



CEU
*Universidad
San Pablo*

Time Series Analysis

Session 0: Course outline

Carlos Óscar Sánchez Sorzano, Ph.D.
Madrid

Motivation for this course

The screenshot shows a Microsoft Internet Explorer window displaying the 'Time Series Data Library' website. The title bar reads 'Time Series Data Library - Microsoft Internet Explorer'. The menu bar includes 'Archivo', 'Edición', 'Ver', 'Favoritos', 'Herramientas', and 'Ayuda'. The toolbar includes standard buttons for back, forward, search, and file operations. The address bar shows the URL 'http://www-personal.buseco.monash.edu.au/~hyndman/TSDL/'. Below the address bar, a Google search bar is present. The main content area features a large blue background with a white wavy line graph on the left. The right side contains a sidebar with a list of subjects. A large black arrow points from the word 'Subjects' in the text below to the sidebar list. Another large black arrow points from the word 'Sources' in the text below to the 'Sources' link at the bottom of the sidebar.

Time Series Data Library

This is a collection of about 800 time series drawn from many different fields. To find a series, browse the subjects on the right or try the site search below.

Google SafeSearch Google Search

The Time Series Data Library was created by [Rob Hyndman](#). Please [let us know](#) if you find any broken links or other errors. We also welcome any additional time series to add to the library.

The time series may be freely copied and used, provided this source is clearly acknowledged. Citations to the Time Series Data Library should be as follows:

[Home](#)

[Agriculture](#)
[Chemistry](#)
[Crime](#)
[Demography](#)
[Ecology](#)
[Finance](#)
[Health](#)
[Hydrology](#)
[Industry](#)
[Labour market](#)
[Macro-Economics](#)
[Meteorology](#)
[Micro-Economics](#)
[Miscellaneous](#)
[Physics](#)
[Production](#)
[Sales](#)
[Simulated series](#)
[Sport](#)
[Transport & Tourism](#)
[Tree-rings](#)
[Utilities](#)

[Sources](#)

Course outline



Course outline: Session 1

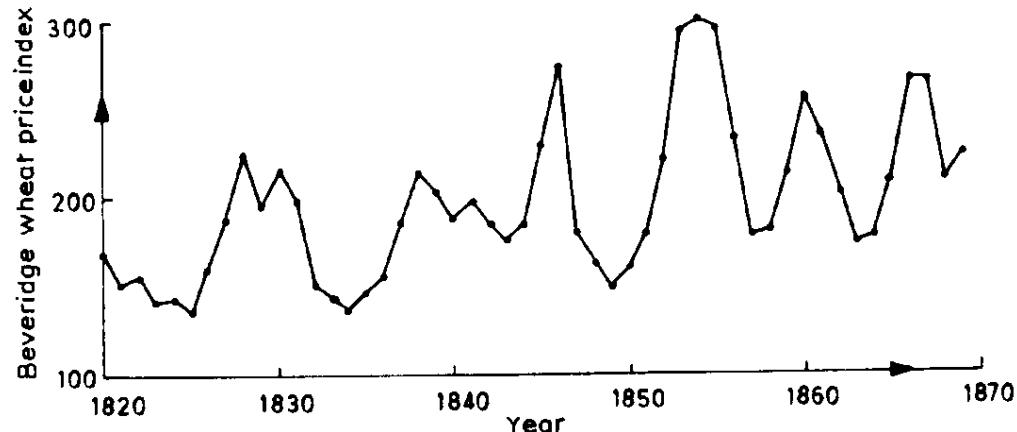
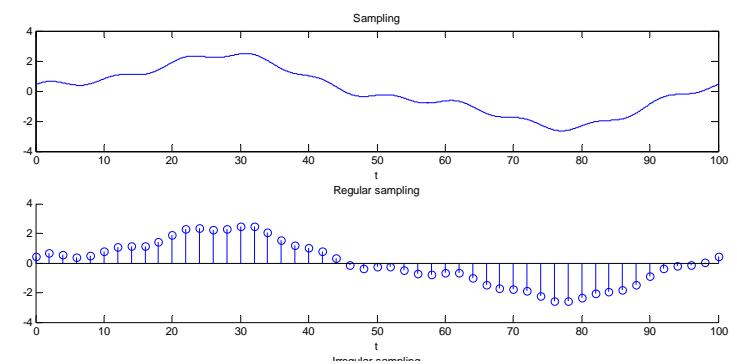


Figure 1.1 Part of the Beveridge wheat price index series.



- Kinds of time series
- Continuous time series sampling
- Data models $x[n] = trend[n] + periodic[n] + random[n]$
- Descriptive analysis: time plots and data preprocessing
- Distributional properties: statistical distribution, stationarity and autocorrelation
- Outlier detection and rejection

Course outline: Session 2

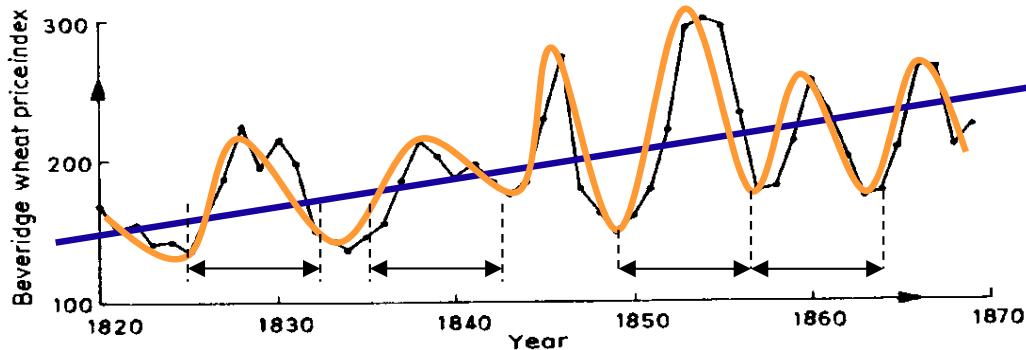


Figure 1.1 Part of the Beveridge wheat price index series.

$$x[n] = \boxed{\text{trend}[n] + \text{periodic}[n] + \text{random}[n]}$$

Trend analysis:

- Linear and non-linear regression
- Polynomial fitting
- Cubic spline fitting

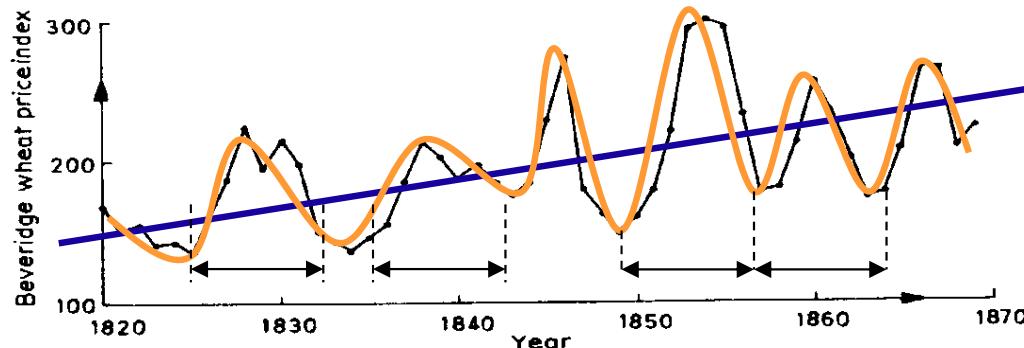
Seasonal component analysis:

- Spectral representation of stationary processes

Spectral signal processing:

- Detrending and filtering
- Non-stationary signal processing

Course outline: Session 3



$$x[n] = \text{trend}[n] + \text{periodic}[n] + \text{random}[n]$$

Model definition:

- Moving Average processes (MA)
- Autoregressive processes (AR)
- Autoregressive, Moving Average (ARMA)
- Autoregressive, Integrated, Moving Average (ARIMA, FARIMA)
- Seasonal, Autoregressive, Integrated, Moving Average (SARIMA)
- Known external inputs: System identification
- A family of models
- Nonlinear models

Parameter estimation

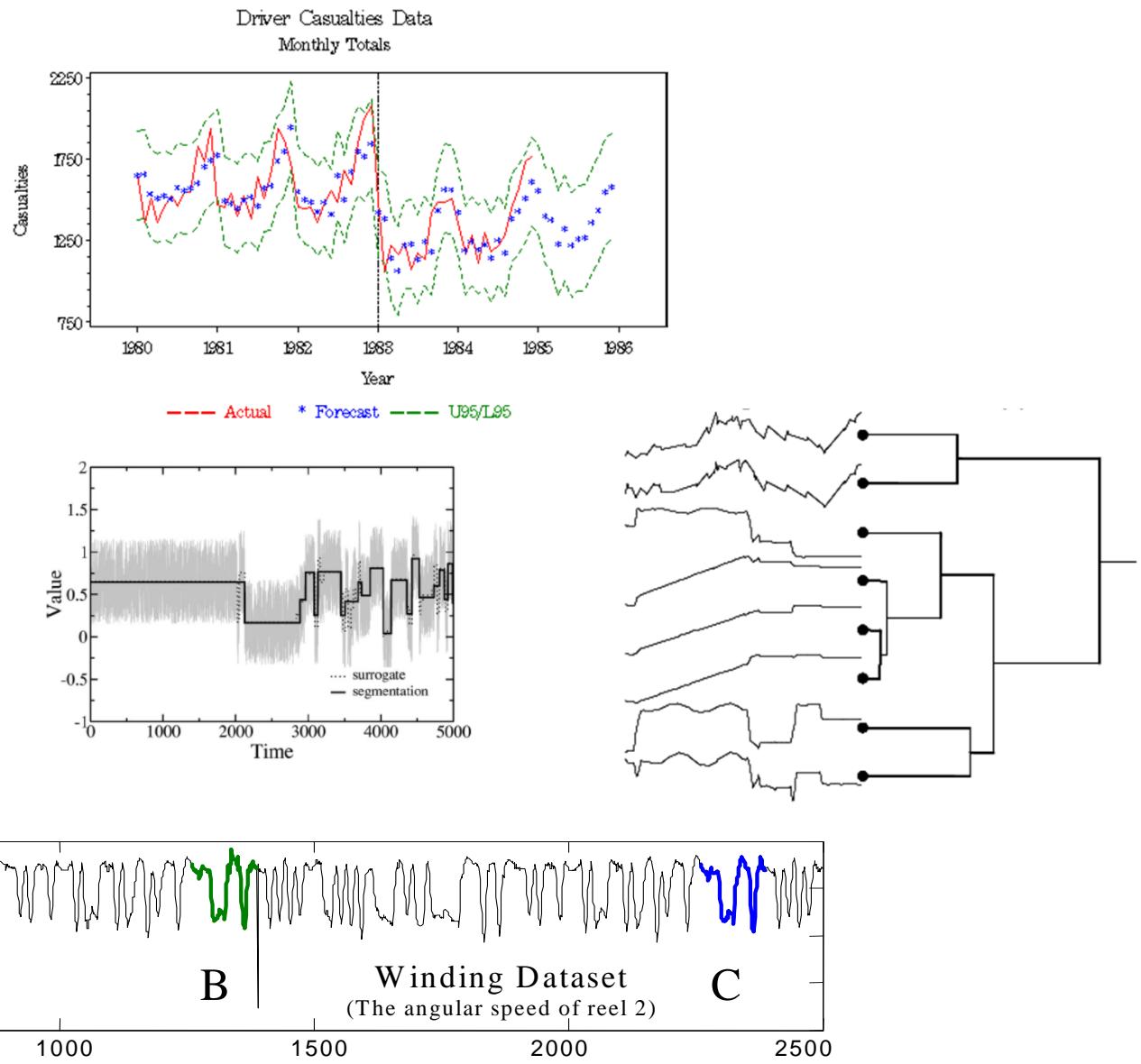
Order selection

Model checking

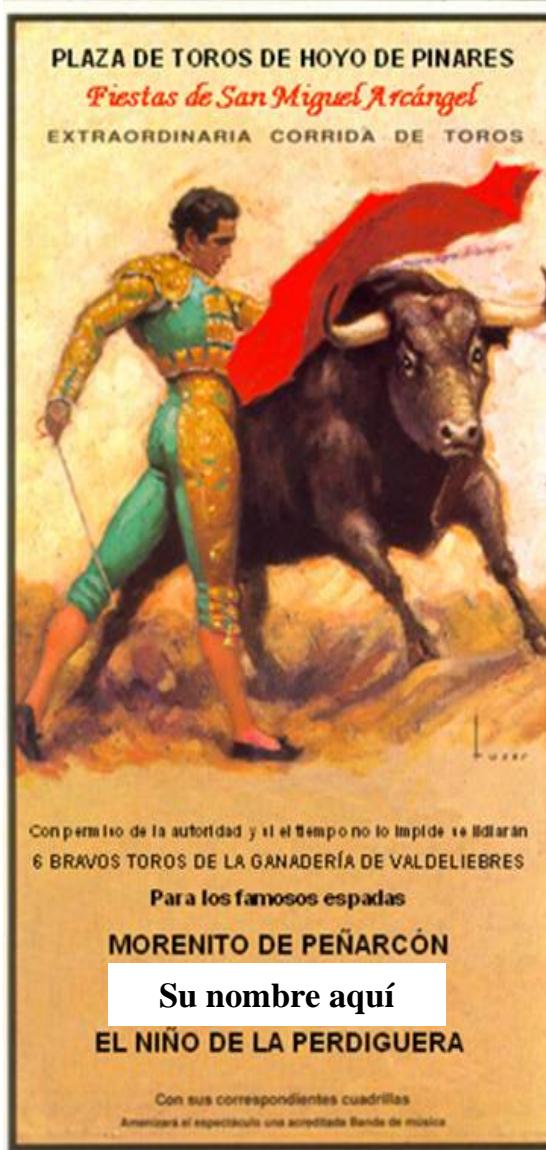
Self-similarity, fractal dimension and chaos theory

Course outline: Session 4

- Forecasting
 - Univariate forecasting
- Intervention modelling
- State-space modelling
- Time series data mining
 - Time series representation
 - Distance measure
 - Anomaly/Novelty detection
 - Classification/Clustering
 - Indexing
 - Motif discovery
 - Rule extraction
 - Segmentation
 - Summarization



Course Outline: Session 5



Bring your own data if possible!

Suggested readings

It is suggested to read (before coming):

- Geo4990: Time series analysis
- Adler1998: Analysing stable time series
- Leonard: Mining Transactional and Time Series Data
- Chattarjee2006: Simple Linear Regression

Resources

Data sets

<http://www.york.ac.uk/depts/mathsls/data/ts>

<http://www-personal.buseco.monash.edu.au/~hyndman/TSDL>

Competition

<http://www.neural-forecasting-competition.com>

Links to organizations, events, software, datasets

<http://www.buseco.monash.edu.au/units/forecasting/links.php>

http://www.secondmoment.org/time_series.php

Lecture notes

<http://www.econphd.net/notes.htm#Econometrics>

Bibliography

- D. Peña, G. C. Tiao, R. S. Tsay. *A course in time series analysis*. John Wiley and Sons, Inc. 2001
- C. Chatfield. *The analysis of time series: an introduction*. Chapman & Hall, CRC, 1996.
- C. Chatfield. *Time-series forecasting*. Chapman & Hall, CRC, 2000.
- D.S.G. Pollock. *A handbook of time-series analysis, signal processing and dynamics*. Academics Press, 1999.
- J. D. Hamilton. *Time series analysis*. Princeton Univ. Press, 1994.
- C. Pérez. *Econometría de las series temporales*. Prentice Hall (2006)
- A. V. Oppenheim, R. W. Schafer, J. R. Buck. *Discrete-time signal processing, 2nd edition*. Prentice Hall, 1999.
- A. Papoulis, S. U. Pillai. *Probability, random variables and stochastic processes, 4th edition*. McGraw Hill, 2002.



CEU
*Universidad
San Pablo*

Time Series Analysis

Session I: Introduction

Carlos Óscar Sánchez Sorzano, Ph.D.
Madrid

Session outline

1. Features and objectives of the time series
2. Sampling
3. Components of time series: data models
4. Descriptive analysis
5. Distributional properties
6. Detection and removal of outliers
7. Time series methods

1. Features and objectives of the time series

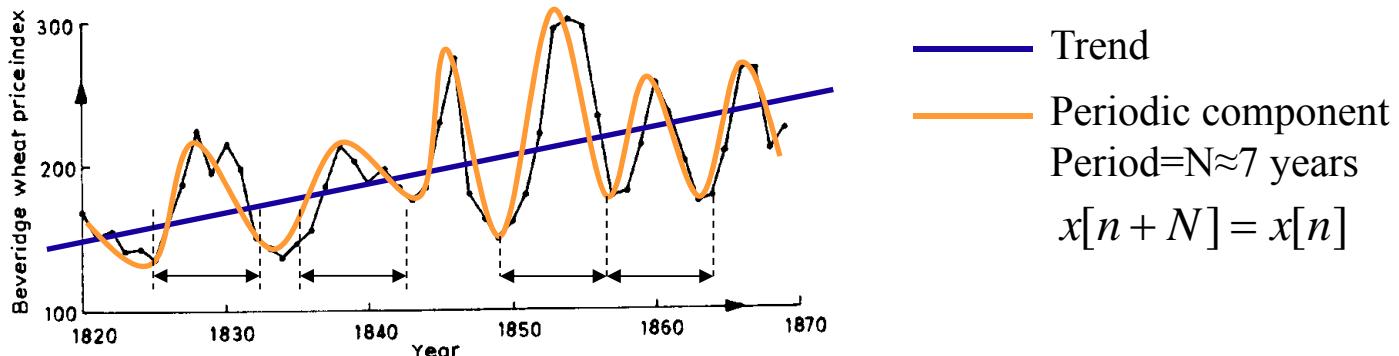


Figure 1.1 Part of the Beveridge wheat price index series.

