AALBORG UNIVERSITY

STOCHASTIC SIGNALS AND PROCESSES

Lecture 6: Estimating The Parameters of Random Processes from Data

WELCOME

www.hst.aau.dk

Focus on estimation of:

- Autocorrelation functions
- Power spectral density functions

Model based or parametric
Model free or nonparametric

Ergodicity and Stationarity are assumed

Stationarity

Practical random sequences can be classified as:

1) Stationary over long periods of time
2) Stationary for short periods of time
3) Nonstationary

<u>Tests for stationarity</u>

1. Simple: The mean and varaince should not vary. Tested using *t-test* or *F-test*
2. Run test

Model-free Estimation

Directly estimate from the data

Autocorrelation Function estimation

$$\widehat{R_{XX}}(k) = \frac{1}{N-k}\sum_{i=0}^{N-k-1} X(i)X(i+k), k = 0 \ldots N-1$$

$$\widehat{R_{XX}}(-k) = \widehat{R_{XX}}(k)$$

1. We can only estimate $R_{XX}(k)$ for values of k less than N
2. As $E\left[\widehat{R_{XX}}(k)\right] = R_{XX}(k), k < N$. Thus unbiased for $k < N$

Model-free Estimation

Estimation of Power spectral density Function

For stationary processes,

$$S_{XX}(f) = \int_{-\infty}^{\infty} R_{XX}(\tau) \exp(-j2\pi f\tau)\, d\tau$$

Thus the estimator of the psd in the discrete case is

$$\widehat{S_{XX}}(f) = \sum_{k=-(N-1)}^{N-1} \widehat{R_{XX}}(k) \exp(-j2\pi f\tau), |f| < \frac{1}{2}$$

Problems:

1. Convolution is involved, thus high computational load
2. Large variance and Non negative values may occur for the estimate of psd.

Periodogram

$$\hat{\hat{S}}_{XX}(f) = \frac{1}{N}\left|X_F(f)\right|^2, \qquad |f| < \frac{1}{2}$$

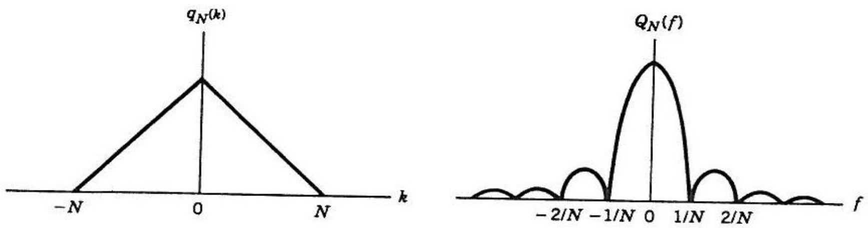
$X_F(f)$: Fourier transform of the data sequence.

It can be shown that the periodogram is a biased estimator of the psd.

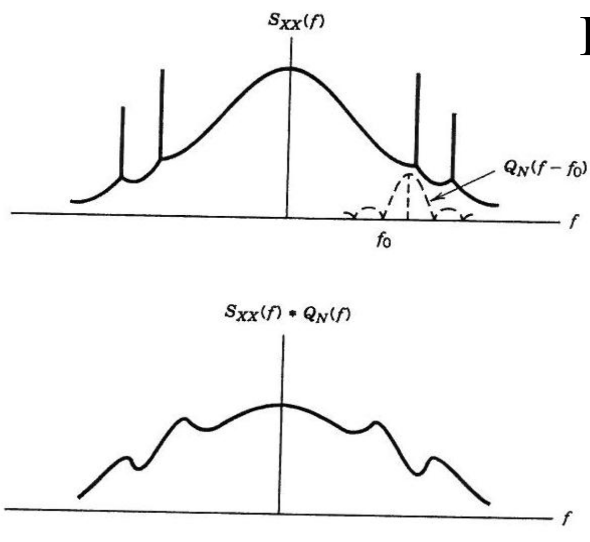
$$E\left\{\hat{\hat{S}}_{XX}(f)\right\} = \sum_{k=-\infty}^{\infty} q_N(k) r_{XX}(k) \exp(-j2\pi kf), \qquad |f| < \frac{1}{2}$$

$$= \int_{-1/2}^{1/2} S_{XX}(\alpha) Q_N(f-\alpha) d\alpha$$

$q_N(k)$: Triangular window function.
$Q_N(f)$: $(\mathrm{sinc})^2$ function.



