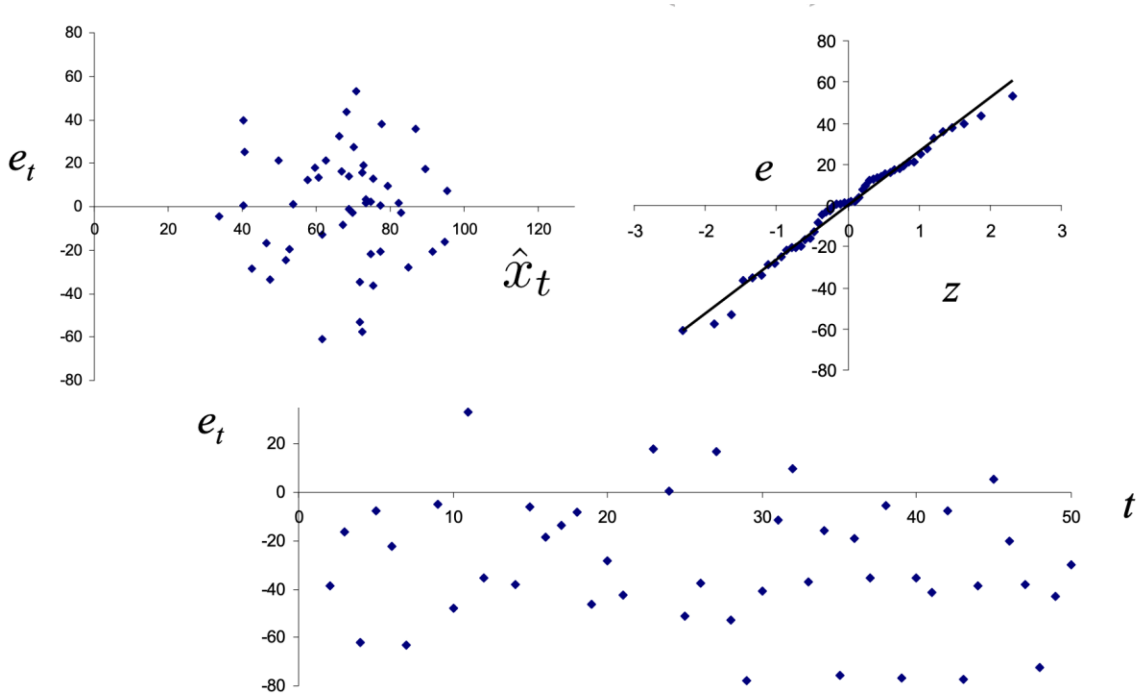
Time Series Data Analysis

- **Visual Tests**
    - x_t vs. xt-1 for linearity
    - Errors e_t vs. predicted values for additivity
    - Q-Q Plot of errors for Normality
    - Errors e_t vs. t for Stationarity
    - Correlations for Independence



Time Series Data Analysis

- **Visual Tests**



Autoregression (AR) Model

- **AR(p) Model**
  - x_t is a function of the last p values:
    $$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + \cdots + a_p x_{t-p} + e_t$$
  - AR(2): $x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + e_t$
  - AR(3): $x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + a_3 x_{t-3} + e_t$

Autoregression (AR) Model

- **Backward Shift Operator**
  - Similarly, $\mathrm{B}(x_t) = x_{t-1}$

    $\mathrm{B}(\mathrm{B}(x_t)) = \mathrm{B}(x_{t-1}) = x_{t-2}$

    $\mathrm{B}^2 x_t = x_{t-2}$
  - Or $\mathrm{B}^3 x_t = x_{t-3}$

    $\mathrm{B}^k x_t = x_{t-k}$

  - Using this notation, AR(p) model is:

$$
\begin{aligned}
x_t - a_1 x_{t-1} - a_2 x_{t-2} - \cdots - a_p x_{t-p} &= a_0 + e_t \\
x_t - a_1 B x_t - a_2 B^2 x_t - \cdots - a_p B^p x_t &= a_0 + e_t \\
(1 - a_1 B - a_2 B^2 - \cdots - a_p B^p) x_t &= a_0 + e_t \\
\phi_p(B) x_t &= a_0 + e_t
\end{aligned}
$$

  - Here, Φ p is a polynomial of degree p.

Autoregression (AR) Model (Visulize, parameters)

- **AR(p) Parameter Estimation**

$$x_t = a_0 + a_1 x_{t-1} + a_2 x_{t-2} + e_t$$

  - The coefficients a_i's can be estimated by minimizing SSE using Multiple Linear Regression.

$$\text{SSE} = \sum e_t^2 = \sum_{t=3}^{n} (x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2})^2$$

  - Optimal a_0, a_1, and $a_2 \Rightarrow$ Minimize SSE
    $\Rightarrow$Set the first differential to zero:

$$\begin{aligned}
\frac{d}{da_0}\text{SSE} &= \sum_{t=3}^{n} -2(x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2}) = 0 \\
\frac{d}{da_1}\text{SSE} &= \sum_{t=3}^{n} -2x_{t-1}(x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2}) = 0 \\
\frac{d}{da_2}\text{SSE} &= \sum_{t=3}^{n} -2x_{t-2}(x_t - a_0 - a_1 x_{t-1} - a_2 x_{t-2}) = 0
\end{aligned}$$

Autoregression (AR) Model

- **AR(p) Parameter Estimation**
  - The equations can be written as:

$$\begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix} \begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

  - Note: All sums are for t=3 to n. n-2 terms.