Goal: Explain history

Features:

1. Samples (discrete) $x[n]$
2. Bounded
3. Finite support

$$x[n] = 0 \quad \forall n \notin \{n_0, n_0 + 1, \dots, n_F\}$$

$$x[n] \in l_1 = \left\{ s[n] \middle| \sum_{n=-\infty}^{\infty} |s[n]| < \infty \right\}$$

1. Features and objectives of the time series

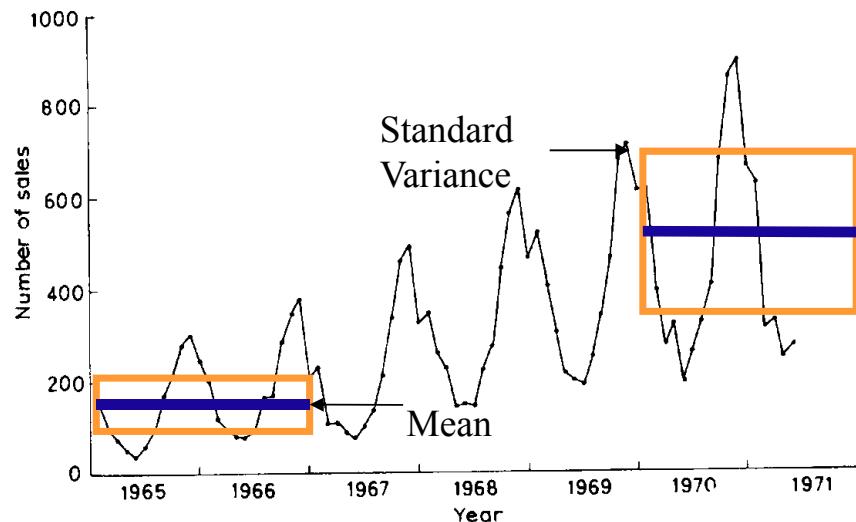


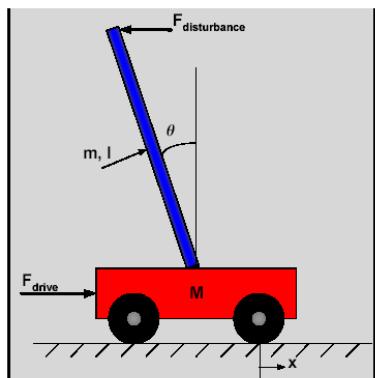
Figure 1.3 Sales of a certain engineering company in successive months.

Goal: Forecast demand

Features:

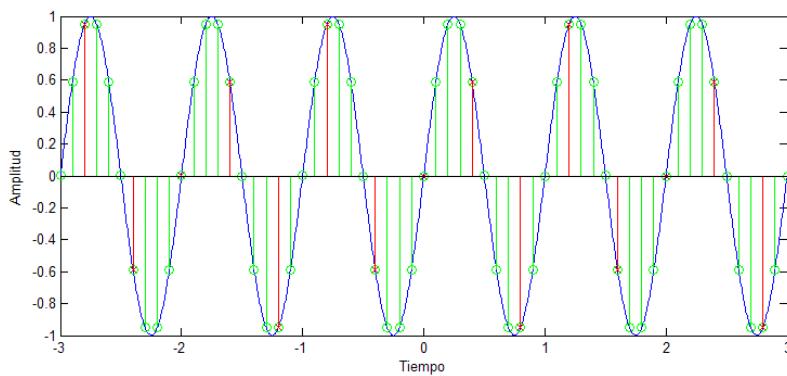
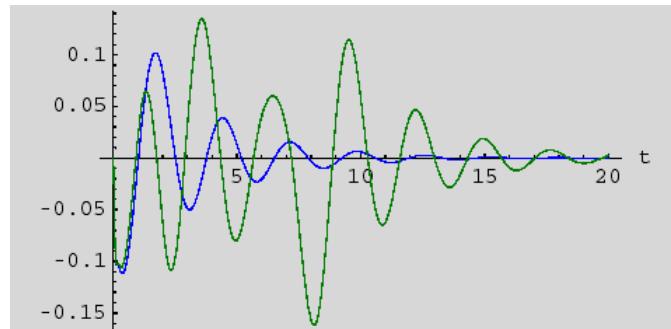
1. Seasonal
2. Non stationary
3. Non independent samples

1. Features and objectives of the time series



Sketch : Inverted Pendulum

Pendulum angle with different controllers



Goal:

Control=Forecast(+correct)

Features:

1. Continuous signal → Regular sampling

$$x[n] = x(nT_s)$$

T_s =Sampling period



1. Features and objectives of the time series

Other kind of times series not covered in this course

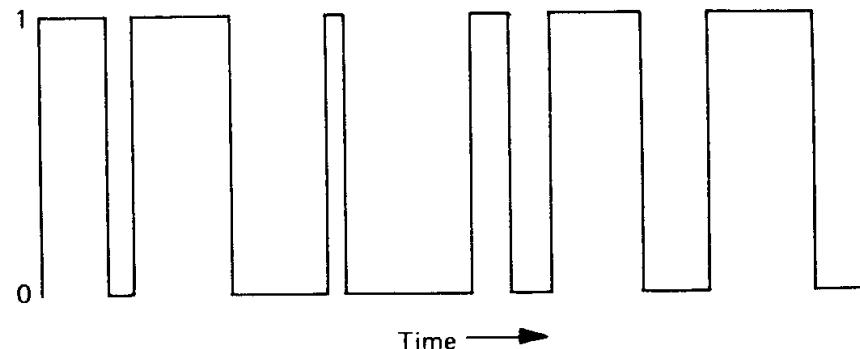


Figure 1.5 A realization of a binary process.

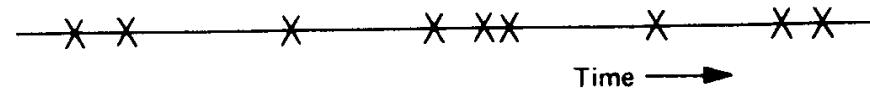
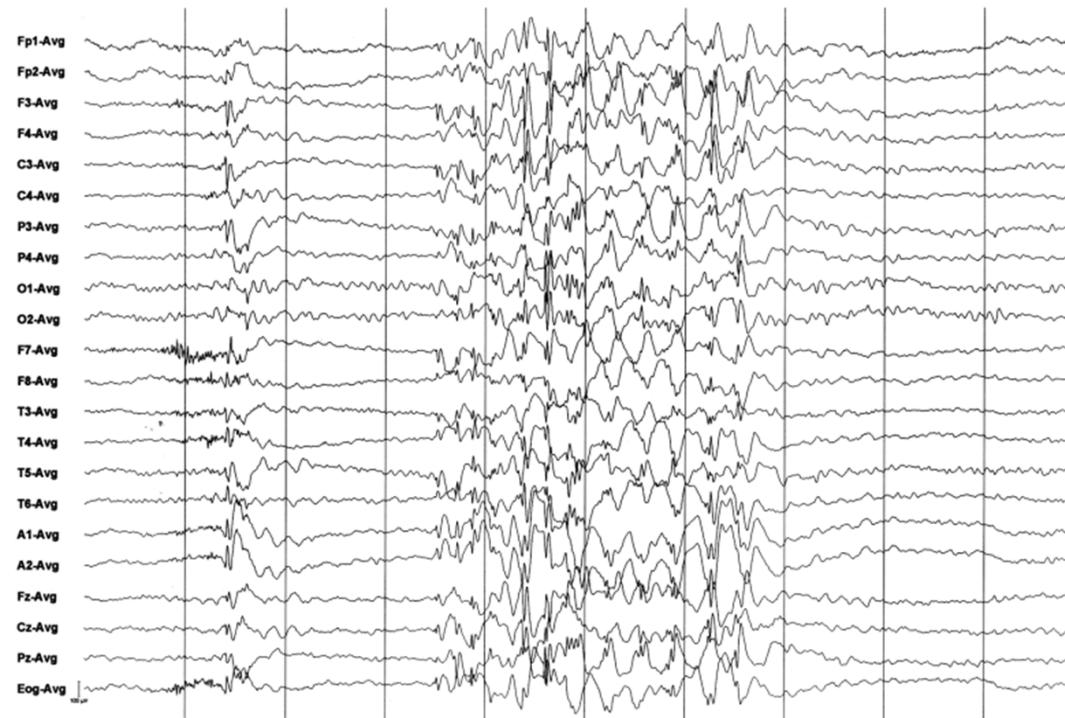


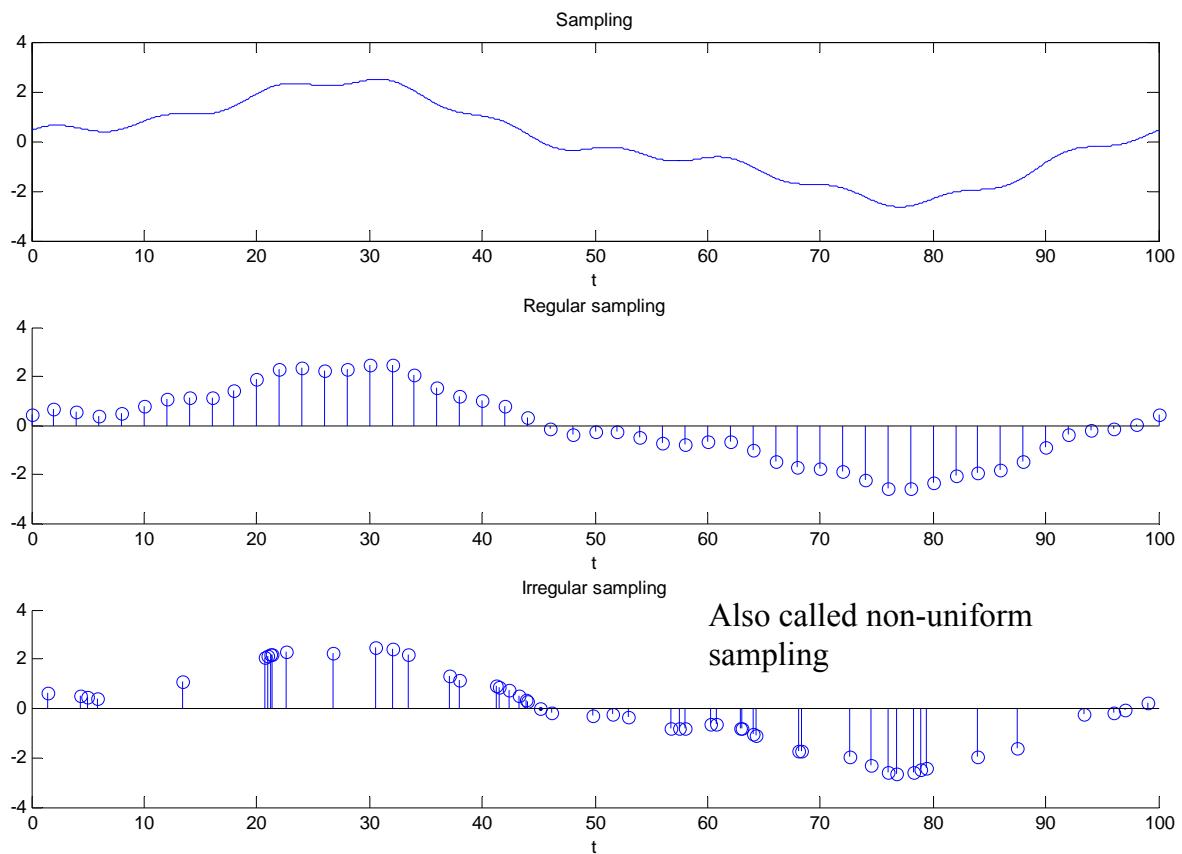
Figure 1.6 A realization of a point process (\times denotes an event).

1. Features and objectives of the time series

Other kind of times series not covered in this course



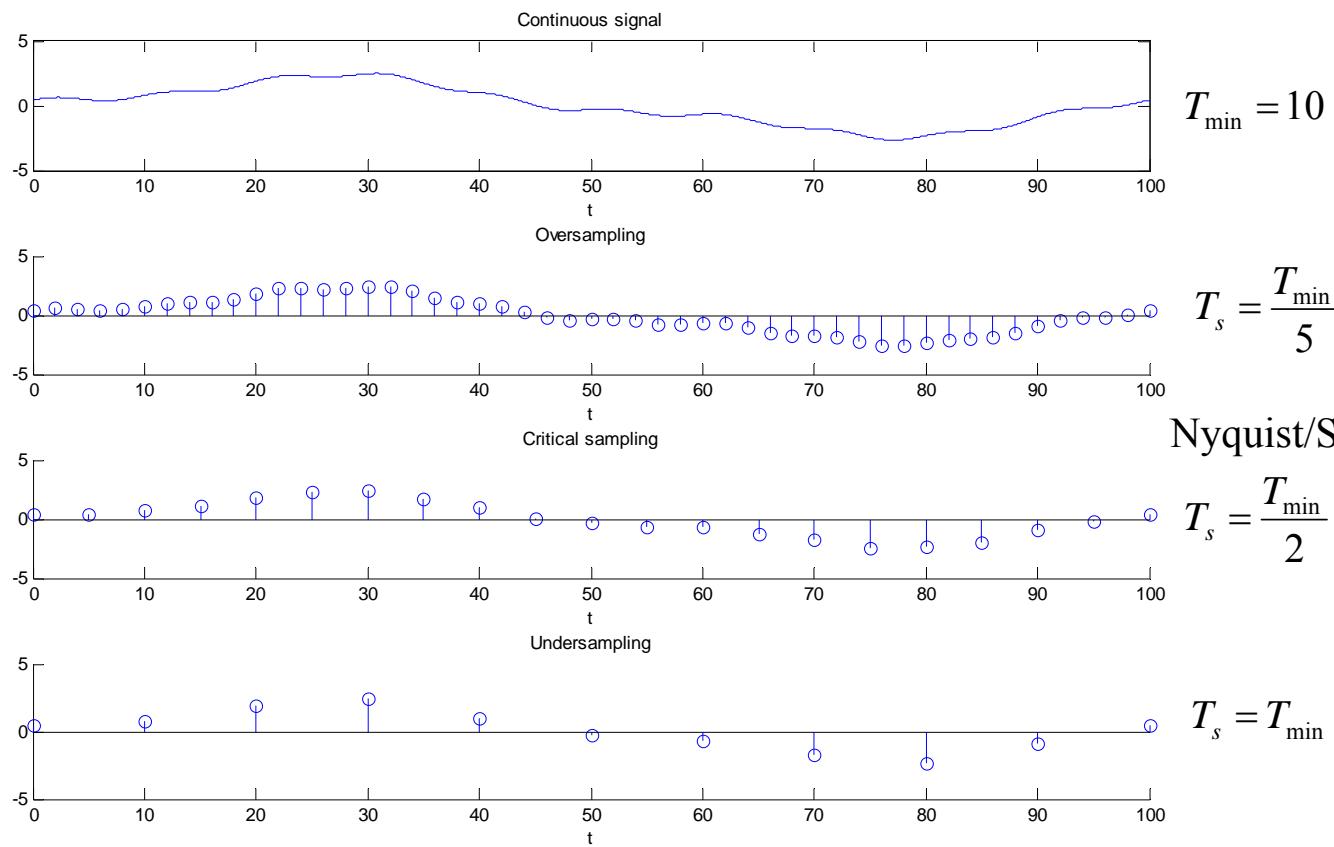
2. Sampling



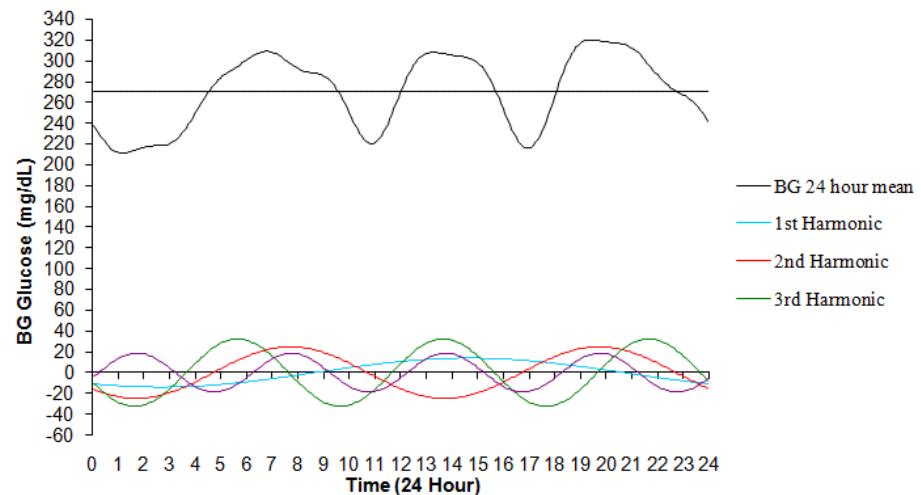
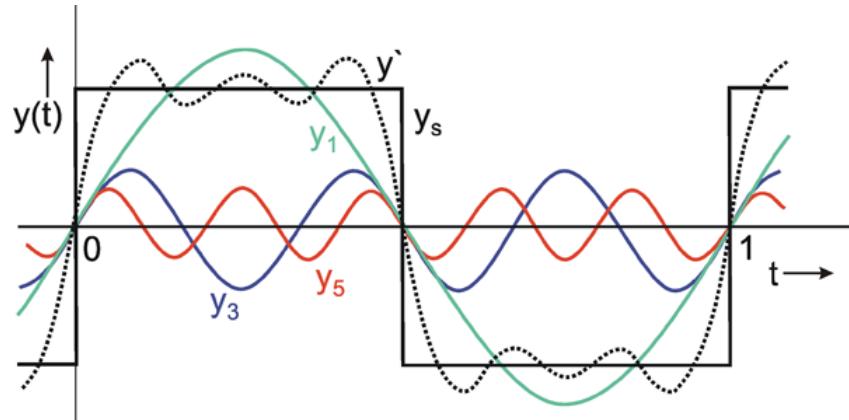
$$x[n] = x(nT_s)$$