(a)



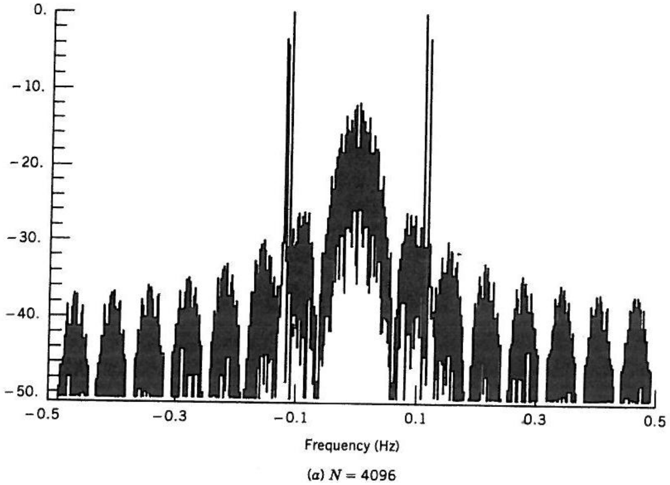
(b)

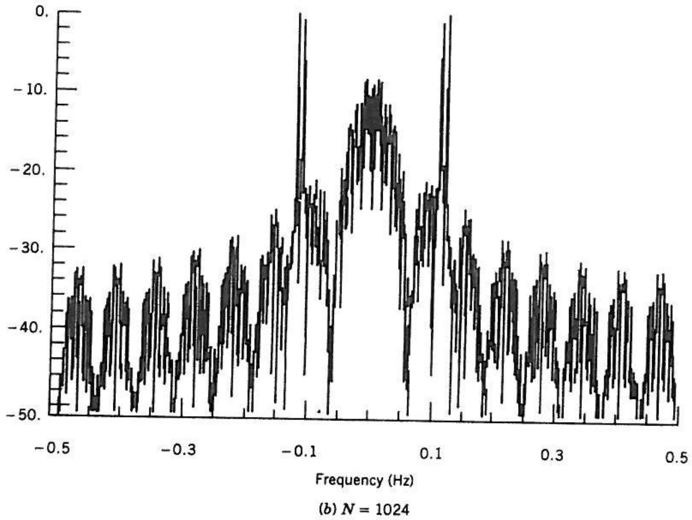
$$E\left\{\hat{\hat{S}}_{XX}(f)\right\} = \int_{-1/2}^{1/2} S_{XX}(\alpha) Q_N(f-\alpha)\, d\alpha$$

Spectrual leakage

Normalized standard error

$$\varepsilon_r = \frac{\sqrt{\mathrm{Var}\ \hat{\hat{S}}_{XX}(f)}}{S_{XX}(f)} = \frac{\sigma^2}{\sigma^2} = 100\%$$





Poor estimator as the variance does not depend on N. However increased N provides better resolution in freq.

Smoothing the periodogram

Segmentation in the time domain:

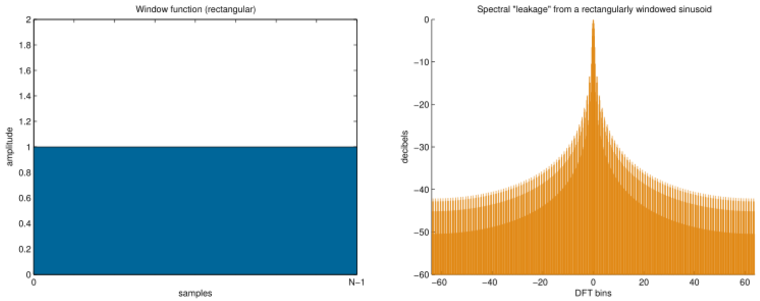
- Segment the data sequence
- Compute the psd of each segment
- Average all psds.

Running average in the frequency domain.

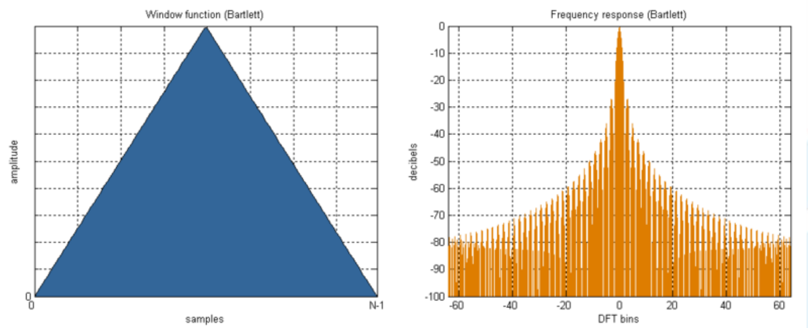
Effect: May decrease the variance, but increased bias