    Multiplying by the inverse of the first matrix, we get:

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

Autoregression (AR) Model

- **Example**
  - Consider the data of previous example and fit an AR(2) model:
    - The number of disk access for 50 database queries were measured to be: 73, 67, 83, 53, 78, 88, 57, 1, 29, 14, 80, 77, 19, 14, 41, 55, 74, 98, 84, 88, 78, 15, 66, 99, 80, 75, 124, 103, 57, 49, 70, 112, 107, 123, 79, 92, 89, 116, 71, 68, 59, 84, 39, 33, 71, 83, 77, 37, 27, 30.

$$\begin{bmatrix} a_0 \\ a_1 \\ a_2 \end{bmatrix} = \begin{bmatrix} n-2 & \sum x_{t-1} & \sum x_{t-2} \\ \sum x_{t-1} & \sum x_{t-1}^2 & \sum x_{t-1}x_{t-2} \\ \sum x_{t-2} & \sum x_{t-1}x_{t-2} & \sum x_{t-2}^2 \end{bmatrix}^{-1} \begin{bmatrix} \sum x_t \\ \sum x_t x_{t-1} \\ \sum x_t x_{t-2} \end{bmatrix}$$

$$= \begin{bmatrix} 48 & 3283 & 3329 \\ 3283 & 266773 & 247337 \\ 3329 & 247337 & 271373 \end{bmatrix}^{-1} \begin{bmatrix} 3246 \\ 243256 \\ 229360 \end{bmatrix} = \begin{bmatrix} 39.979 \\ 0.587 \\ -0.180 \end{bmatrix}$$

  - SSE= 31969.99
  - (3% lower than 32995.57 for AR(1) model)

Autoregression (AR) Model

- **Assumptions and Tests for AR(p)**
  - Assumptions:
    - Linear relationship between xt and {xt-1, ..., xt-p}
    - Normal Independent identically distributed errors:
      - Normal errors
      - Independent errors
    - Additive errors
    - xt is stationary
  - Visual Tests: Similar to AR(1).

Autoregression (AR) Model

- **Autocorrelation**

  - Covariance of xt and xt-k = Auto-covariance at lag k

$$\text{Autocovariance of } x_t \text{ at lag } k = \text{Cov}[x_t, x_{t-k}] = E[(x_t - \mu)(x_{t-k} - \mu)]$$

  - For a stationary series, the statistical characteristics do not depend upon time t.
  - Therefore, the autocovariance depends only on lag k and not on time t.
  - Similarly,

$$\begin{aligned} \text{Autocorrelation of } x_t \text{ at lag } k \;\; r_k &= \frac{\text{Autocovariance of } x_t \text{ at lag } k}{\text{Variance of } x_t} \\ &= \frac{\text{Cov}[x_t, x_{t-k}]}{\text{Var}[x_t]} \\ &= \frac{E[(x_t - \mu)(x_{t-k} - \mu)]}{E[(x_t - \mu)^2]} \end{aligned}$$

Autoregression (AR) Model

- **Autocorrelation**

Autocorrelation is dimensionless and is easier to interpret than autocovariance.

It can be shown that autocorrelations are normally distributed with mean: $E[r_k] \approx \frac{-1}{n}$

and variance: $\mathrm{Var}[r_k] \approx \frac{-1}{n}$

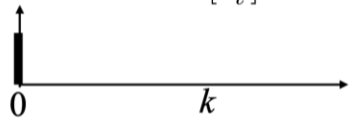
Therefore, their 95% confidence interval is $-1/n \mp 1.96/\sqrt{n}$
This is generally approximated as $\mp 2/\sqrt{n}$

Autoregression (AR) Model (What happen after normalized)

- **White noise**
  - Errors et are normal independent and identically distributed (IID) with zero mean and variance σ2
  - Such IID sequences are called "white noise" sequences.
  - Properties:

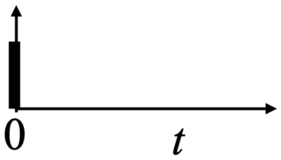
$$
\begin{aligned}
E[e_t] &= 0 \qquad \forall t \\
\mathrm{Var}[e_t] &= E[e_t^2] = \sigma^2 \qquad \forall t
\end{aligned}
$$

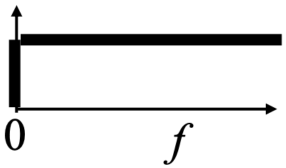
$$
\begin{aligned}
\mathrm{Cov}[e_t, e_{t-k}] &= E[e_t e_{t-k}] = \begin{cases} \sigma^2 & k = 0 \\ 0 & k \neq 0 \end{cases} \\
\mathrm{Cor}[e_t, e_{t-k}] &= \frac{E[e_t e_{t-k}]}{E[e_t^2]} = \begin{cases} 1 & k = 0 \\ 0 & k \neq 0 \end{cases}
\end{aligned}
$$



Autoregression (AR) Model

- **White noise**
  - The autocorrelation function of a white noise sequence is a spike (δ function) at k=0.
  - The Laplace transform of a δ function is a constant. So in frequency domain white noise has a flat frequency spectrum.