2. Sampling

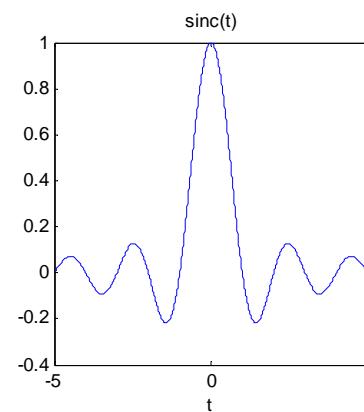
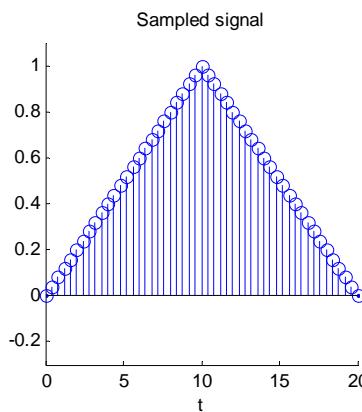
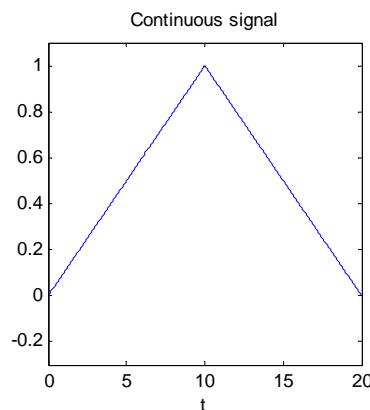
$$x(t) = 2 \sin\left(2\pi \frac{1}{100}t\right) - \frac{1}{2} \sin\left(2\pi \frac{3}{100}t - \frac{\pi}{4}\right) + 0.2 \sin\left(2\pi \frac{10}{100}t + \frac{\pi}{8}\right)$$



2. Sampling



2. Sampling: Signal reconstruction

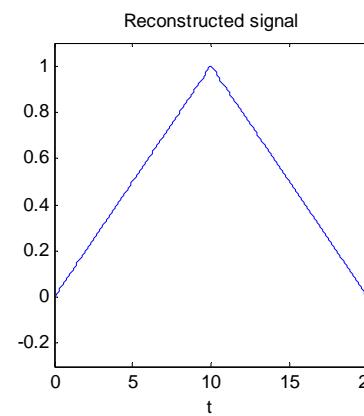
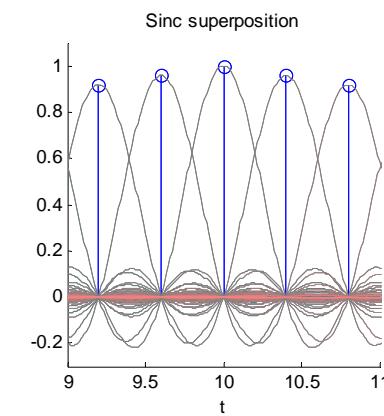
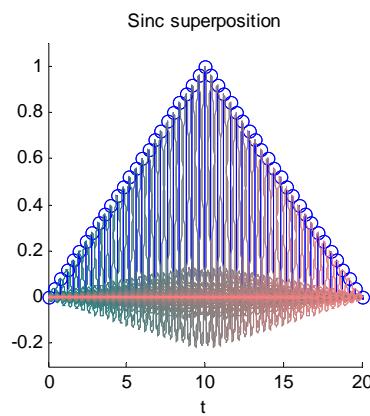


Reconstruction formula

$$x_r(t) = \sum_{n=-\infty}^{\infty} x[n]h_r(t-nT_s)$$

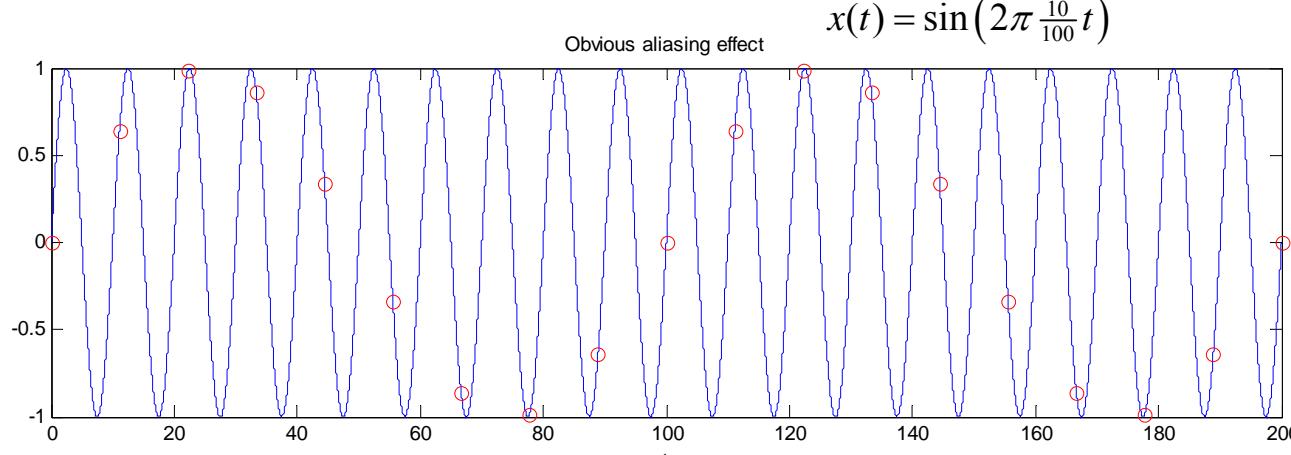
Reconstruction kernel
for bandlimited
signals

$$h_r(t) = \sin c\left(\frac{t}{T_s}\right)$$



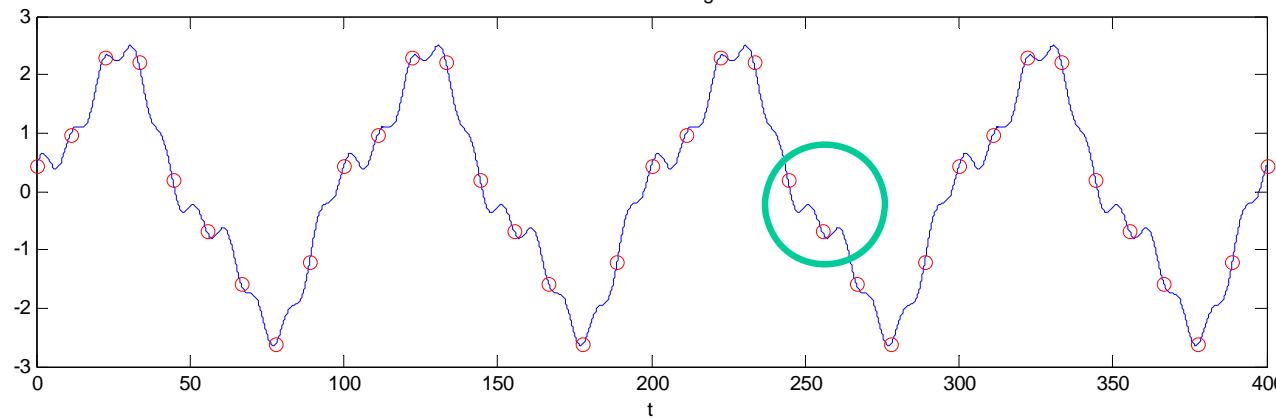
2. Sampling: Aliasing

Obvious aliasing effect



$$T_s = \frac{T_{\min}}{1.1} > \frac{T_{\min}}{2}$$

Not so clear aliasing effect



$$T_s = \frac{T_{\min}}{1.1} > \frac{T_{\min}}{2}$$



2. Aliasing: Conclusions

From the digression above, two are the main consequences that must be kept in mind:

1. Any continuous time series can be safely treated as a discrete time series as long as the Nyquist criterion is satisfied.
2. Once our discrete analysis is finished, we can always return to the continuous “world” by reconstructing our output sequence.

Although not discussed, once the continuous signal is discretized, one can arbitrarily change the sampling period without having to go back to the continuous signal with two operations called “upsampling” (going for a finer discretization) and “downsampling (coarser discretization).

3. Components of a time series: data models

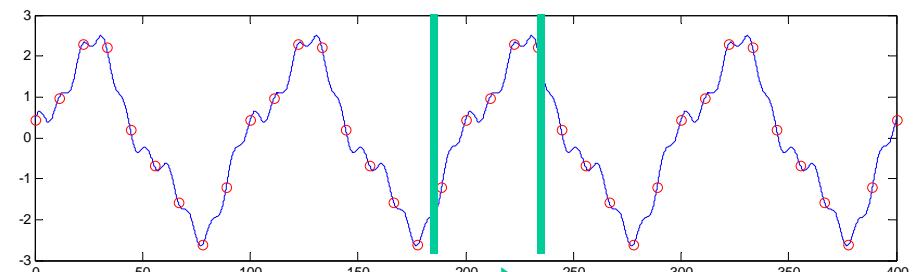
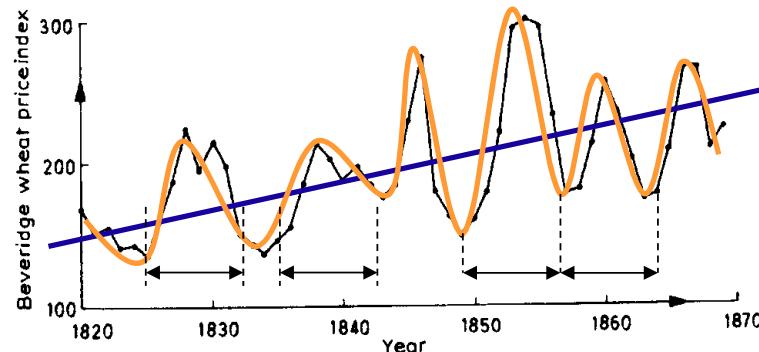


Figure 1.1 Part of the Beveridge wheat price index series.

$$x[n] = \text{trend}[n] + \text{periodic}[n] + \text{random}[n]$$

Seasonal

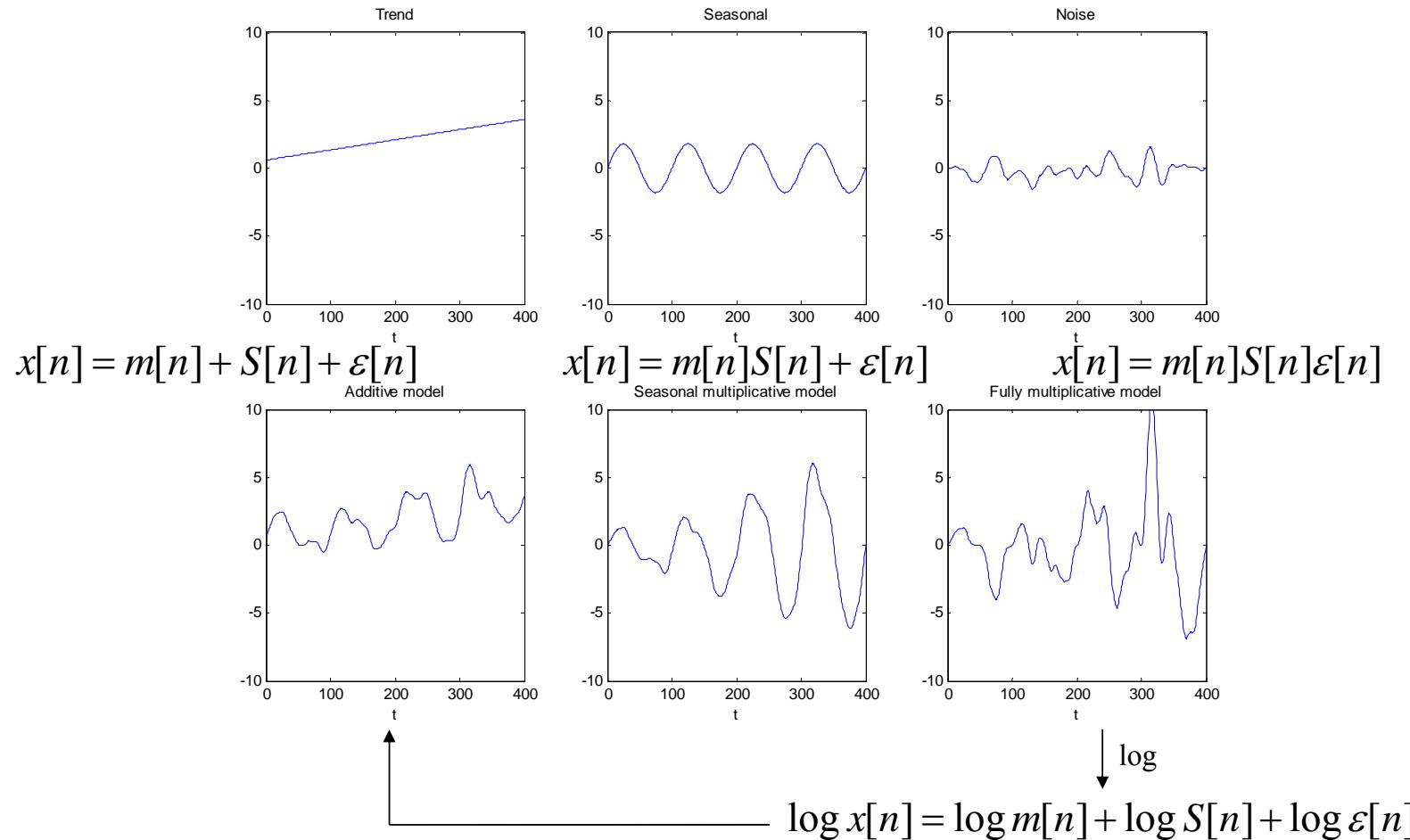
Ex: Unemployment is low in summer

Ex: Temperature is high in the middle of the day

Long-term change. What is long-term?

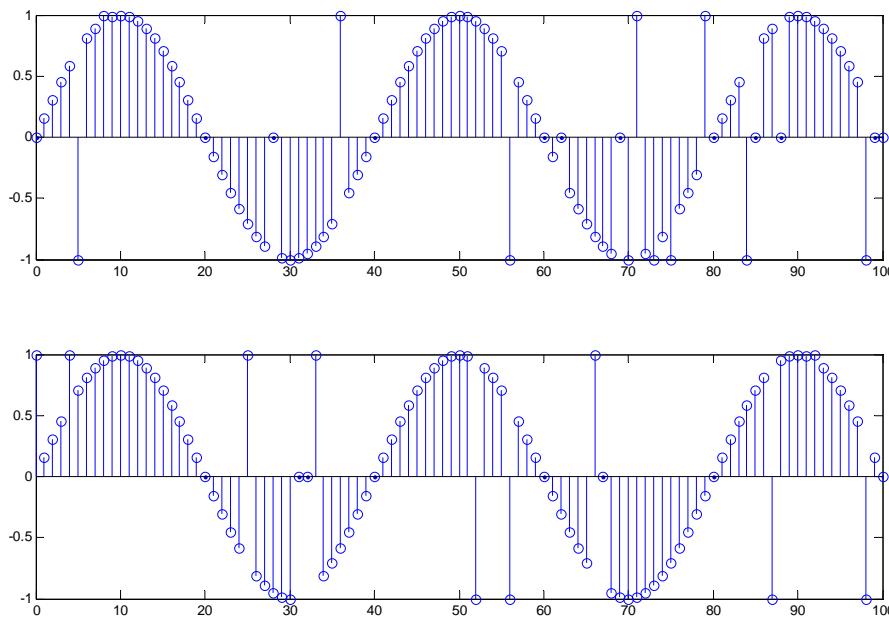
Explained by statistical models (AR, MA, ...)

3. Components of a time series: data models

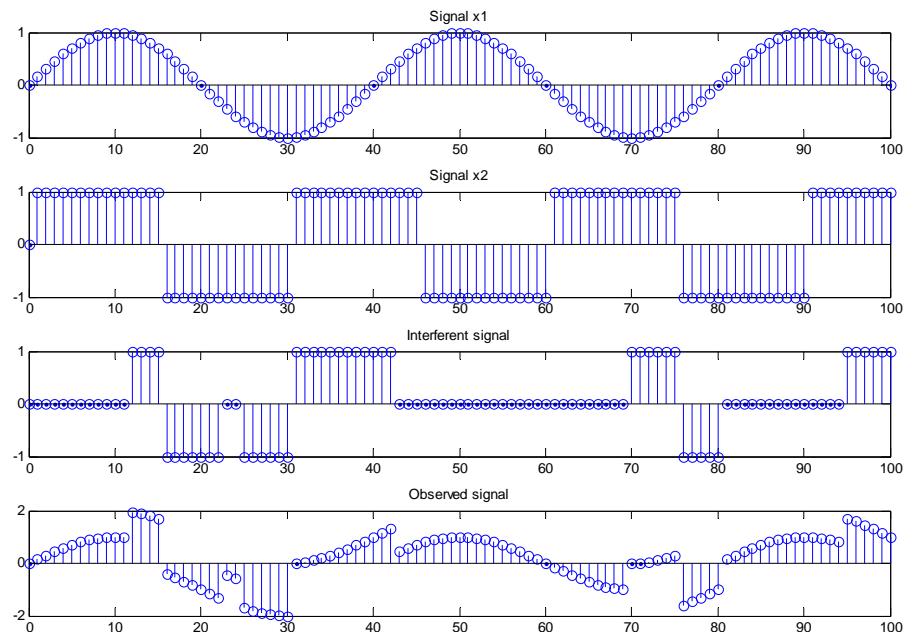


3. Components of a time series: data model

Noise is dependent of signal?