Windowing

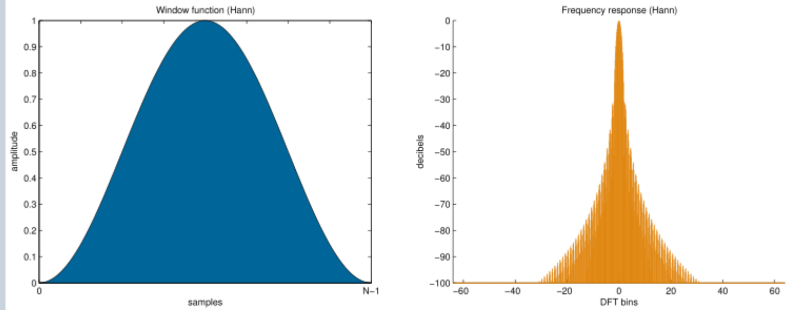
Rectangular window



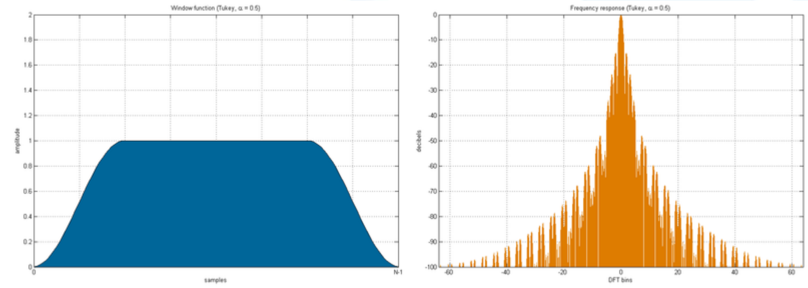
Bartlett window



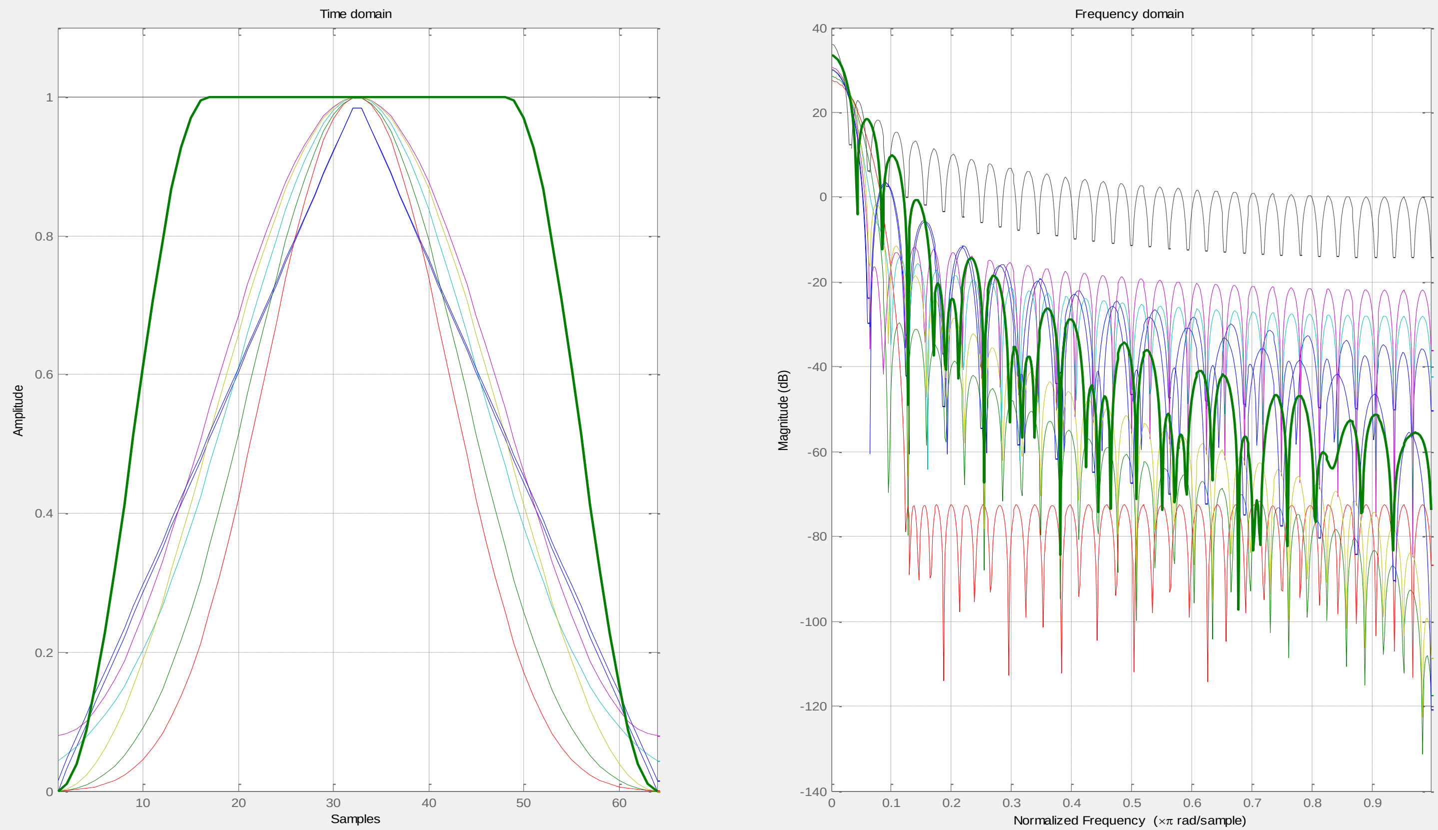
Hann window



Tukey window



AALBORG UNIVERSITY



Choose a window with most of the energy in a narrow main lobe and smaller side lobes => Reduced the amount of leakage.