  - It was incorrectly assumed that white light has no color and, therefore, has a flat frequency spectrum and so random noise with flat frequency spectrum was called white noise.
  - Ref: http://en.wikipedia.org/wiki/Colors_of_noise

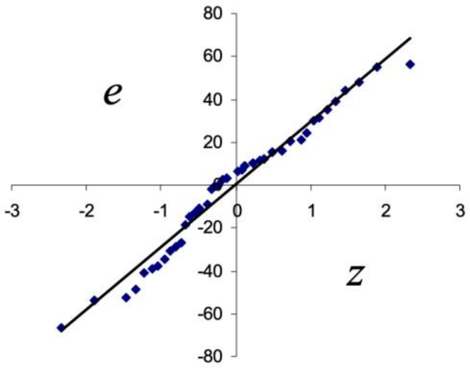
Autoregression (AR) Model

- **Example**
  - Consider the data of previous example. The AR(0) model is:
    - The number of disk access for 50 database queries were measured to be: 73, 67, 83, 53, 78, 88, 57, 1, 29, 14, 80, 77, 19, 14, 41, 55, 74, 98, 84, 88, 78, 15, 66, 99, 80, 75, 124, 103, 57, 49, 70, 112, 107, 123, 79, 92, 89, 116, 71, 68, 59, 84, 39, 33, 71, 83, 77, 37, 27, 30.

$$x_t = a_0 + e_t$$

$$\sum x_t = na_0 + \sum e_t$$

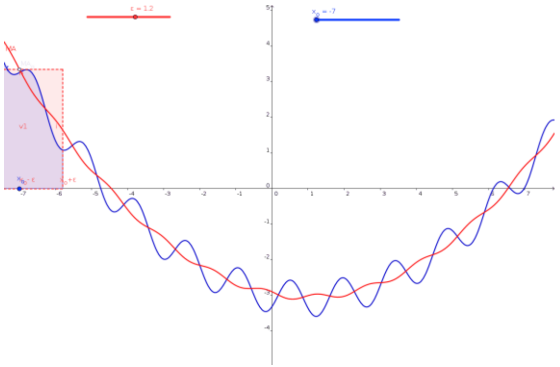
$$a_0 = \frac{1}{n}\sum x_t = 67.72$$



  - SSE = 43702.08

Moving Average (MA) Models

- **MA Models**



  - Moving Average of order 1: MA(1)

$$x_t - a_0 = e_t + b_1 e_{t-1}$$

  - Moving Average of order 2: MA(2)

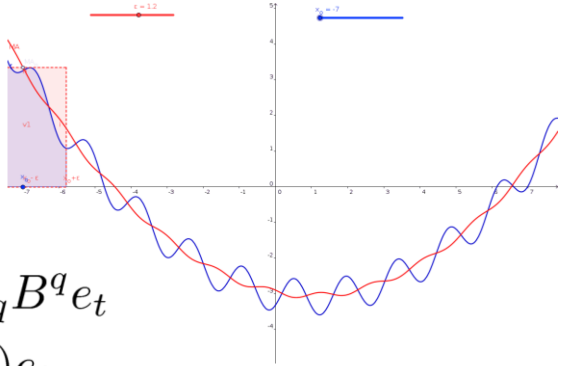
$$x_t - a_0 = e_t + b_1 e_{t-1} + b_2 e_{t-2}$$

  - Moving Average of order q: MA(q)

$$x_t - a_0 = e_t + b_1 e_{t-1} + b_2 e_{t-2} + \cdots + b_q e_{t-q}$$

  - Moving Average of order 0: MA(0) (Note: This is also AR(0)) xt-a0 is a white noise. a0 is the mean of the time series. $x_t - a_0 = e_t$

Moving Average (MA) Models

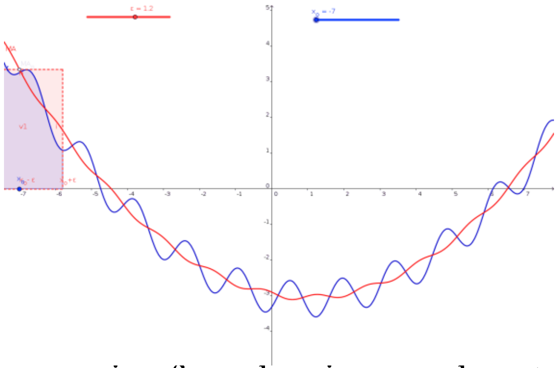


- **MA Models**
  - Using the backward shift operator B, MA(q):

$$\begin{aligned} x_t - a_0 &= e_t + b_1 B e_t + b_2 B^2 e_t + \cdots + b_q B^q e_t \\ &= (1 + b_1 B + b_2 B^2 + \cdots + b_q B^q) e_t \\ &= \psi_q(B) e_t \end{aligned}$$

  - Here, ϕ q is a polynomial of order q.