Noise is dependent of noise



$$\hat{x}[n] = x[n] + \varepsilon[n]$$

4. Descriptive analysis

- Time plot
 - Take care of appearance (scale, point shape, line shape, etc.)
 - Take care of labelling axes specifying units (be careful with the scientific notation, e.g. from 0.5e+03 to 0.1e+04)
- Data Preprocessing
 - Consider transforming the data to enhance a certain feature, although this is a controversial topic:
 - Logarithm: Stabilizes variance in the fully multiplicative model

$$y[n] = \log(x[n])$$

- Box-Cox: the transformed data tends to be normally distributed

$$y[n] = \begin{cases} \frac{x^\lambda[n]-1}{\lambda} & \lambda > 0 \\ \log(x[n]) & \lambda = 0 \end{cases}$$

4. Descriptive analysis

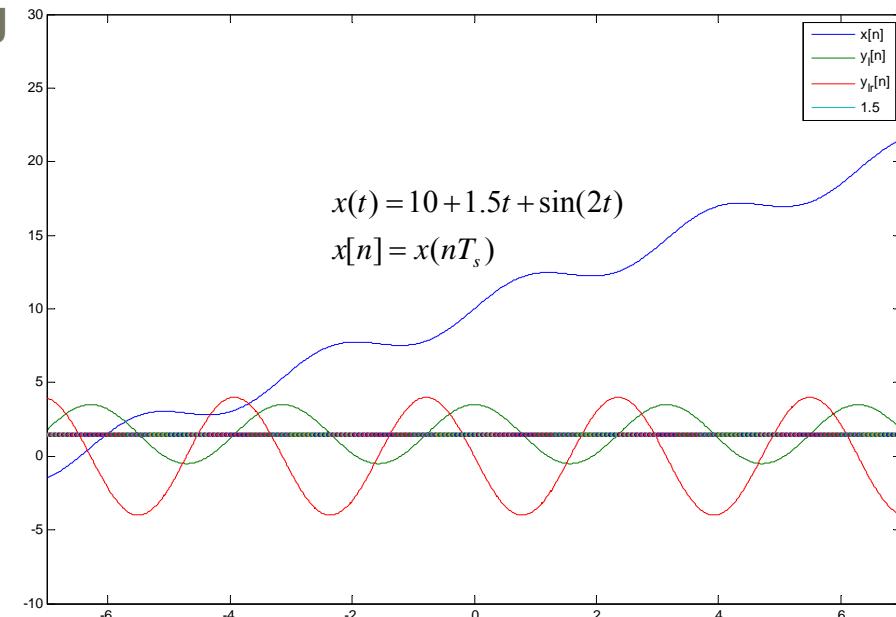
- Data preprocessing:
 - (Removal of trend) **Detrending**: if the trend clearly follows a known curve (line, polynomial, logistic, Gaussian, Gompertz, etc.), fit a model to the time series and detrend (either by subtracting or dividing).
 - (Removal of trend) **Differencing**

$$y_l[n] = \nabla_l x[n] = \frac{x[n] - x[n-1]}{T_s}$$

$$y_r[n] = \nabla_r x[n] = \frac{x[n+1] - x[n]}{T_s}$$

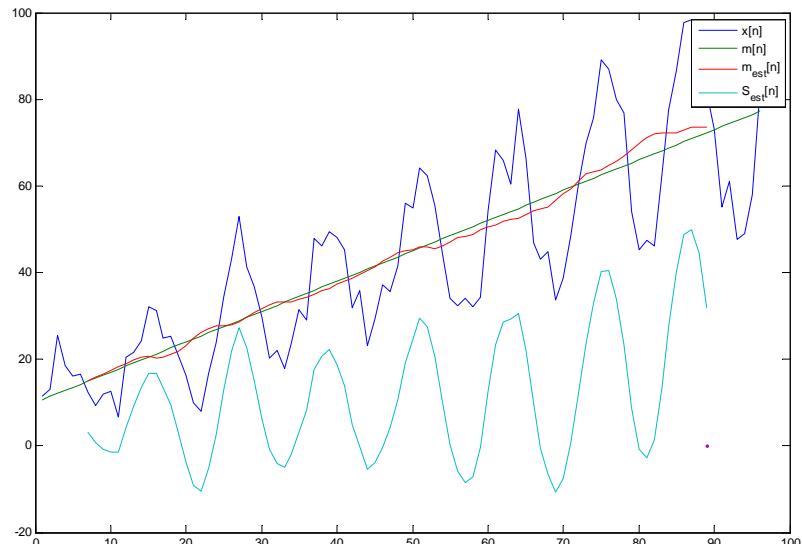
$$y_{lr}[n] = \nabla_r y_l[n] = \frac{x[n+1] - 2x[n] + x[n-1]}{T_s^2}$$

$$y_{ll}[n] = \nabla_l y_l[n] = \frac{x[n] - 2x[n-1] + x[n-2]}{T_s^2}$$



4. Descriptive analysis

- Data preprocessing:
 - (Removal of season) **Deseasoning**: average over the seasonal period to remove its effect.
 - (Estimate of season) **Estimating seasoning**: subtract the deseasoned time series from a local estimate of the current sample.



$$x[n] = m[n] + S[n] + e[n]$$

$$m_{est}[n] = \frac{1}{12} \left(\frac{1}{2} x[n-6] + x[n-5] + x[n-4] + \dots + x[n] \right. \\ \left. x[n+1] + \dots + x[n+5] + \frac{1}{2} x[n+6] \right)$$

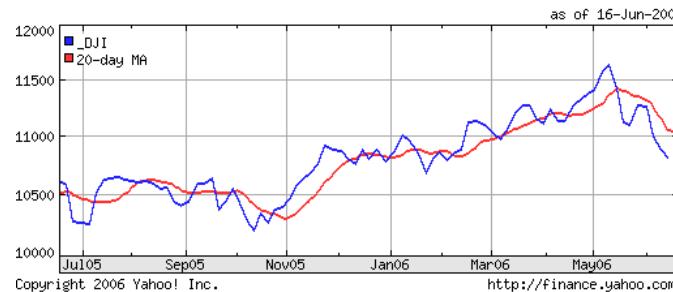
$$S_{est}[n] = \left(\sum_{k=-2}^2 \frac{1}{2^{|k|+1}} x[n-k] \right) - m_{est}[n]$$

$$S_{est}[n] = \frac{1}{8} x[n-2] + \frac{1}{4} x[n-1] + \frac{1}{2} x[n] + \frac{1}{4} x[n+1] + \frac{1}{8} x[n+2] - m_{est}[n]$$

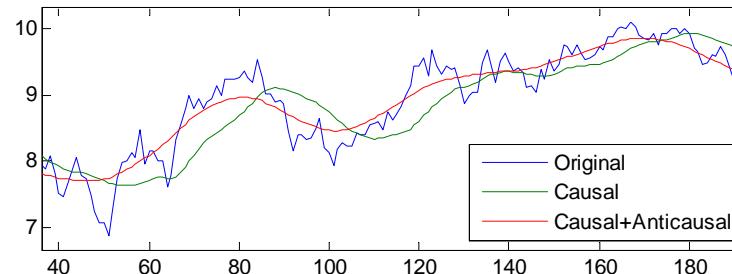
4. Descriptive analysis

- Data preprocessing:
 - (Removal or isolation of trend, seasonal or noise) **Filtering**: filtering aims at the removal of any of the components, for example, the moving average is a common filtering operation in stock analysis to remove noise.

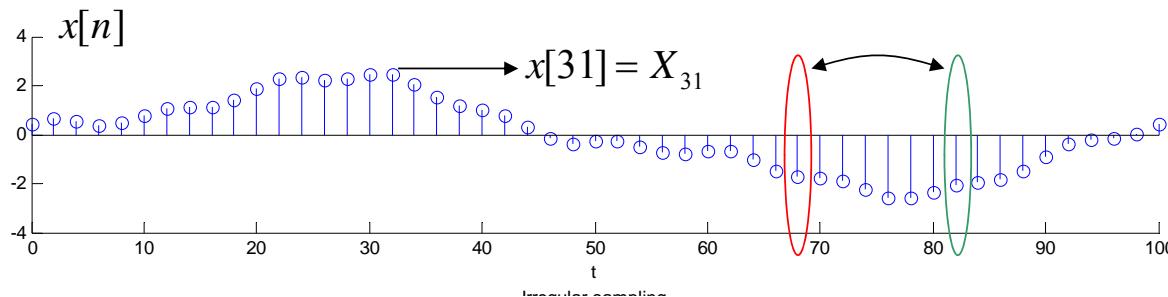
$$y[n] = \frac{1}{20} \sum_{k=1}^{20} x[n-k]$$



$$y[n] = \frac{1}{21} \sum_{k=-10}^{10} x[n-k]$$



5. Distributional properties



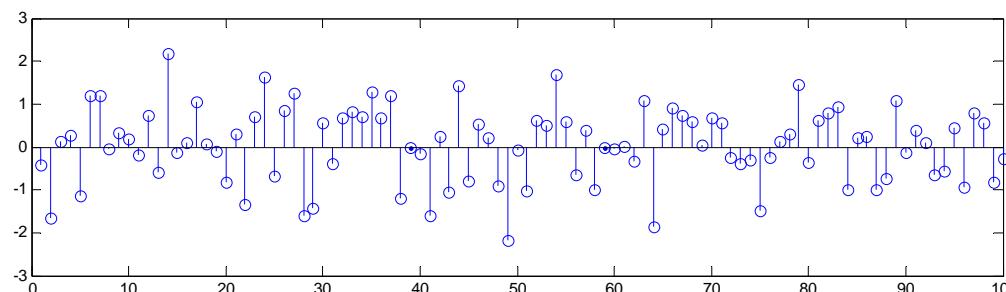
All variables are identically distributed → Stationarity

A variable can be normally distributed or Poisson or etc.

It is important to characterize their distribution

All variables are independent → White process

Independency is measured through the autocorrelation function



5. Distributional properties

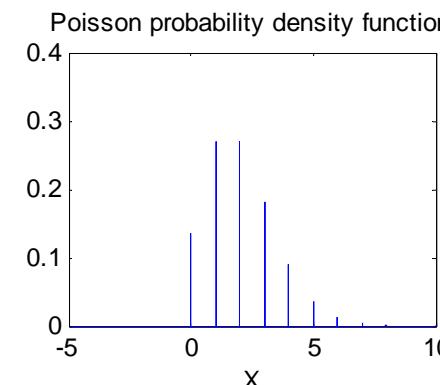
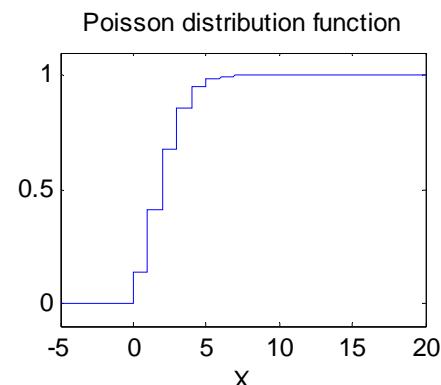
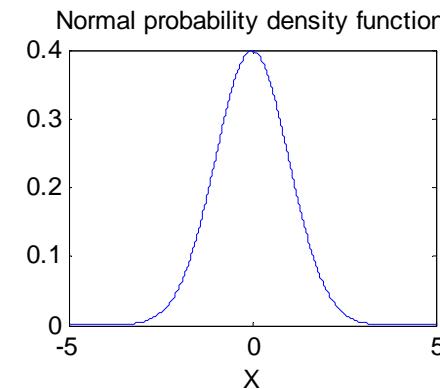
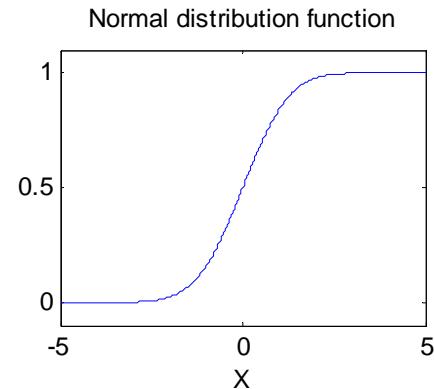
$$\begin{aligned} F_X(x) &= \Pr\{X \leq x\} \\ &= \int_{-\infty}^x f(t)dt \\ &= \sum_{x_i \leq x} \Pr\{X = x_i\} \end{aligned}$$

$$F_{X,Y}(x, y) = \Pr\{X \leq x, Y \leq y\}$$

$$\begin{aligned} E\{X^r\} &= \int x^r F_X(dx) \\ &= \int_{-\infty}^{\infty} x^r f(t)dt \\ &= \sum_{x_i \leq x} x^r \Pr\{X = x_i\} \end{aligned}$$

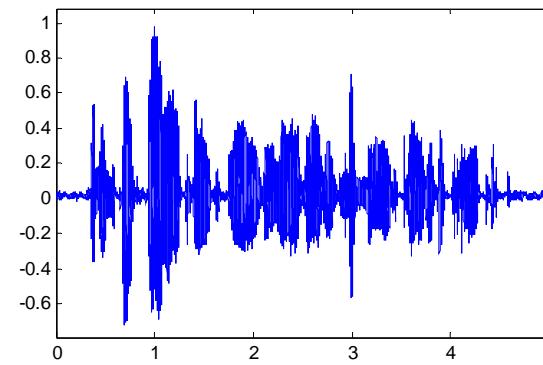
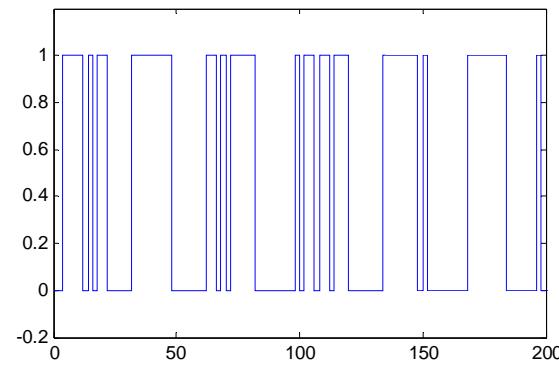
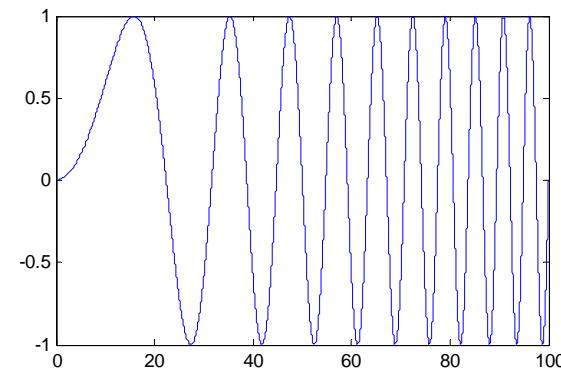
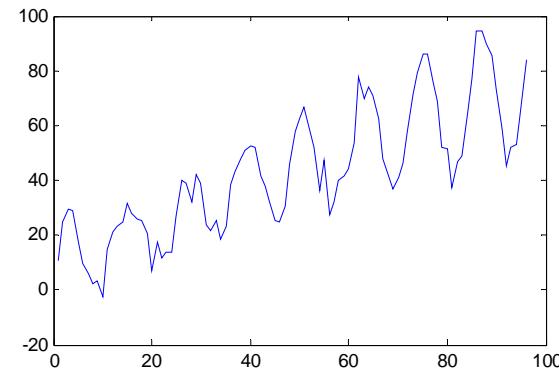
Characteristic

$$\varphi_X(t) = E\{e^{itX}\} = 1 + \sum_{k=1}^{\infty} \frac{i^k E\{X^k\}}{k!} t^k \longleftrightarrow F_X(b) - F_X(a) = \frac{1}{2\pi} \lim_{\tau \rightarrow \infty} \int_{-\tau}^{\tau} \frac{e^{-ita} - e^{-itb}}{it} \varphi_X(t) dt$$



5. Distributional properties

Examples of nonstationary variables



5. Distributional properties

Strictly stationary $F_{X_{n_1}, X_{n_2}, \dots, X_{n_k}}(x_{n_1}, x_{n_2}, \dots, x_{n_k}) = F_{X_{n_1+N}, X_{n_2+N}, \dots, X_{n_k+N}}(x_{n_1+N}, x_{n_2+N}, \dots, x_{n_k+N})$

Consequences: $E\{X_n\} = \mu \quad \forall n \quad \forall k, n_1, n_2, \dots, n_k, N$

Autocorrelation function (ACF) $\rightarrow \text{Corr}(X_n, X_{n+n_0}) = E\{X_n X_{n+n_0}^*\} = \Gamma[n_0] \quad \forall n, n_0$

$$\Gamma[n_0] = \Gamma^*[-n_0] \quad \forall n_0$$

\uparrow lag

$$\Gamma[0] \geq |\Gamma[n_0]| \quad \forall n_0$$

Example: white process $\rightarrow \Gamma[n_0] = \sigma_x^2 \delta[n_0] = \begin{cases} 1 & n_0 = 0 \\ 0 & n_0 \neq 0 \end{cases}$

Autocovariance function $\rightarrow C[n_0] = \text{Cov}(X_n, X_{n+n_0}) = E\{(X_n - \mu)(X_{n+n_0} - \mu)^*\} = \Gamma[n_0] - |\mu|^2$

Correlation coefficient $\rightarrow r[n_0] = \frac{C[n_0]}{C[0]}$

Wide sense stationary $E\{X_n\} = \mu \quad \forall n$

$$\text{Corr}(X_n, X_{n+n_0}) = \Gamma[n_0] \quad \forall n, n_0$$

5. Distributional properties

Correlation coefficients (Zero-order correlations)

$$r[n_0] = \frac{C[n_0]}{C[0]}$$

Under the null hypothesis ($H_0 \equiv r[n_0] = 0$) it is distributed as Student's t with N-2 degrees of freedom.