Model-based estimation

Requires fewer data, the model is used as partial substitute of more data.

Steps:

1. Assume a form of the autocorrelation function or a model structure (type)
2. Estimate the parameters of this model
3. Is the model consistent with the data?

Model-based estimation

Many random sequences can be approximated by a rational transfer function

ARMA(p, q):

$$X(n) = \sum_{k=1}^{p} -a_k X(n-k) + e(n) + \sum_{l=1}^{q} b_l e(n-l)$$

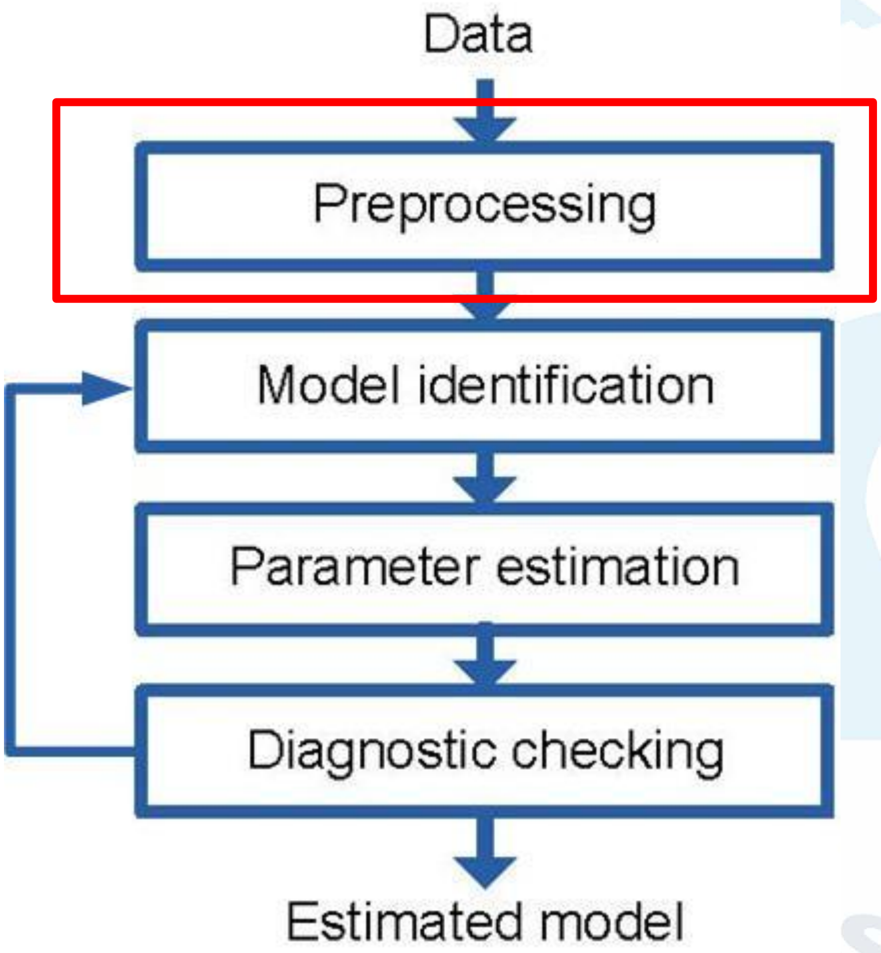
$$S_{XX}(f) = |H(f)|^2 S_{ee}(f), \qquad |f| < \frac{1}{2}$$

$$S_{ee}(f) = \sigma_n^2$$

Problem: Estimate the model parameters

Box-Jenkins iterative procedure

- Test for stationarity
- From nonstationary to quasistationary sequences



Preprocessing

- Logarithm transformation: $Y(n)=\ln X(n)$
  - ❖ Approximately constant variance but variation in mean still exist
- Differencing (Remove linear mean trends and periodic components):