Moving Average (MA) Models



- **Determining MA Parameters**
  - Consider MA(1):

$$x_t - a_0 = e_t + b_1 e_{t-1}$$

  - The parameters a0 and b1 cannot be estimated using standard regression formulas since we do not know errors. The errors depend on the parameters.
  - So the only way to find optimal a0 and b1 is by iteration.
  - Start with some suitable values and change a0 and b1 until SSE is minimized and average of errors is zero.

Moving Average (MA) Models

- **Example**
  - Consider the data of previous example. The number of disk access for 50 database queries were measured to be: 73, 67, 83, 53, 78, 88, 57, 1, 29, 14, 80, 77, 19, 14, 41, 55, 74, 98, 84, 88, 78, 15, 66, 99, 80, 75, 124, 103, 57, 49, 70, 112, 107, 123, 79, 92, 89, 116, 71, 68, 59, 84, 39, 33, 71, 83, 77, 37, 27, 30.
  - For this data: $\bar{x} = \frac{1}{50}\sum_{t=1}^{50} x_t = 67.72$
  - We start with $a_0 = 67.72$, $b_1 = 0.4$,
  - Assuming e0=0, $\bar{e} = \frac{1}{50}\sum_{t=1}^{50} e_t = -0.152$ compute all the errors and SSE.
  - We then adjust a0 and b1 until SSE is minimized and mean error is close to zero.

Moving Average (MA) Models

- **Example**
  - The steps are: Starting with $a_0 = \bar{x}$ and b1=0.4, 0.5, 0.6

a_0	b_1	$\bar{e}$	SSE	Decision
67.72	0.4	-0.15	33542.65	
67.72	0.5	-0.17	33274.55	
67.72	0.6	-0.18	34616.85	0.5 is the lowest. Try 0.45 and 0.55
67.72	0.55	-0.18	33686.88	
67.72	0.45	-0.16	33253.62	Lowest. Try 0.475 and 0.425
67.72	0.475	-0.17	33221.06	Lowest. Try 0.4875 and 0.4625
67.72	0.4875	-0.17	33236.41	
67.72	0.4625	-0.16	33227.19	b_1=0.475 is lowest. Adjust a_0
67.35	0.475	0.08	33223.45	Close to minimum SSE and zero mean.

Moving Average (MA) Models

- **Autocorrelations for MA**
  - For this series, the mean is:

$$\mu = E[x_t] = a_0 + E[e_t] + b_1 E[e_{t-1}] = a_0$$

  - The variance is:

$$\begin{aligned}
\mathrm{Var}[x_t] &= E[(x_t - \mu)^2] = E[(e_t + b_1 e_{t-1})^2] \\
&= E[e_t^2 + 2b_1 e_t e_{t-1} + b_1^2 e_{t-1}^2] \\
&= E[e_t^2] + 2b_1 E[e_t e_{t-1}] + b_1^2 E[e_{t-1}^2] \\
&= \sigma^2 + 2b_1 \times 0 + b_1^2 \sigma^2 = (1 + b_1^2)\sigma^2
\end{aligned}$$

  - The autocovariance at lag 1 is:

$$\begin{aligned}
\text{Covar at lag 1} &= E[(x_t - \mu)(x_{t-1} - \mu)] \\
&= E[e_t + b_1 e_{t-1})(e_{t-1} + b_1 e_{t-2})] \\
&= E[e_t e_{t-1} + b_1 e_{t-1} e_{t-1} + b_1 e_t e_{t-2} + b_1^2 e_{t-1} e_{t-2}] \\
&= E[0 + b_1 E[e_{t-1}^2] + 0 + 0] \\
&= b_1 \sigma^2
\end{aligned}$$

Moving Average (MA) Models

- **Autocorrelations for MA**
  - The autocovariance at lag 2 is:

$$\begin{aligned}\text{Covar at lag 2} &= E[(x_t-\mu)(x_{t-2}-\mu] \\ &= E[(e_t+b_1e_{t-1})(e_{t-2}+b_1e_{t-3})] \\ &= E[e_te_{t-2}+b_1e_{t-1}e_{t-2}+b_1e_te_{t-3}+b_1^2e_{t-1}e_{t-3}] \\ &= 0+0+0+0=0\end{aligned}$$

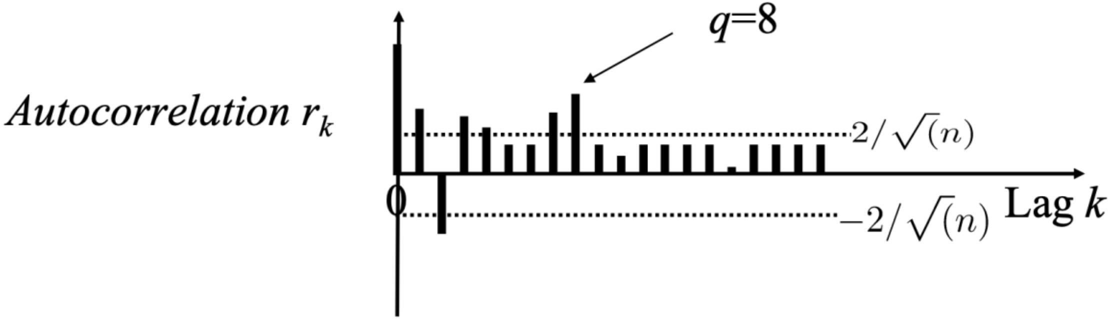
  - For MA(1), the autocovariance at all higher lags (k>1) is 0.

$$r_k=\begin{cases}1 & k=0\\ \frac{b_1}{1+b_1^2} & k=1\\ 0 & k>1\end{cases}$$

  - The autocorrelation is:
  - The autocorrelation of MA(q) series is non-zero only for lags k< q and is zero for all higher lags.