Number of samples upon which $r[n_0]$ is estimated

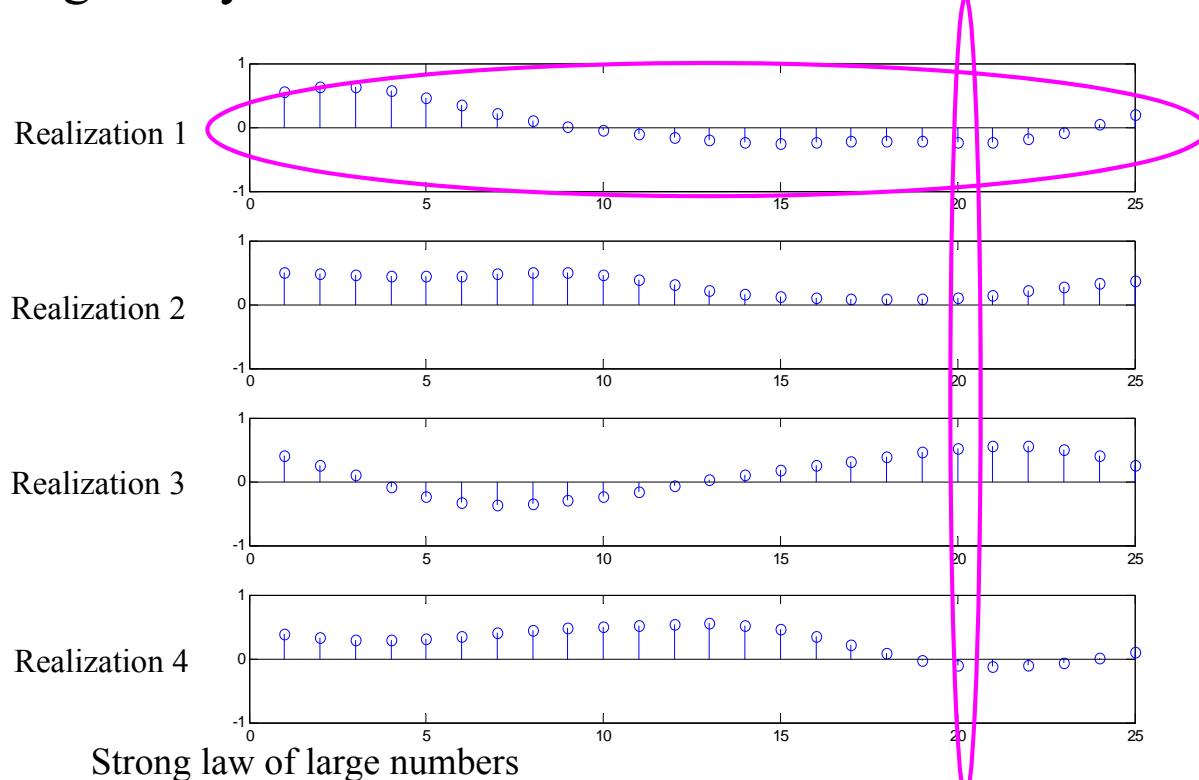
The probability of observing $r[n_0]$ if the null hypothesis is true is given by

$$\Pr\left\{ |t| \geq r[n_0] \sqrt{\frac{N-2}{1-r^2[n_0]}} \right\}$$



5. Distributional properties

Ergodicity



$$\hat{\mu} = \frac{1}{25} \sum_{n=1}^{25} x[n]$$

Time average

Equal if
stationary and
ergodic for the
mean

$$\mu_{20} = \lim_{N \rightarrow \infty} \frac{1}{N} \sum_{i=1}^N x_{20,i} \quad \longleftrightarrow \quad \hat{\mu}_{20} = \frac{1}{4} \sum_{i=1}^4 x_{20,i}$$

Ensemble average

5. Distributional properties

Stationarity $\not\Rightarrow$ Ergodicity

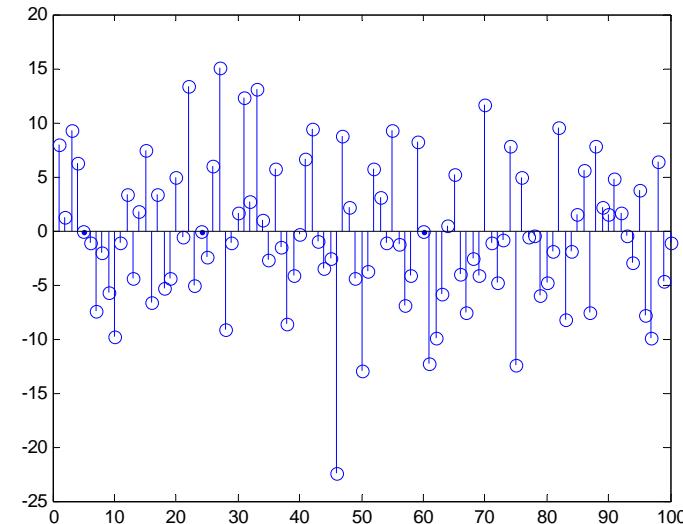
Example:

$$x[n] = m[n] + e[n]$$

$N(0, \sigma_\mu^2)$ $N(0, \sigma_\varepsilon^2)$

$$E\{X_n\} = 0$$

$$\text{Cov}(X_n, X_{n+n_0}) = \sigma_\mu^2 + \sigma_\varepsilon^2$$



The time average still converges to the ensemble average

5. Distributional properties

How to detect non-stationarity in the mean?

There is a variation in the local mean

Solutions:

Polynomial trends: By differentiating p times

$$x^{(1)}[n] = x[n] - x[n-1] \longrightarrow \text{Removal of a constant (or piece-wise constant)}$$



$$x^{(2)}[n] = x^{(1)}[n] - x^{(1)}[n-1] \longrightarrow \text{Removal of a linear trend}$$



$$x^{(3)}[n] = x^{(2)}[n] - x^{(2)}[n-1] \longrightarrow \text{Removal of a quadratic trend}$$

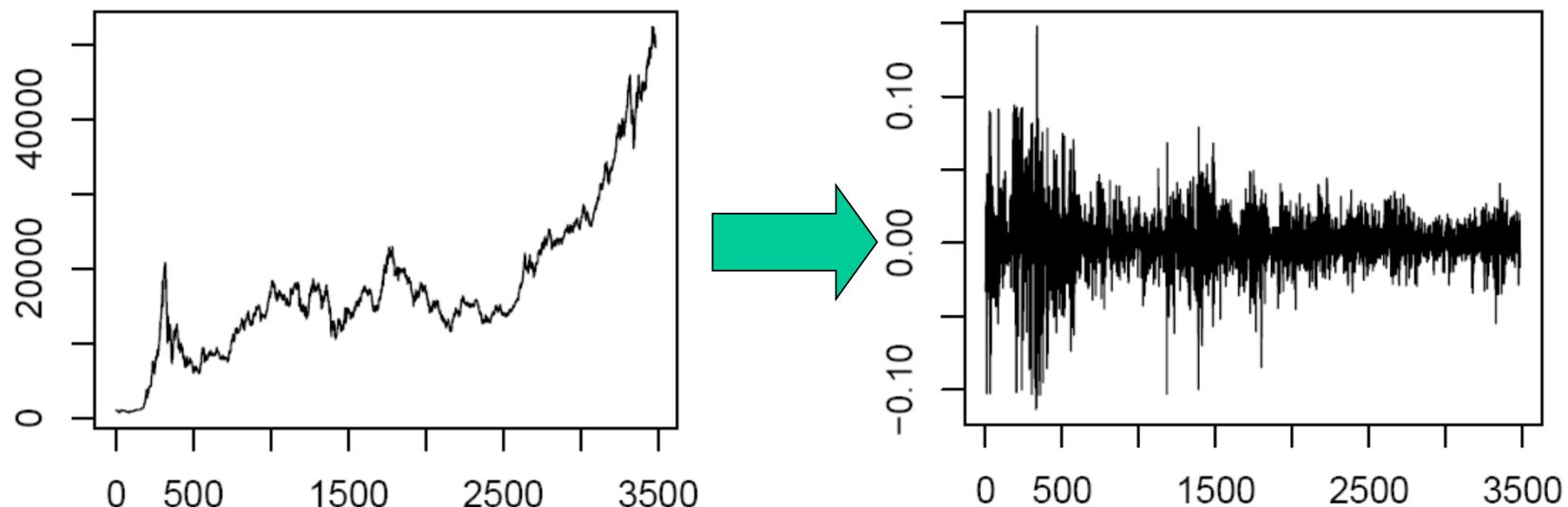
...

$$x^{(p)}[n] = x^{(p-1)}[n] - x^{(p-1)}[n-1] \longrightarrow \text{Removal of a p-th order polynomial trend}$$

5. Distributional properties

How to detect non-stationarity in the mean?

Exponential trends: Taking logs \longrightarrow Financial time series

$$y[n] = \log \frac{x[n]}{x[n-1]}$$


5. Distributional properties

How to detect non-stationarity due to seasonality?

There is a periodic variation in the local mean

Solutions:

Polynomial trends: By differentiating p times with that seasonality

$$x^{(1)}[n] = x[n] - x[n-12] \longrightarrow \text{Monthly data: removal of a yearly seasonality}$$

$$x^{(1)}[n] = x[n] - x[n-7] \longrightarrow \text{Daily data: removal of a weekly seasonality}$$

How to detect non-stationarity in the variance?

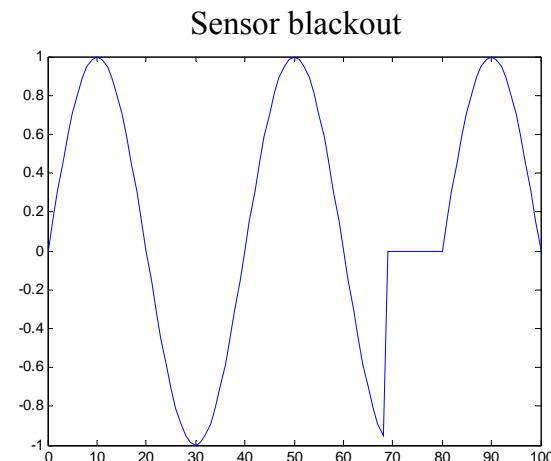
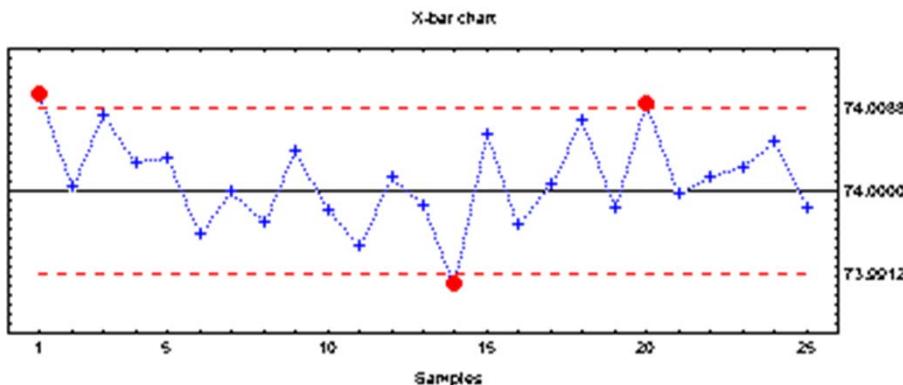
There is a variation in the local variance

Solutions: Box-Cox transformation tends to stabilize variance

How to detect non-stationarity?

Solutions: Unit root tests

6. Outlier detection and rejection



- Common procedures:
 1. Visual inspection
 2. $\text{Mean} \pm k \text{ Standard Deviation}$
 3. $\text{Median} \pm k \text{ Median Abs. Deviation}$
 4. Robust detection
 5. Robust model fitting
- Common actions:
 1. Remove observation
 2. Substitute by an estimate

→ { Do

- Estimate mean and Std. Dev.
- Remove samples outside a given interval

Until (No sample is removed)

7. Time series methods

- Time-domain methods:
 - Based on classical theory of correlation (autocovariance)
 - Include parametric methods (AR, MA, ARMA, ARIMA, etc.)
 - Include regression (linear, nonlinear)
- Frequency-domain (spectral) methods:
 - Based on Fourier analysis
 - Include harmonic methods
- Neural networks and Fuzzy neural networks
- Other fancy approaches: fractal models, wavelet models, bayesian networks, etc.

Bibliography

- C. Chatfield. *The analysis of time series: an introduction*. Chapman & Hall, CRC, 1996.
- D.S.G. Pollock. *A handbook of time-series analysis, signal processing and dynamics*. Academics Press, 1999.
- J. D. Hamilton. *Time series analysis*. Princeton Univ. Press, 1994.
- A. V. Oppenheim, R. W. Schafer, J. R. Buck. *Discrete-time signal processing, 2nd edition*. Prentice Hall, 1999.
- A. Papoulis, S. U. Pillai. *Probability, random variables and stochastic processes, 4th edition*. McGraw Hill, 2002.



CEU
*Universidad
San Pablo*

Time Series Analysis

Session II: Regression and Harmonic Analysis

Carlos Óscar Sánchez Sorzano, Ph.D.
Madrid

Session outline

1. Goal
2. Linear and non-linear regression
3. Polynomial fitting
4. Cubic spline fitting
5. A short introduction to system analysis
6. Spectral representation of stationary processes
7. Detrending and filtering
8. Non-stationary processes

1. Goal

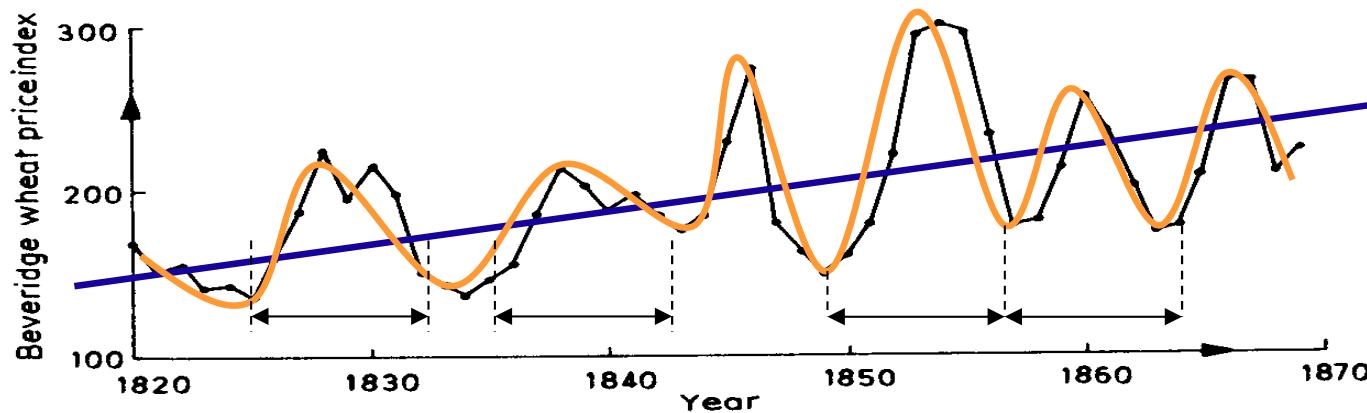


Figure 1.1 Part of the Beveridge wheat price index series.

$$x[n] = \underbrace{\text{trend}[n] + \text{periodic}[n]}_{\text{Session 2}} + \text{random}[n]$$

Session 3

2. Linear and non linear regression

Linear regression

Year (n)	Price (x[n])
1500	17
1501	19
1502	20
1503	15
1504	13
1505	14
1506	14
1507	14
...	...