$Y(n)=X(n) - X(n-1)$

If Y(n) is ARMA model, than X(n) is ARIMA model

$X(n) = X(n-1) + Y(n)$

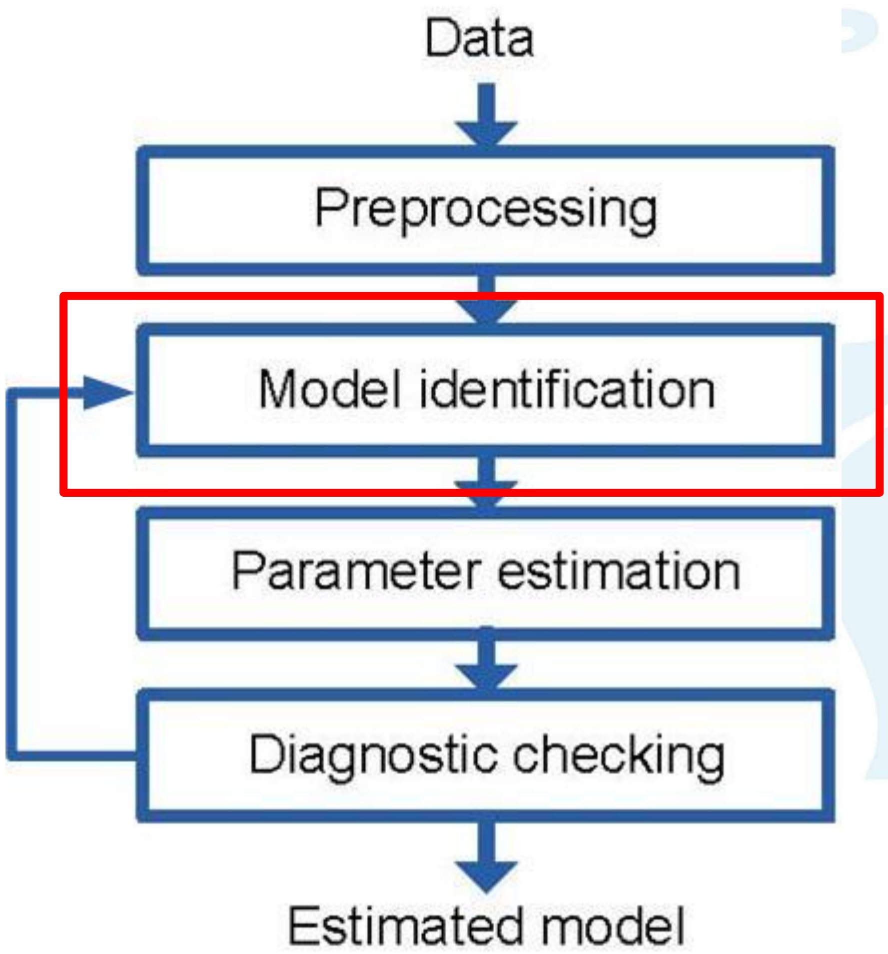
$= X(0) + Y(1) + Y(2) + \ldots + Y(n)$

X(n) is called the integrative version of Y(n)

AALBORG UNIVERSITY

Box-Jenkins iterative procedure

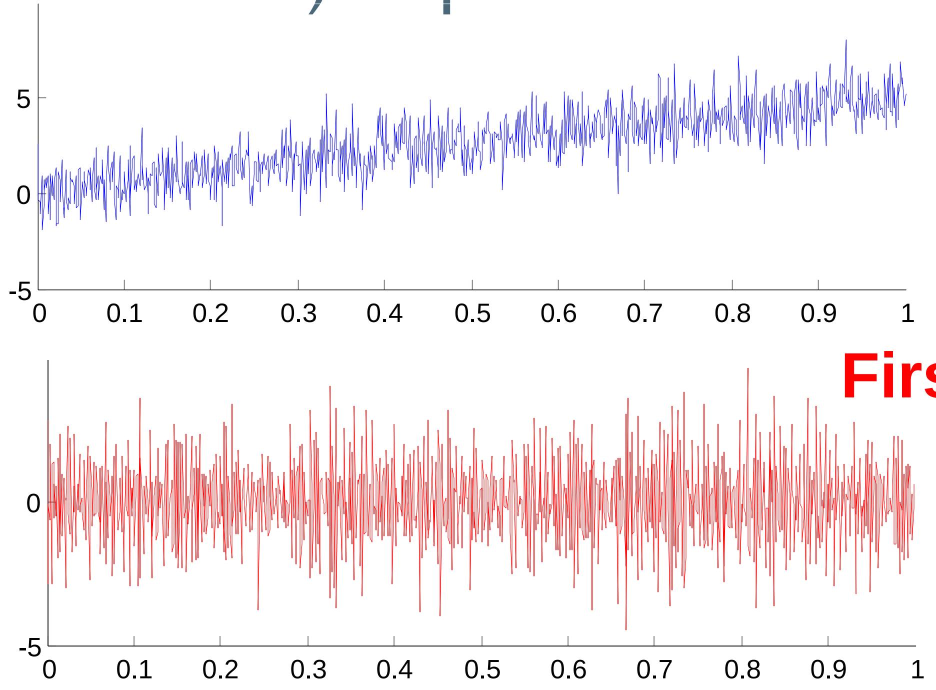
- Order identification:
- ARIMA(p,d,q)
- Order of differencing (d)
- Order ot the AR(p) part
- Order of the MA(q) part



www.hst.aau.dk

Order of differencing

Differencing as many times as necessary (usually not more than 2) to produce stationarity.