Moving Average (MA) Models

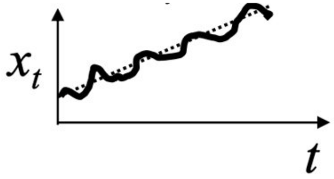
- **Determining the Order MA(q)**

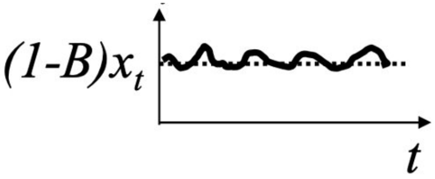


  - The order of the last significant rk determines the order of the MA(q) model.

Non-Stationarity: Integrated Models

- In the white noise model AR(0): $x_t = a_0 + e_t$
- The mean a0 is independent of time.
- If it appears that the time series in increasing approximately linearly with time, the first difference of the series can be modeled as white noise: $(x_t - x_{t-1}) = a_0 + e_t$
- Or using the B operator: (1-B)xt = xt-xt-1 $(1 - B)x_t = a_0 + e_t$
- This is called an "integrated" model of order 1 or I(1). Since the errors are integrated to obtain x.
- Note that xt is not stationary but (1-B)xt is stationary.



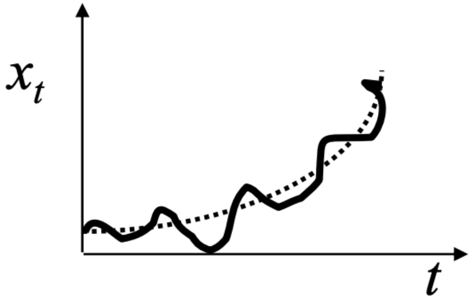


Non-Stationarity: Integrated Models

- If the time series is parabolic, the second difference can be modeled as white noise:

$$(x_t - x_{t-1}) - (x_{t-1} - x_{t-2}) = a_0 + e_t$$

- Or $(1 - B)^2 x_t = a_0 + e_t$
- This is an I(2) model.



ARMA and ARIMA Models

- It is possible to combine AR, MA, and I models
- ARMA(p, q) Model:

$$\begin{aligned} x_t - a_1 x_{t-1} - \ldots - a_p x_{t-p} &= a_0 + e_t + b_1 e_{t-1} + \ldots + b_q e_{t-q} \\ \phi_p(B) x_t &= a_0 + \psi_q(B) e_t \end{aligned}$$

- ARIMA(p,d,q) Model:

$$\phi_p(B)(1-B)^d x_t = a_0 + \psi_q(B) e_t$$

Non-Stationarity due to Seasonality

- The mean temperature in December is always lower than that in November and in May it always higher than that in March
- Temperature has a yearly season.
- One possible model could be I(12):

$$x_t - x_{t-12} = a_0 + e_t$$

- Or

$$(1 - B^{12})x_t = a_0 + e_t$$

Seasonal ARIMA (SARIMA) Models

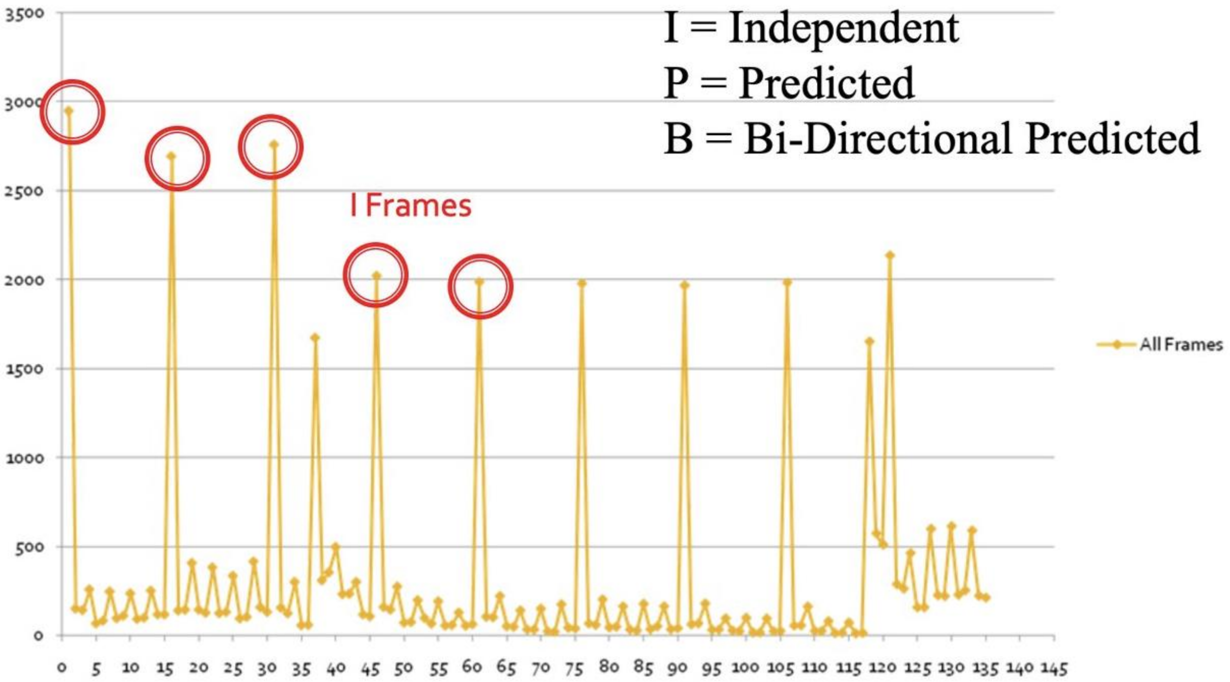
- SARIMA $(p, d, q) \times (P, R, Q)^s$ Model:

$$\phi_p(B)\Phi_P(B^s)(1 - B^s)^R(1 - B)^d x_t = a_0 + \psi_q(B)\Psi_Q(B^s)e_t$$

- Fractional ARIMA (FARIMA) Models ARIMA(p, d+δ, q) $-0.5<\delta<0.5$

=> Fractional Integration allowed.

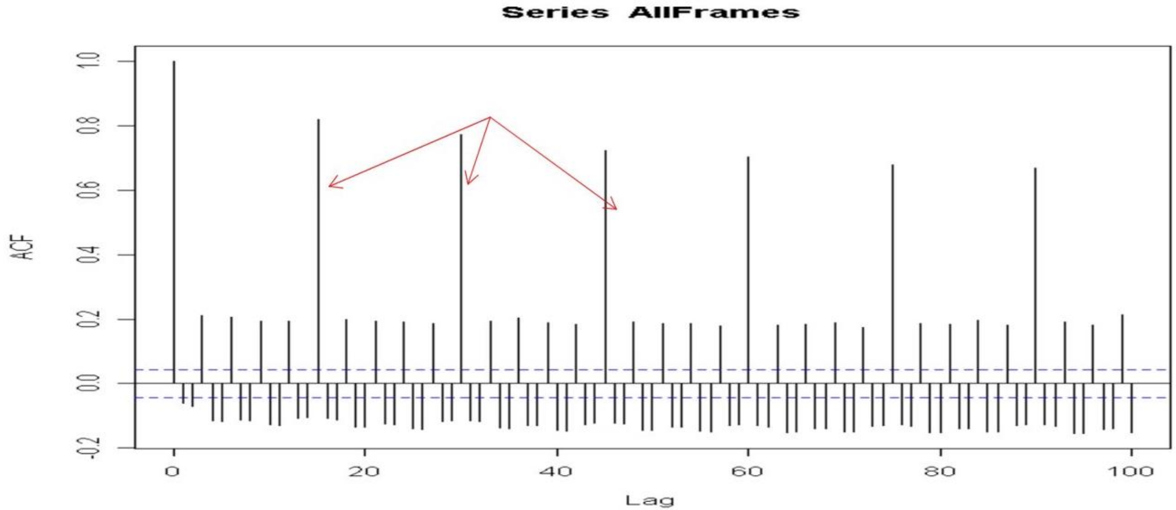
Case Study: Mobile Video



Observation: Every 15th frame is a large (I) frame.

Traffic Modeling – All Frames

A closer look at the ACF graph shows a strong continual correlation every 15 lag ➔ GOP size



Result: SARIMA (1, 0, 1)x$(1,1,1)^s$ Model, s=group size =15

Summary

- AR(1) Model:

$$x_t = a_0 + a_1 x_{t-1} + e_t$$

- MA(1) Model:

$$x_t - a_0 = e_t + b_1 e_{t-1}$$

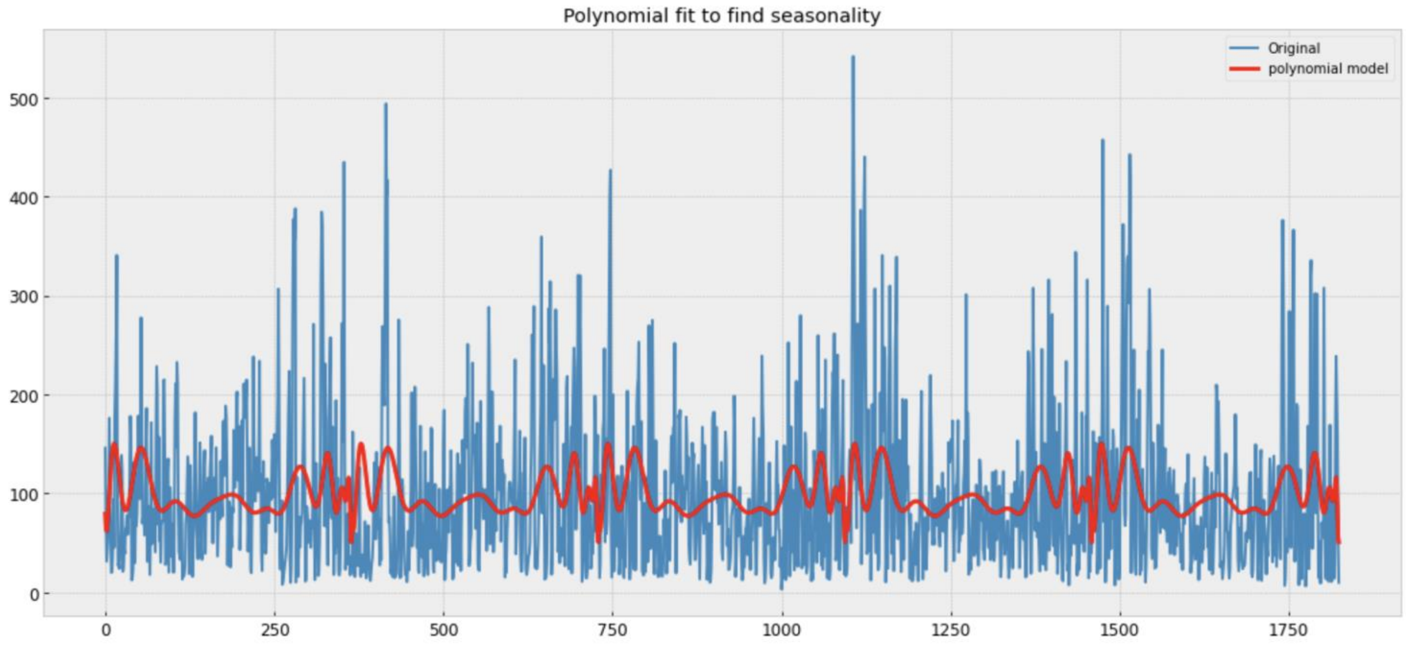
- ARIMA(1,1,1) Model:

$$x_t - x_{t-1} = a_0 + a_1(x_{t-1} - x_{t-2}) + e_t + b_1 e_{t-1}$$

- Seasonal ARIMA (1,0,1)x(0,1,0)12 model:

$$x_t - x_{t-12} = a_0 + a_1(x_{t-1} - x_{t-13}) + e_t + b_1 e_{t-1}$$

Time series forecasting with python

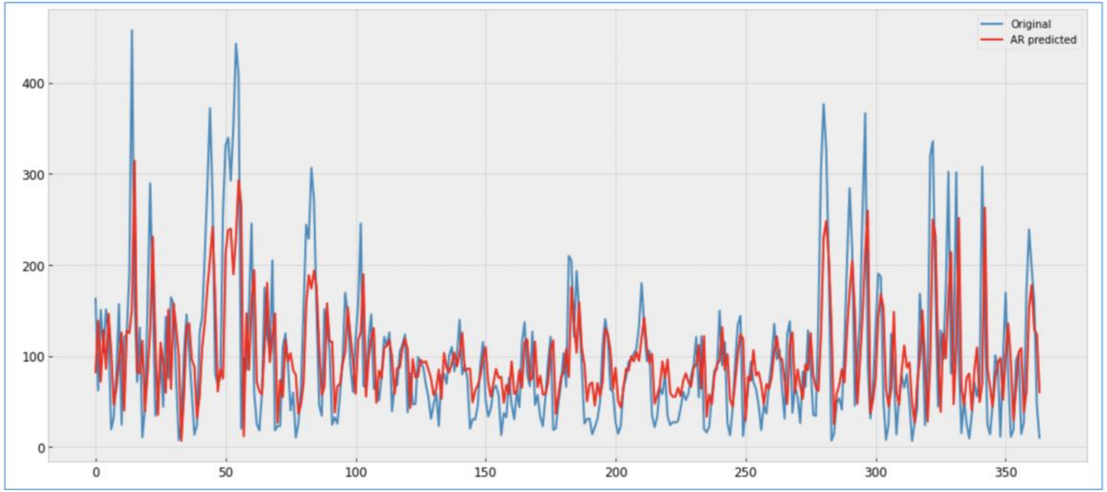


Time series forecasting with python

- Autoregression (AR)

```python
# Walk throught the test data, training and predicting 1 day ahead for all the test data
index = len(df_training)
yhat = list()
for t in tqdm(range(len(df_test.pollution_today))):
    temp_train = air_pollution[:len(df_training)+t]
    model = AR(temp_train.pollution_today)
    model_fit = model.fit()
    predictions = model_fit.predict(
        start=len(temp_train), end=len(temp_train),
    yhat = yhat + [predictions]

yhat = pd.concat(yhat)
resultsDict['AR'] = evaluate(df_test.pollution_toda
predictionsDict['AR'] = yhat.values
```