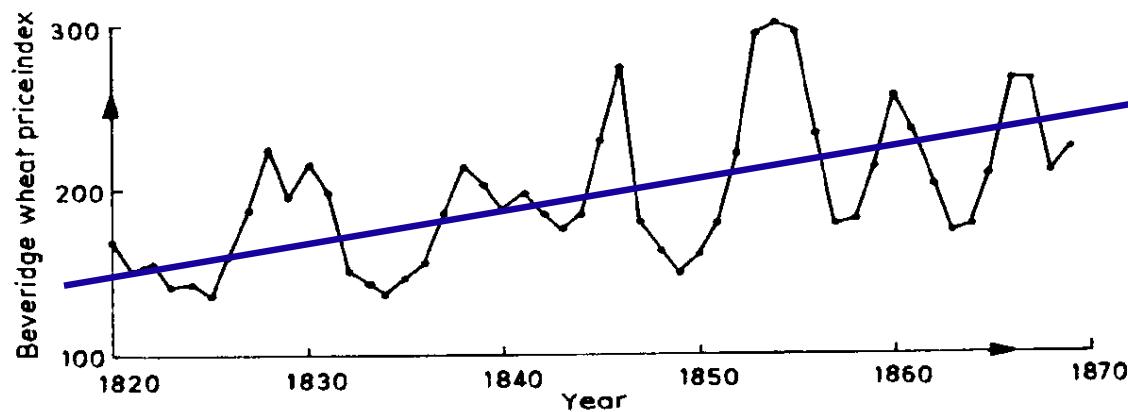


Figure 1.1 Part of the Beveridge wheat price index series.

$$\begin{aligned} & \left\{ \begin{array}{l} x[n] = \text{trend}[n] + \text{random}[n] \\ \text{trend}[n] = \beta_0 + \beta_1 n \end{array} \right\} \rightarrow \left\{ \begin{array}{l} 17 = \beta_0 + 1500\beta_1 \\ 19 = \beta_0 + 1501\beta_1 \\ 20 = \beta_0 + 1502\beta_1 \\ 15 = \beta_0 + 1503\beta_1 \\ \dots \end{array} \right\} \rightarrow \begin{pmatrix} 1 & 1500 \\ 1 & 1501 \\ 1 & 1502 \\ 1 & 1503 \\ \dots & \dots \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \begin{pmatrix} 17 \\ 19 \\ 20 \\ 15 \\ \dots \end{pmatrix} \\ & \boxed{\mathbf{X}\boldsymbol{\beta} = \mathbf{x}} \end{aligned}$$

2. Linear and non linear regression

Linear regression

$$x[n] = trend[n] + random[n] \xrightarrow{\quad} \mathbf{X}\boldsymbol{\beta} = \mathbf{x}$$

↓ Least Squares Estimate

$$\hat{\boldsymbol{\beta}} = \arg \min_{\boldsymbol{\beta}} \sigma_{\varepsilon}^2 = \mathbf{X}^+ \mathbf{x} = (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{X}^t \mathbf{x}$$

Let's assume $E\{\varepsilon[n]\} = 0$
 $\Gamma_{\varepsilon}[n_0] = \sigma_{\varepsilon}^2 \delta[n_0]$
 Homocedasticity

Properties: $E\{\hat{\boldsymbol{\beta}}\} = \boldsymbol{\beta}$

$$Cov\{\hat{\boldsymbol{\beta}}\} = \sigma_{\varepsilon}^2 (\mathbf{X}^t \mathbf{X})^{-1}$$

$$\hat{\sigma}_{\varepsilon}^2 = \frac{1}{N-k} (\mathbf{X}\boldsymbol{\beta} - \mathbf{x})^t (\mathbf{X}\boldsymbol{\beta} - \mathbf{x})$$

Degree of fit

$$R^2 = 1 - \frac{\|\mathbf{x} - \mathbf{X}\boldsymbol{\beta}\|^2}{\|\mathbf{x} - \bar{\mathbf{x}}\mathbf{l}\|^2}$$

$$R_{adjusted}^2 = 1 - (1 - R^2) \frac{N-1}{N-k-1}$$

Linear regression with constraints

$$\hat{\boldsymbol{\beta}}_R = \arg \min_{\boldsymbol{\beta}} \sigma_{\varepsilon}^2 \quad s.t. \quad \mathbf{R}\boldsymbol{\beta} = \mathbf{r} = \hat{\boldsymbol{\beta}} + (\mathbf{X}^t \mathbf{X})^{-1} \mathbf{R}^t \left(\mathbf{R}(\mathbf{X}^t \mathbf{X})^{-1} \mathbf{R}^t \right) (\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}})$$

Example:
 $\beta_2 = 2\beta_1 \Rightarrow 2\beta_1 - \beta_2 = 0$

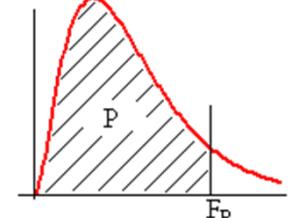
2. Linear and non linear regression

First test:

$$\begin{cases} H_0 \equiv \beta = \beta_0 \\ H_1 \equiv \beta \neq \beta_0 \end{cases}$$

$$\rightarrow F = \frac{1}{\hat{\sigma}_\varepsilon^2 k} (\hat{\beta} - \beta_0)^t \mathbf{X}^t \mathbf{X} (\hat{\beta} - \beta_0)$$

$$\xrightarrow{H_0 \text{ is true}} F(k, N-k)$$



If $F > F_p$, reject H_0

Let's assume $\varepsilon[n] \rightarrow N(0, \sigma_\varepsilon^2)$

$$\Gamma_\varepsilon[n_0] = \sigma_\varepsilon^2 \delta[n_0]$$

$$\begin{aligned} x[n] &= \text{trend}[n] + \text{random}[n] \\ \text{trend}[n] &= \beta_0 + \beta_1 n + \beta_2 n^2 \end{aligned}$$

$$\left. \begin{array}{l} 17 = \beta_0 + 1500\beta_1 + 1500^2\beta_2 \\ 19 = \beta_0 + 1501\beta_1 + 1501^2\beta_2 \\ 20 = \beta_0 + 1502\beta_1 + 1502^2\beta_2 \\ 15 = \beta_0 + 1503\beta_1 + 1503^2\beta_2 \\ \dots \end{array} \right\} \rightarrow$$

$$\begin{pmatrix} 1 & 1500 & 1500^2 \\ 1 & 1501 & 1501^2 \\ 1 & 1502 & 1502^2 \\ 1 & 1503 & 1503^2 \\ \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 17 \\ 19 \\ 20 \\ 15 \\ \dots \end{pmatrix}$$

$$\mathbf{X}\beta = \mathbf{x}$$

$$\mathbf{P}_1 = \mathbf{X}_1 (\mathbf{X}_1^t \mathbf{X}_1)^{-1} \mathbf{X}_1^t \quad \begin{pmatrix} \mathbf{X}_1 & \mathbf{X}_2 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \beta_2 \end{pmatrix} = \mathbf{x}$$

Second test: $H_0 \equiv \beta_2 = \beta_{20}$

$$\begin{cases} H_0 \equiv \beta_2 = \beta_{20} \\ H_1 \equiv \beta_2 \neq \beta_{20} \end{cases}$$

$$\rightarrow F = \frac{1}{\hat{\sigma}_\varepsilon^2 k_2} (\hat{\beta}_2 - \beta_{20})^t \mathbf{X}_2^t (\mathbf{I} - \mathbf{P}_1) \mathbf{X}_2 (\hat{\beta}_2 - \beta_{20}) \xrightarrow{H_0 \text{ is true}} F(k_2, N-k)$$

$$\begin{cases} H_0 \equiv \beta_i = \beta_{i0} \\ H_1 \equiv \beta_i \neq \beta_{i0} \end{cases}$$

$$\rightarrow t = \frac{\hat{\beta}_i - \beta_{i0}}{\hat{\sigma}_\varepsilon \sqrt{\omega_{ii}}}$$

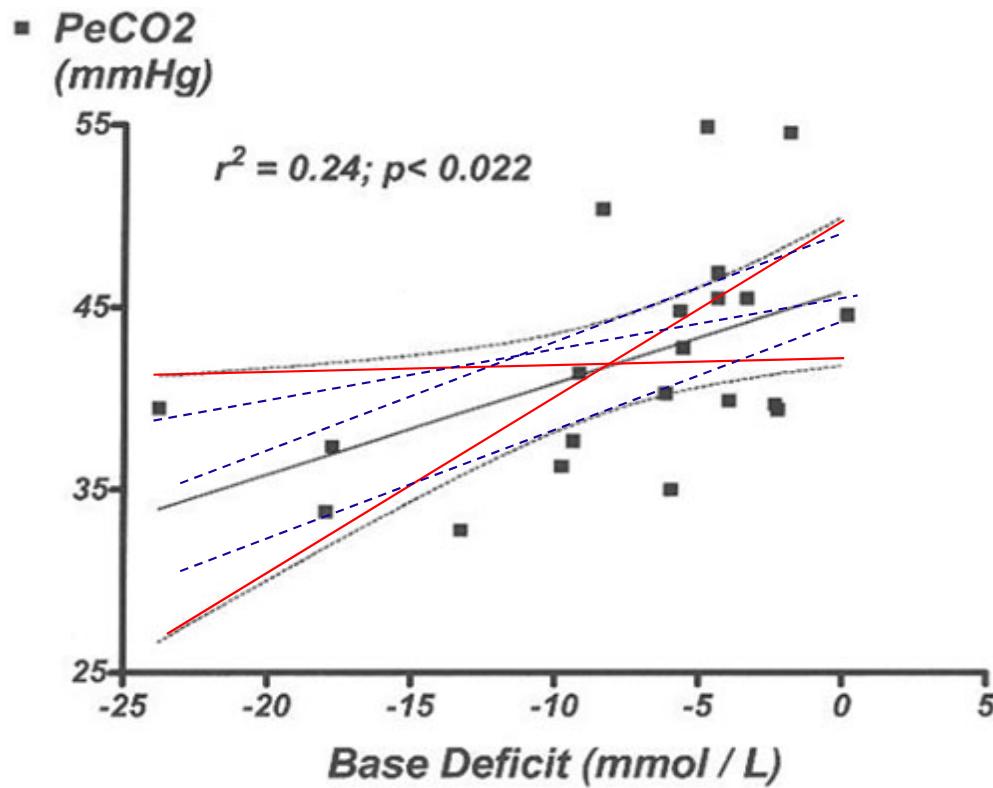
$$\left[\omega_{ii} = (\mathbf{X}^t \mathbf{X})^{-1}_{ii} \right] \xrightarrow{H_0 \text{ is true}} t(N-k)$$

2. Linear and non linear regression

Confidence intervals for the coefficients

$$\hat{\sigma}_j^2 = \hat{\sigma}^2 (\mathbf{X}^t \mathbf{X})_{jj}^{-1} \leftarrow \text{Unbiased variance of the j-th regression coefficient}$$

$$\beta_j \in \hat{\beta}_j + t_{1-\frac{\alpha}{2}, N-k-1} \hat{\sigma}_j^2 \leftarrow \text{Confidence interval for the j-th regression coefficient}$$

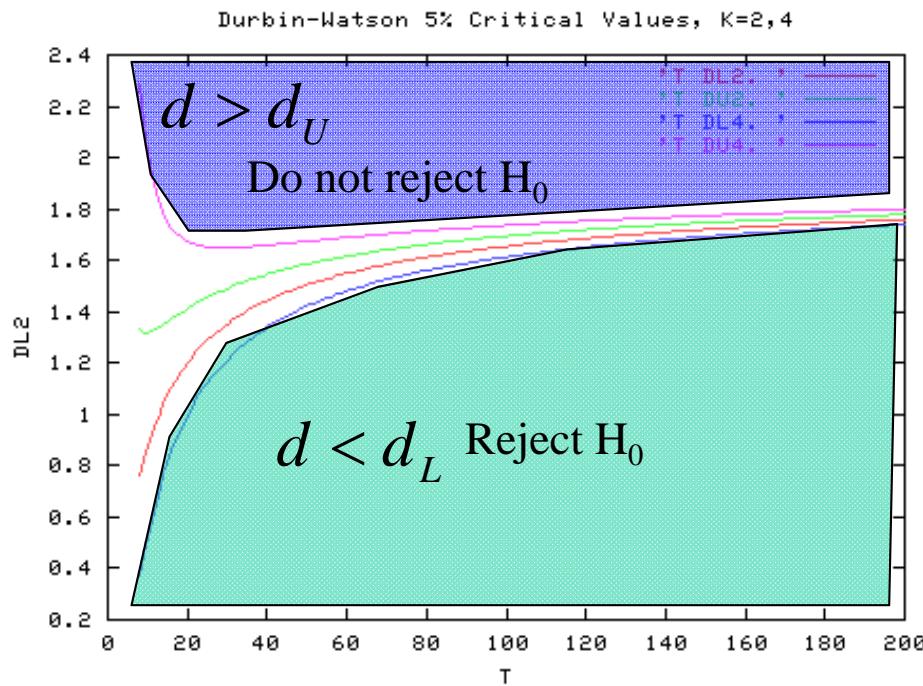


We got a certain regression line but the true regression line lies within this region with a 95% confidence.

2. Linear and non linear regression

Test that the residuals are certainly uncorrelated

Durbin-Watson test: $H_0 \equiv \Gamma[n_0] = 0$ $\forall n_0 \neq 0$ $H_1 \equiv \Gamma[n_0] > 0$ $d = \frac{\sum_{n=2}^N (\varepsilon[n] - \varepsilon[n-1])^2}{\sum_{n=1}^N \varepsilon^2[n]}$



Other tests: Durbin's h, Wallis' D₄, Von Neumann's ratio, Breusch-Godfrey

2. Linear and non linear regression

Cochrane-Orcutt method of regression with correlated residues

1. Estimate a first model
2. Estimate residuals (correlated)
3. Estimate the correlation of the residues
4. $i=1$
5. Estimate “uncorrelated” residuals
6. Estimate “uncorrelated” output
7. Estimate “uncorrelated” input
8. Reestimate model
9. Estimate residuals
10. $i=i+1$ until convergence in $\beta^{(i)}$

$$\begin{aligned}\hat{y}^{(0)}[n] &= \boldsymbol{\beta}^{(0)} \mathbf{x}[n] \\ \varepsilon^{(0)}[n] &= y[n] - \hat{y}^{(0)}[n] \\ \hat{\Phi}_\varepsilon &\\ a^{(i)}[n] &= \varepsilon^{(i)}[n] - \hat{\Phi}_\varepsilon[i-1] \varepsilon^{(i)}[n-1] \\ y^{(i)}[n] &= y[n] - \hat{\Phi}_\varepsilon[i-1] y^{(i)}[n-1] \\ \mathbf{x}^{(i)}[n] &= \mathbf{x}[n] - \hat{\Phi}_\varepsilon[i-1] \mathbf{x}^{(i)}[n-1] \\ \hat{y}^{(i)}[n] - a^{(i)}[n] &= \boldsymbol{\beta}^{(i)} \mathbf{x}^{(i)}[n] \\ \varepsilon^{(i)}[n] &= y^{(i)}[n] - \hat{y}^{(i)}[n]\end{aligned}$$

2. Linear and non linear regression

Assumptions of regression

- The sample is representative of your population
- The dependent variable is noisy, but the predictors are not!! **Solution:** Total Least Squares
- Predictors are linearly independent (i.e., no predictor can be expressed as a linear combination of the rest), although they can be correlated. If it happens, this is called multicollinearity. **Solution:** add more samples, remove dependent variable, PCA
- The errors are homoscedastic. **Solution:** Weighted Least Squares
- The errors are uncorrelated to the predictors and to itself. **Solution:** Generalized Least Squares
- The errors follow a normal distribution. **Solution:** Generalized Linear Models

2. Linear and non linear regression

More linear regression

$$x[n] = \beta_0 + \beta_1 n + w[n]$$

$$x[n] = \beta_0 + \beta_1 n + \beta_2 x[n-1] + w[n]$$

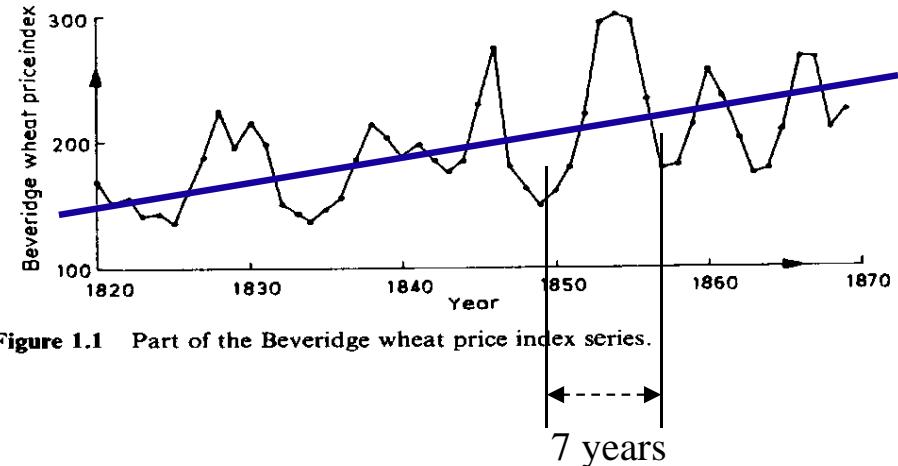
$$x[n] = \beta_0 + \beta_1 n + \beta_2 x[n-1] + \beta_3 x[n-2] + w[n]$$

$$x[n] = \beta_0 + \beta_1 n + \beta_2 (x[n-1] - x[n-2]) + w[n]$$

$$x[n] = \beta_0 + \beta_1 n + \alpha_0 M_0[n] + \alpha_1 M_1[n] + \dots + \alpha_6 M_6[n] + w[n]$$

$$\uparrow M_0[n] = \begin{cases} 1 & n = \dot{\gamma} + 0 \\ 0 & \text{otherwise} \end{cases}$$

$$\uparrow M_6[n] = \begin{cases} 1 & n = \dot{\gamma} + 6 \\ 0 & \text{otherwise} \end{cases}$$