First order derivative

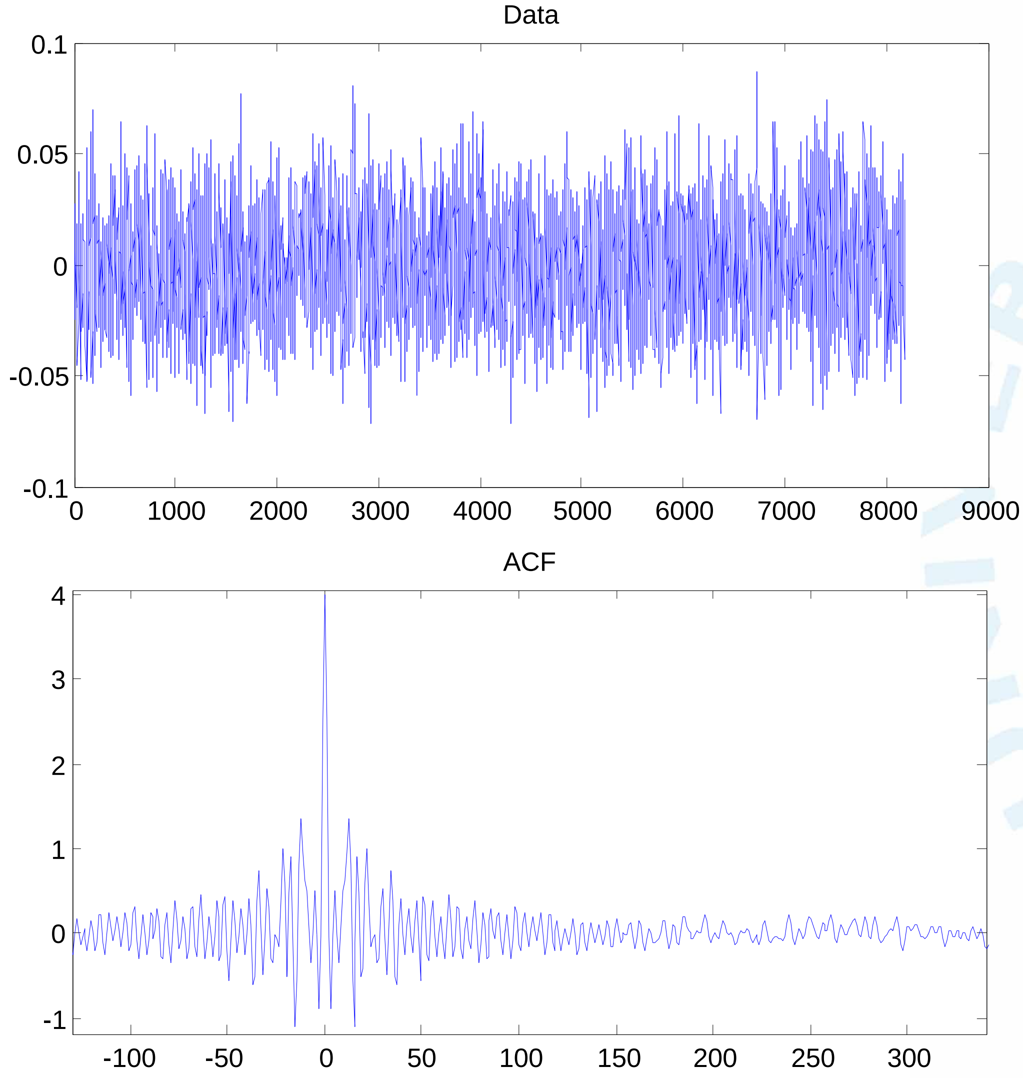
Order of ARMA model

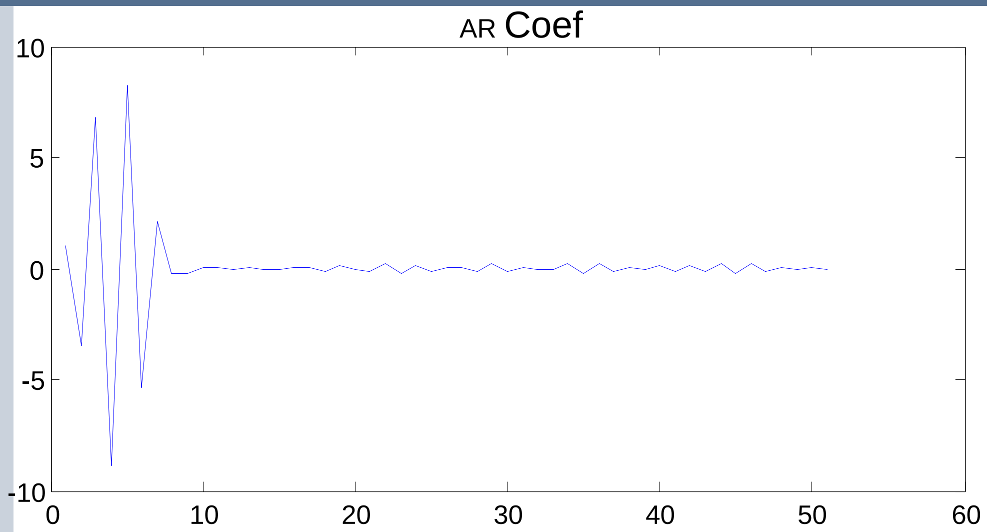
Partial autocorrelations function: The coefficient $a_{p,p}$ found from the Yule-Walker equation when $p = k$, is defined as the k^{th} partial autocorrelation coefficient.

$$\begin{bmatrix} \hat{r}(1) \\ \hat{r}(2) \\ \vdots \\ \hat{r}(p) \end{bmatrix} = \begin{bmatrix} \hat{r}(0) & \hat{r}(1) & \hat{r}(2) & \cdots & \hat{r}(p-1) \\ \hat{r}(1) & \hat{r}(0) & \hat{r}(1) & \cdots & \hat{r}(p-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}(p-1) & \hat{r}(p-2) & \hat{r}(p-3) & \cdots & \hat{r}(0) \end{bmatrix} \begin{bmatrix} a_{p,1} \\ a_{p,2} \\ \vdots \\ a_{p,p} \end{bmatrix}$$