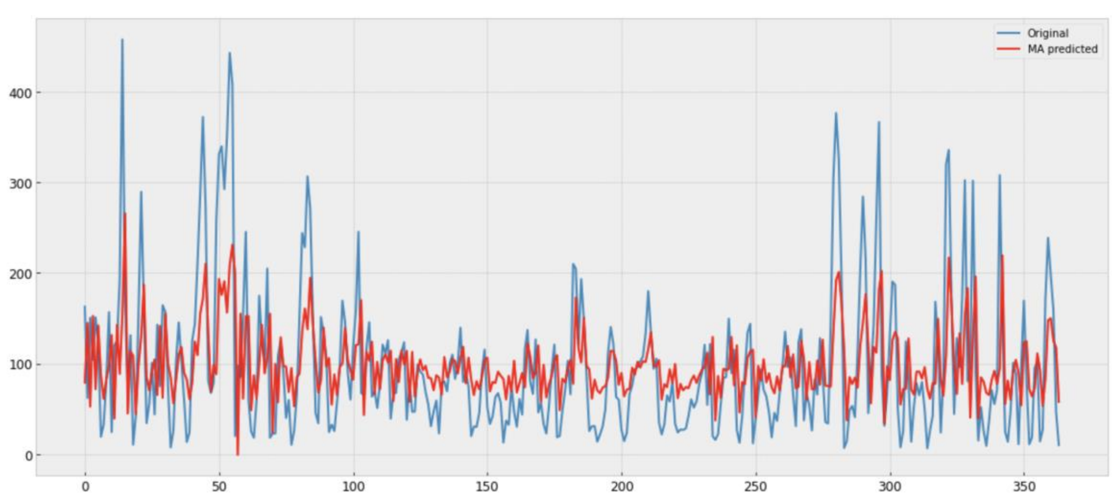
Time series forecasting with python

- Moving Average (MA)

```
# MA example

# Walk throught the test data, training and predicting 1 day ahead for all the test data
index = len(df_training)
yhat = list()
for t in tqdm(range(len(df_test.pollution_today))):
    temp_train = air_pollution[:len(df_training)+t]
    model = ARMA(temp_train.pollution_today, order=(0, 1))
    model_fit = model.fit(disp=False)
    predictions = model_fit.predict(
        start=len(temp_train), end=len(temp_tr
    yhat = yhat + [predictions]

yhat = pd.concat(yhat)
resultsDict['MA'] = evaluate(df_test.pollution
predictionsDict['MA'] = yhat.values
```



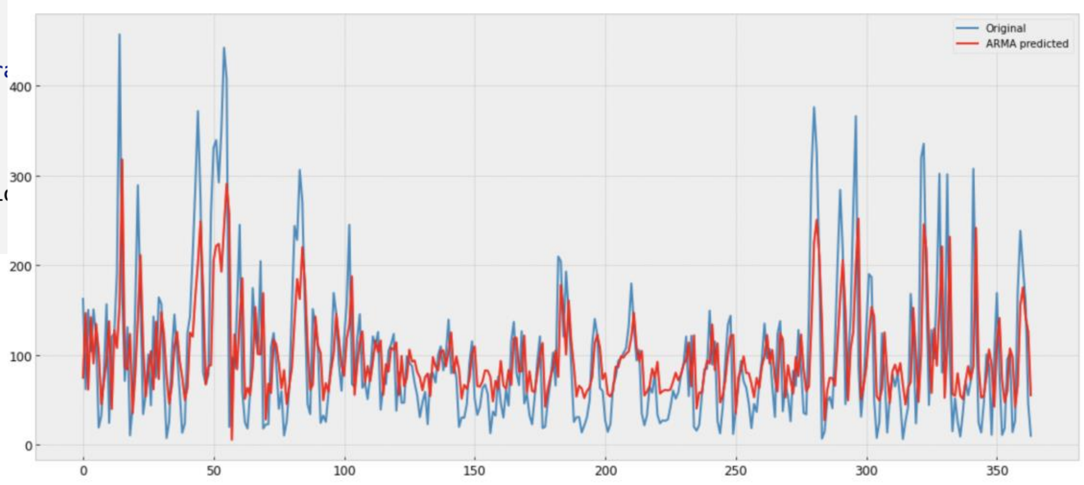
Time series forecasting with python

- Autoregressive Moving Average (ARMA)

```
# ARMA example

# Walk throught the test data, training and predicting 1 day ahead for all the test data
index = len(df_training)
yhat = list()
for t in tqdm(range(len(df_test.pollution_today))):
    temp_train = air_pollution[:len(df_training)+t]
    model = ARMA(temp_train.pollution_today, order=(1, 1))
    model_fit = model.fit(disp=False)
    predictions = model_fit.predict(
        start=len(temp_train), end=len(temp_tra
    yhat = yhat + [predictions]

yhat = pd.concat(yhat)
resultsDict['ARMA'] = evaluate(df_test.polluti
predictionsDict['ARMA'] = yhat.values
```



Time series forecasting with python

- Autoregressive integrated moving average (ARIMA)

```
# ARIMA example

# Walk throught the test data, training and predicting 1 day ahead for all the test data
index = len(df_training)
yhat = list()
for t in tqdm(range(len(df_test.pollution_today))):
    temp_train = air_pollution[:len(df_training)+t]
    model = ARIMA(temp_train.pollution_today, order=(1, 0, 0))
    model_fit = model.fit(disp=False)
    predictions = model_fit.predict(
        start=len(temp_train), end=len(temp_tra
    yhat = yhat + [predictions]

yhat = pd.concat(yhat)
resultsDict['ARIMA'] = evaluate(df_test.pollut
predictionsDict['ARIMA'] = yhat.values
```