Non linear regression

$$x[n] = \beta_0 + \beta_1 n + \beta_2 \sin(\alpha_0 n + \alpha_1) + w[n]$$

$$x[n] = \beta_0 n^{\beta_1} w[n] \longrightarrow \log x[n] = \log \beta_0 + \beta_1 \log n + \log w[n] \longrightarrow x'[n] = \beta'_0 + \beta'_1 \log n + w'[n]$$

3. Polynomial fitting

Polynomial trends

$$\begin{aligned}
 x[n] &= trend[n] + random[n] \\
 trend[n] &= \beta_0 + \beta_1 n + \beta_2 n^2
 \end{aligned}
 \quad \left\{ \rightarrow \begin{array}{l}
 17 = \beta_0 + 1500\beta_1 + 1500^2 \beta_2 \\
 19 = \beta_0 + 1501\beta_1 + 1501^2 \beta_2 \\
 20 = \beta_0 + 1502\beta_1 + 1502^2 \beta_2 \\
 15 = \beta_0 + 1503\beta_1 + 1503^2 \beta_2 \\
 ...
 \end{array} \right. \rightarrow \begin{pmatrix} 1 & 1500 & 1500^2 \\ 1 & 1501 & 1501^2 \\ 1 & 1502 & 1502^2 \\ 1 & 1503 & 1503^2 \\ \dots & \dots & \dots \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \\ \beta_2 \end{pmatrix} = \begin{pmatrix} 17 \\ 19 \\ 20 \\ 15 \\ \dots \end{pmatrix} \downarrow \boxed{\mathbf{X}\boldsymbol{\beta} = \mathbf{x}}
 \end{math>$$

Orthogonal polynomials

$$17 = \phi_0(1500)\alpha_0 + \phi_1(1500)\alpha_1 + \phi_2(1500)\alpha_2$$

$$19 = \phi_0(1501)\alpha_0 + \phi_1(1501)\alpha_1 + \phi_2(1501)\alpha_2$$

$$20 = \phi_0(1502)\alpha_0 + \phi_1(1502)\alpha_1 + \phi_2(1502)\alpha_2$$

$$15 = \phi_0(1503)\alpha_0 + \phi_1(1503)\alpha_1 + \phi_2(1503)\alpha_2$$

...

$$trend[n] = \beta_0 + \beta_1 n + \beta_2 n^2 + \dots + \beta_q n^q$$

$$trend[n] = \alpha_0 \phi_0(n) + \alpha_1 \phi_1(n) + \alpha_2 \phi_2(n) + \dots + \alpha_q \phi_q(n)$$

Such that $i \neq j \Rightarrow \int_{-\infty}^{\infty} \phi_i(t)\phi_j(t)dt = 0$

$$\left. \begin{array}{l}
 \mathbf{X}\boldsymbol{\beta} = \mathbf{x} \\
 \mathbf{\Phi}\mathbf{a} = \mathbf{x}
 \end{array} \right\} \xrightarrow{\mathbf{\Phi} = \mathbf{X}\mathbf{R}^{-1}} \boxed{\boldsymbol{\beta} = \mathbf{R}^{-1}\mathbf{a}}$$

3. Polynomial fitting

Polynomial trends

Orthogonal polynomials $trend[n] = \alpha_0\phi_0(n) + \alpha_1\phi_1(n) + \alpha_2\phi_2(n) + \dots + \alpha_q\phi_q(n)$

$$\phi_0(t) = 1$$

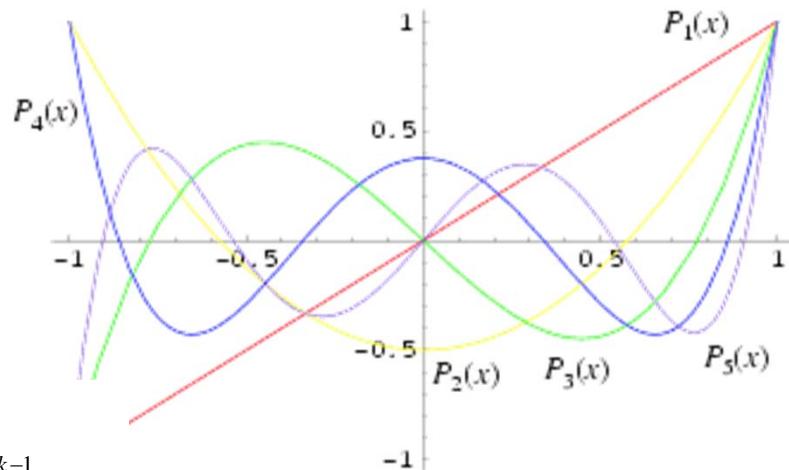
$$\phi_1(t) = t$$

$$\phi_i(t) = \frac{2i+1}{i+1}t\phi_{i-1}(t) - \frac{i}{i+1}\phi_{i-2}(t)$$

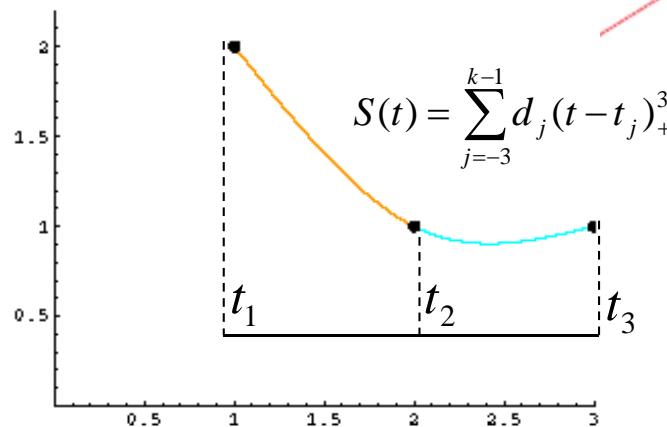
$$\phi_2(t) = \frac{3}{2}t^2 - \frac{1}{2}$$

$$\phi_3(t) = \frac{5}{2}t^3 - \frac{3}{2}t$$

$$i \neq j \Rightarrow \int_{-1}^1 \phi_i(t)\phi_j(t)dt = 0$$



Grafted polynomials

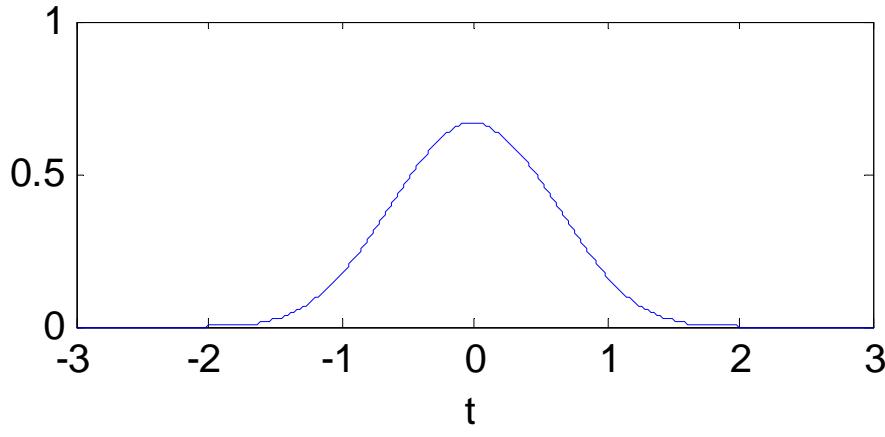


$S(t)$ is a cubic spline in some interval if this interval can be decomposed in a set of subintervals such that $S(t)$ is a polynomial of degree at most 3 on each of these subintervals and the first and second derivatives of $S(t)$ are continuous.

4. Cubic spline fitting

Cubic B-splines

$$S(t) = \sum_{j=-3}^{k-1} d_j B_j^3(t)$$



$$B_{-2}^3(t) = \begin{cases} \frac{2}{3} - t^2 + \frac{|t|^3}{2} & |t| < 1 \\ \frac{(2-|t|)^3}{6} & 1 \leq |t| < 2 \\ 0 & \text{otherwise} \end{cases}$$

$$(t_p \quad t_{p+1} \quad t_{p+2} \quad t_{p+3} \quad t_{p+4}) = (-2 \quad -1 \quad 0 \quad 1 \quad 2)$$

$$B_p^3(t) = \sum_{j=p}^q d_j (t - t_j)_+^3 = \sum_{j=p}^q d_j (t^3 - 3t^2 t_j + 3t t_j^2 - t_j^3)_+$$

$t < t_p \Rightarrow B_p(t) = 0$ By definition

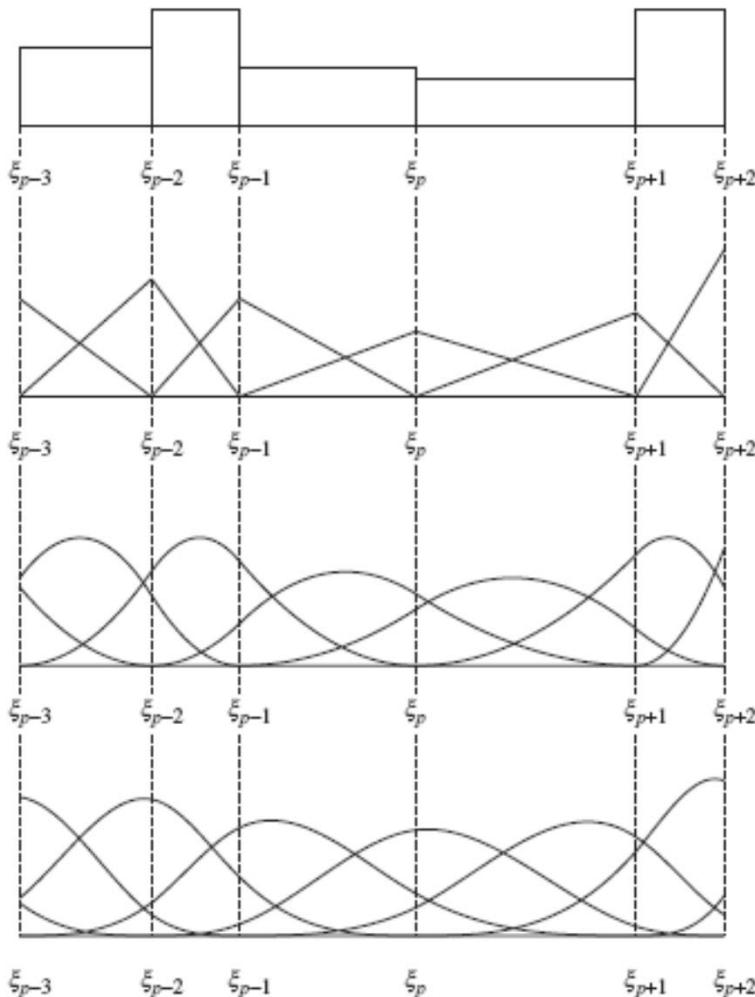
$$t > t_q \Rightarrow B_p(t) = 0 \Rightarrow \sum_{j=p}^q d_j t_j^k = 0 \quad k = 0, 1, 2, 3$$

$$\int_{-\infty}^{\infty} B_p^3(t) dt = 1$$

$$\begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ t_p & t_{p+1} & t_{p+2} & t_{p+3} & t_{p+4} \\ t_p^2 & t_{p+1}^2 & t_{p+2}^2 & t_{p+3}^2 & t_{p+4}^2 \\ t_p^3 & t_{p+1}^3 & t_{p+2}^3 & t_{p+3}^3 & t_{p+4}^3 \end{pmatrix} \begin{pmatrix} d_p \\ d_{p+1} \\ d_{p+2} \\ d_{p+3} \\ d_{p+4} \end{pmatrix} = \begin{pmatrix} 0 \\ 0 \\ 0 \\ 0 \end{pmatrix}$$

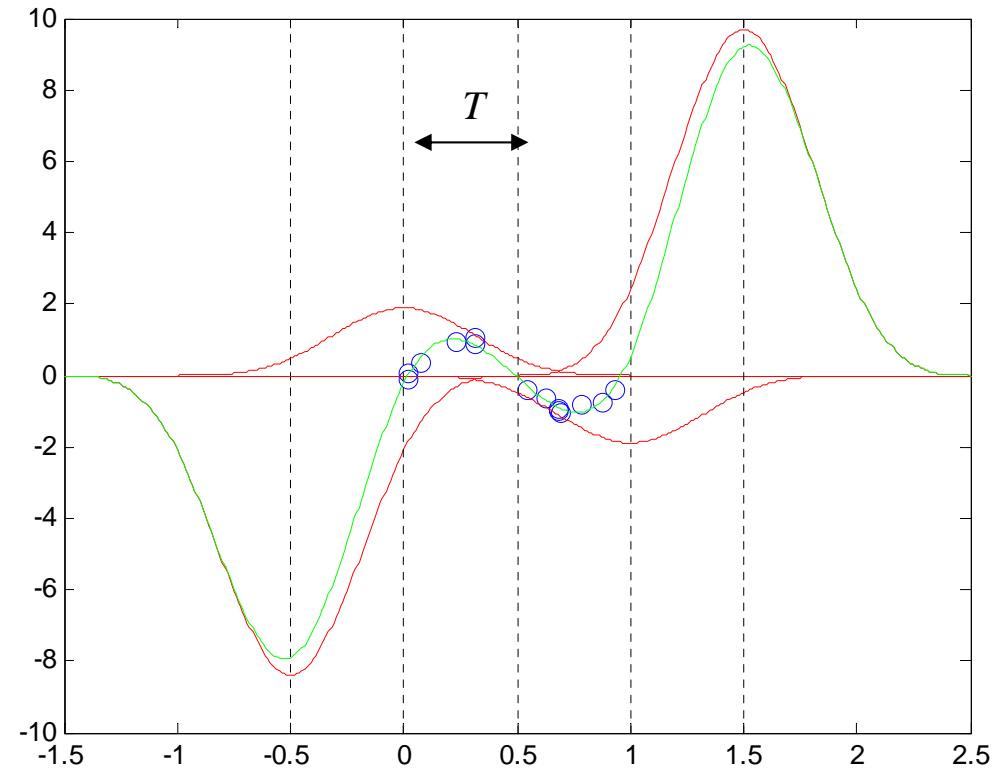
4. Cubic spline fitting

Cubic B-splines



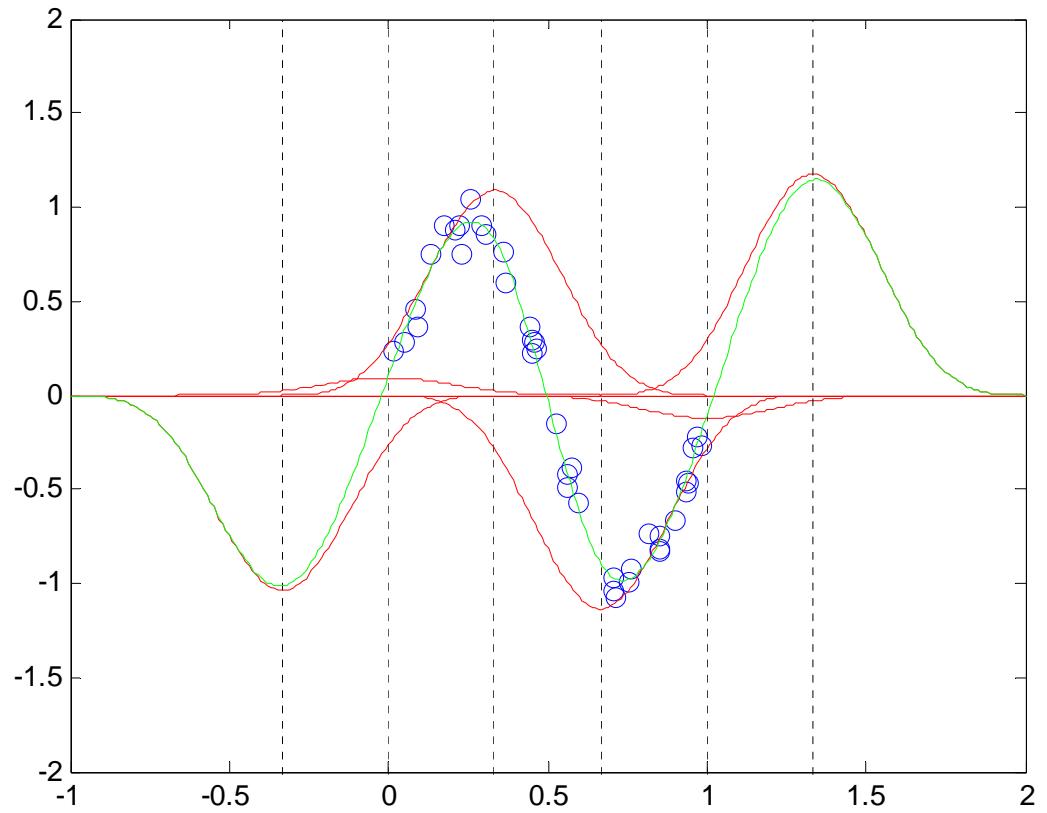
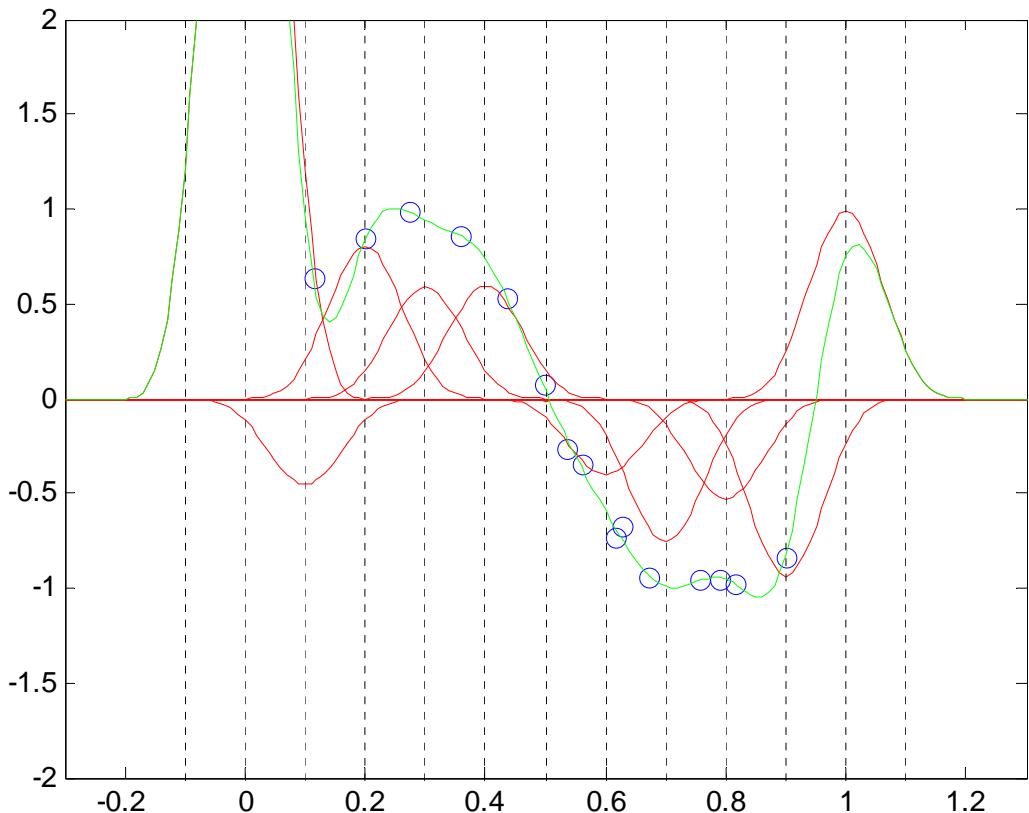
$$S(t) = \sum_{j=-3}^{k-1} d_j B_j^3(t) \xrightarrow{j_0 = \frac{t_0}{T}, j_F = \frac{t_F}{T}} S(t) = \sum_{j=j_0-1}^{j_F+1} d_j B^3\left(\frac{t}{T} - j\right)$$