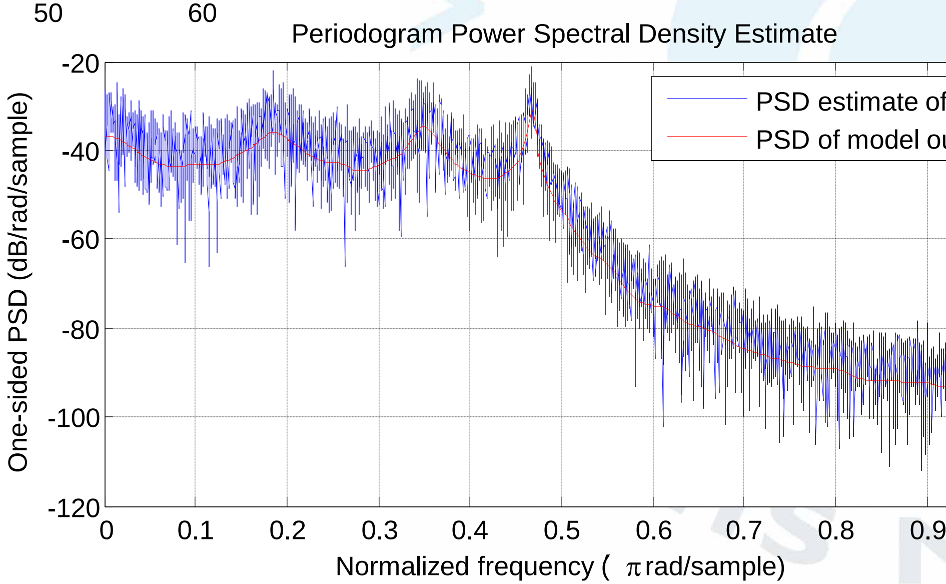
Partial Autocorrelation functions

SHAPE	INDICATED MODEL
Exponential, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to identify the order of the autoregressive model.
Alternating positive and negative, decaying to zero	Autoregressive model. Use the partial autocorrelation plot to help identify the order.
One or more spikes, rest are essentially zero	Moving average model, order identified by where plot becomes zero.
Decay, starting after a few lags	Mixed autoregressive and moving average model.
All zero or close to zero	Data is essentially random.
High values at fixed intervals	Include seasonal autoregressive term.
No decay to zero	Series is not stationary.


Data
0.1
0.05
0
-0.05
-0.1
0
1000
2000
3000
4000
5000
6000
7000
8000
9000
ACF
4
3
2
1
0
-1
-100
-50
0
50
100
150
200
250
300



Trial and errors. ACF has too much variance

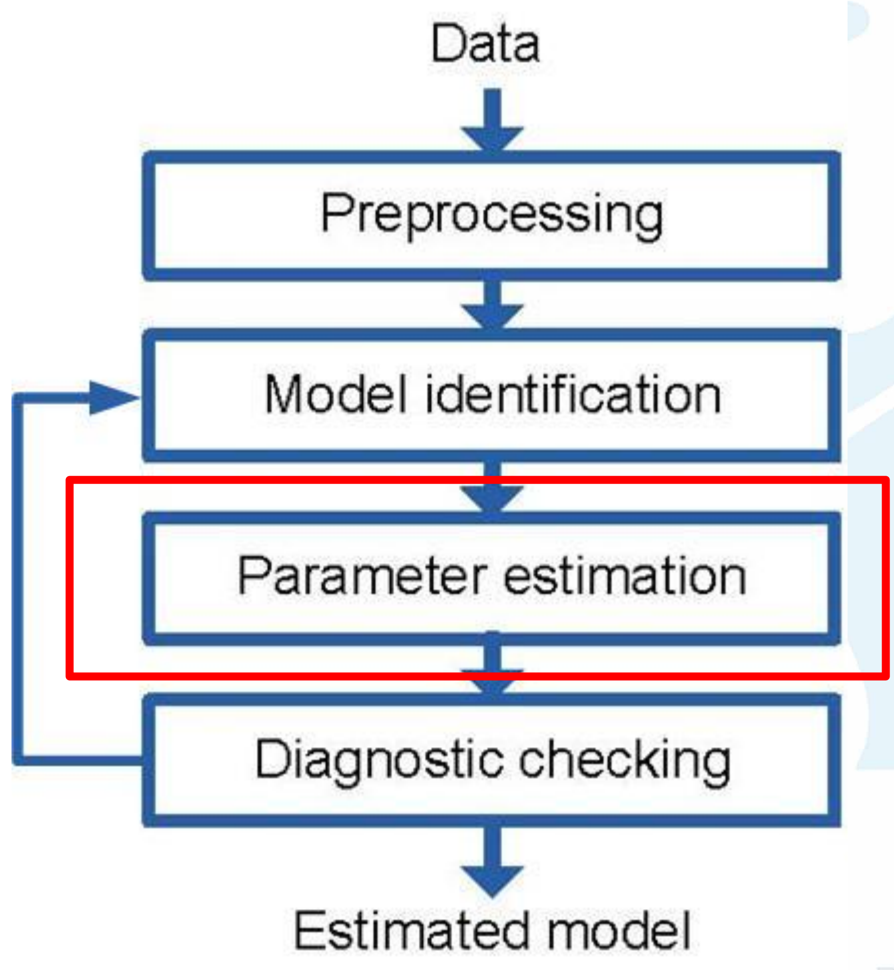


AALBORG UNIVERSITY

www.hst.aau.dk

Box-Jenkins iterative procedure

- Order identification:
- ARIMA(p,d,q)
- Order of differencing (d)
- Order ot the AR(p) part
- Order of the MA(q) part



Estimating the parameters of AR(p)

Yule-Walker equations

$$\begin{bmatrix} \hat{r}(1) \\ \hat{r}(2) \\ \vdots \\ \hat{r}(p) \end{bmatrix} = \begin{bmatrix} \hat{r}(0) & \hat{r}(1) & \hat{r}(2) & \cdots & \hat{r}(p-1) \\ \hat{r}(1) & \hat{r}(0) & \hat{r}(1) & \cdots & \hat{r}(p-2) \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \hat{r}(p-1) & \hat{r}(p-2) & \hat{r}(p-3) & \cdots & \hat{r}(0) \end{bmatrix} \begin{bmatrix} -a_1 \\ -a_2 \\ \vdots \\ -a_p \end{bmatrix}$$

- Yule-walker approach (autocorrelation approach)
- Burg method
- Covariation method
- Improved covariation method

Estimating the parameters of MA(q)

$$X(n)=\sum_{k=0}^{q} b_k e(n-k), \qquad b_0=1$$

More complicated than AR(p)

Apply a product of X(n+m) on both sides

$$r_{XX}(m)=\begin{cases}\sigma_n^2 \sum_{k=m}^{q} b_k b_{k-m}, & m\le q \\ 0, & m>q\end{cases}$$

Non-linear equations

Estimating the parameters of ARMA(p,q)

$$X(n) = \sum_{k=1}^{p} -a_k X(n-k) + e(n) + \sum_{l=1}^{q} b_l e(n-l)$$

Apply a product of *X*(*n*+*m*) on both sides

$$r_{XX}(m) = \begin{cases} -\sum_{k=1}^{p} a_k r_k(m-k) + \sigma_n^2 \sum_{k=m}^{q} b_k b_{k-m}, & m \le q \\ -\sum_{k=1}^{p} a_k r_k(m-k), & m > q \end{cases}$$

Related Matlab functions

m = **ar**(x, p, approach, window)

[a,e] = **aryule**(x, p)

[a,e] = **arburg**(x, p)

[a,e] = **arcov**(x, p)

[a,e] = **armcov**(x, p)

m = **armax**(data, [p, q])