S = Bd

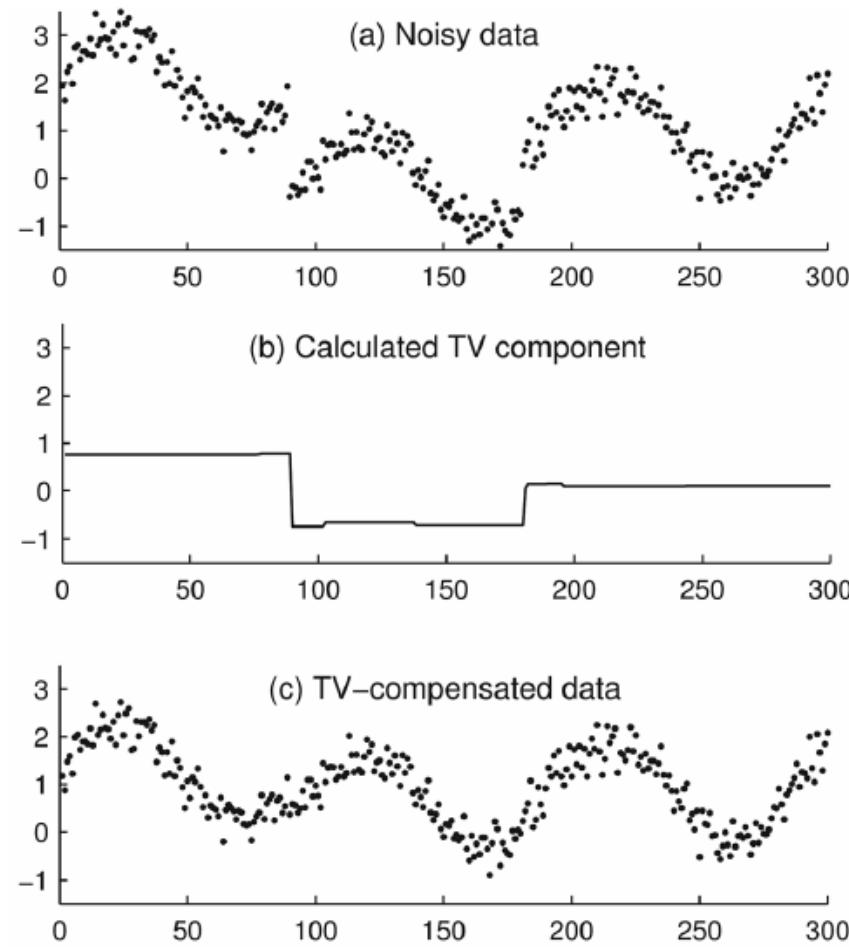


4. Cubic spline fitting

Overfitting and forecasting



4. Polynomial fitting with discontinuities



5. A short introduction to system analysis



Example: $y[n] = 3x[n]$

$$y[n] = x[n - 3]$$

$$y[n] = x^2[n]$$

$$y[n] = \frac{x[n-1] + x[n] + x[n+1]}{3}$$

$$y[n] = T_\lambda(x) = \begin{cases} \frac{x^\lambda[n]-1}{\lambda} & \lambda > 0 \\ \log(x[n]) & \lambda = 0 \end{cases}$$

$$S_{est}[n] = \left(\sum_{k=-2}^2 \frac{1}{2^{|k|+1}} x[n-k] \right) - m_{est}[n]$$

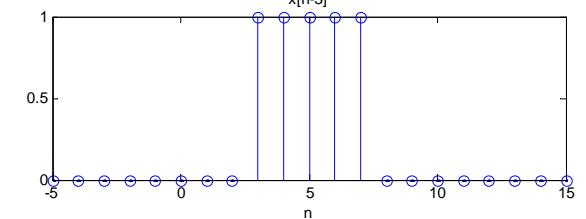
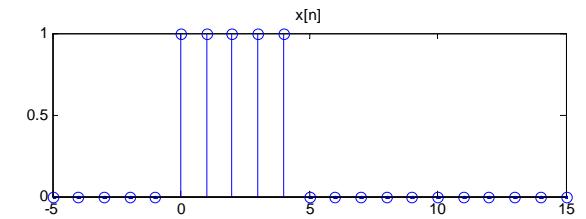
$$m_{est}[n] = \frac{1}{12} \left(\frac{1}{2} x[n-6] + x[n-5] + x[n-4] + \dots \right)$$

Amplifier

Delay

Instant power

Smoother



Box-Cox transformation: family of systems

Season estimation

5. A short introduction to system analysis

Basic system properties

Systems with memory

$$y[n] = T(\overbrace{x[n], x[n-1], x[n-2], \dots}^{\text{Present, Past}})$$

Memoryless systems

$$y[n] = T(x[n])$$

Invertible systems

$$y[n] = T(x[n]) \Rightarrow \exists T^{-1}(y) : x[n] = T^{-1}(y[n]) = T^{-1}(T(x[n]))$$

Causal systems

$$y[n] = T(x[n], x[n-1], x[n-2], \dots)$$

$$y[n] = T(x[n], \overbrace{x[n+1], x[n+2], \dots}^{\text{Future}})$$

Anticausal systems

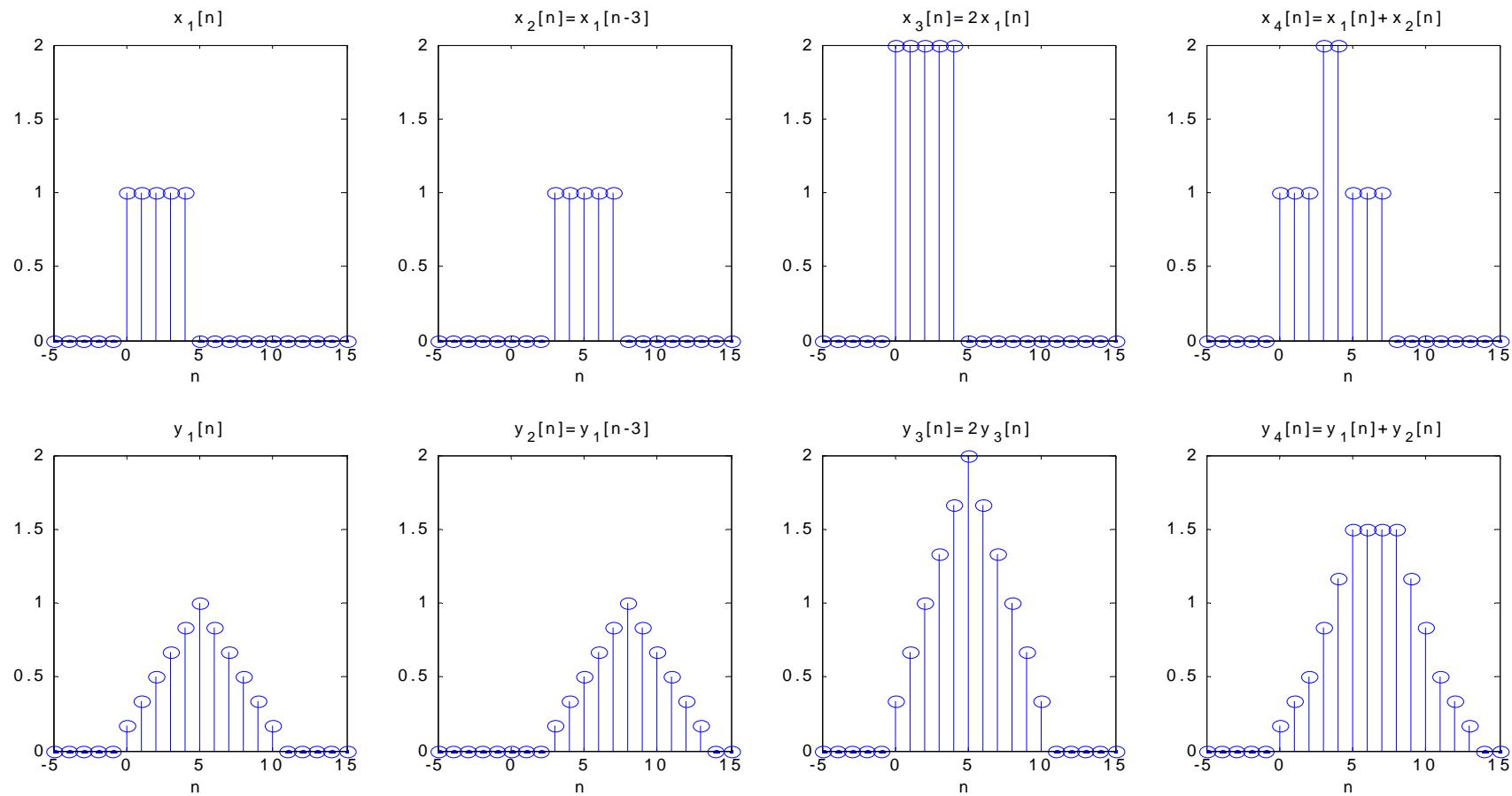
$$|x[n]| < B_x, \forall n \Rightarrow \exists B_T : |T(x[n])| < B_T, \forall n$$

Time invariant
systems

$$\left. \begin{aligned} y[n] &= T(x[n]) \Rightarrow y[n - n_0] = T(x[n - n_0]) \\ T(ax_1[n] + bx_2[n]) &= aT(x_1[n]) + bT(x_2[n]) \end{aligned} \right\} \text{LTI}$$

5. Basic introduction to system analysis

LTI systems



5. A short introduction to system analysis



Example: $y[n] = 3x[n]$

$$y[n] = x[n - 10]$$

$$y[n] = x^2[n]$$

$$y[n] = \frac{x[n-1] + x[n] + x[n+1]}{3}$$

$$y[n] = T_\lambda(x) = \begin{cases} \frac{x^\lambda[n]-1}{\lambda} & \lambda > 0 \\ \log(x[n]) & \lambda = 0 \end{cases}$$

$$S_{est}[n] = \left(\sum_{k=-2}^2 \frac{1}{2^{|k|+1}} x[n-k] \right) - m_{est}[n]$$

$$m_{est}[n] = \frac{1}{12} (\frac{1}{2} x[n-6] + x[n-5] + x[n-4] + \dots)$$

LTI, memoryless, invertible, causal, stable

LTI, with memory, invertible, causal, stable

Non linear, TI, memoryless, non invertible, causal, stable

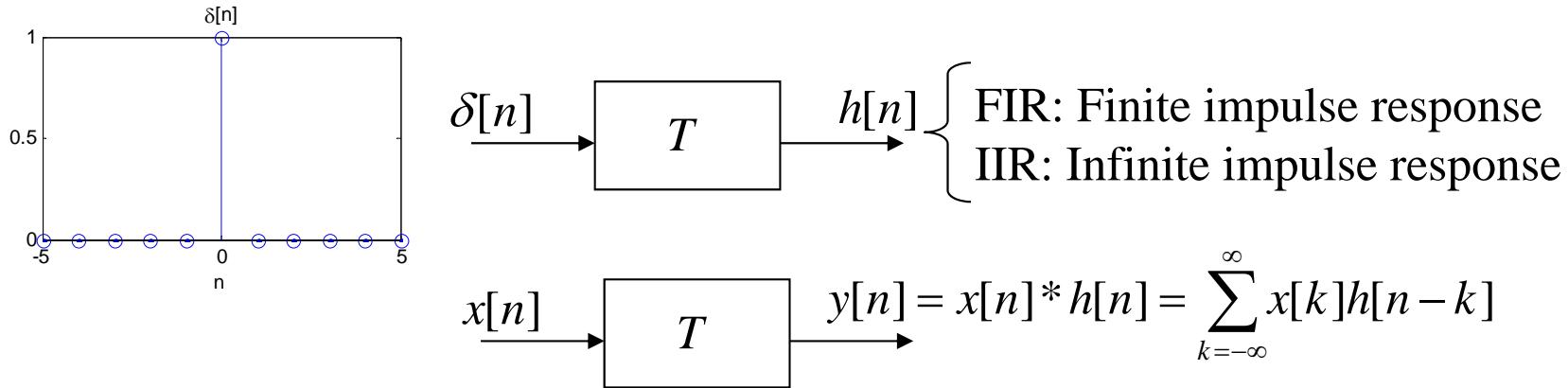
LTI, with memory, invertible, non causal, stable

Non linear, memoryless, invertible, causal, unstable

LTI, with memory, invertible, non causal, stable

5. A short introduction to system analysis

Impulse response of a LTI system

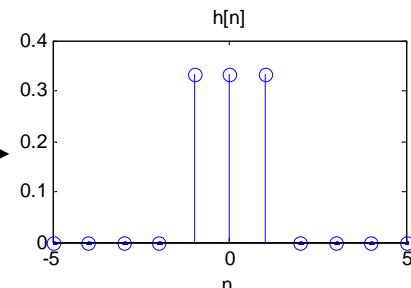


Example:

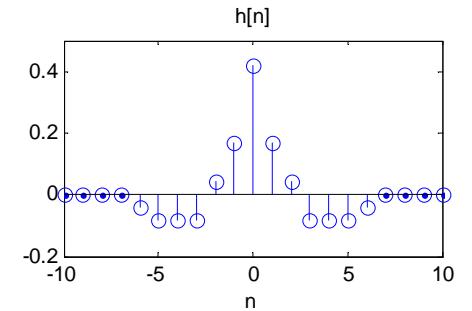
$$y[n] = \frac{x[n-1] + x[n] + x[n+1]}{3}$$

$$S_{est}[n] = \left(\sum_{k=-2}^2 \frac{1}{2^{|k|+1}} x[n-k] \right) - m_{est}[n]$$

$m_{est}[n] = \frac{1}{12} \left(\frac{1}{2} x[n-6] + x[n-5] + x[n-4] + \dots \right)$



$$h[n] = \frac{\delta[n-1] + \delta[n] + \delta[n+1]}{3}$$



5. A short introduction to system analysis

Transfer function

Difference equation

$$\sum_{k=0}^N a_k y[n-k] = \sum_{k=0}^M b_k x[n-k] \longleftrightarrow H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}} \quad z \in C$$

$$\sum_{k=0}^N a_k Y(z)z^{-k} = \sum_{k=0}^M b_k X(z)z^{-k}$$

The diagram illustrates the relationship between a difference equation, its transfer function, and its z-transform representation. At the top, a difference equation is shown: $\sum_{k=0}^N a_k y[n-k] = \sum_{k=0}^M b_k x[n-k]$. This is connected by a double-headed arrow to the transfer function $H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}}$. Below these, another double-headed arrow connects the transfer function to its z-transform representation: $\sum_{k=0}^N a_k Y(z)z^{-k} = \sum_{k=0}^M b_k X(z)z^{-k}$.

Impulse response of a LTI system $Y(z) = H(z)X(z)$

Example: $y[n] + y[n-1] = x[n] - 0.5x[n-1]$

$$Y(z) + Y(z)z^{-1} = X(z) - 0.5X(z)z^{-1}$$

$$H(z) = \frac{Y(z)}{X(z)} = \frac{1 - 0.5z^{-1}}{1 + z^{-1}}$$

$x[n + n_0] \xleftarrow{ZT} z^{n_0} X(z)$

6. Spectral representation of stationary processes

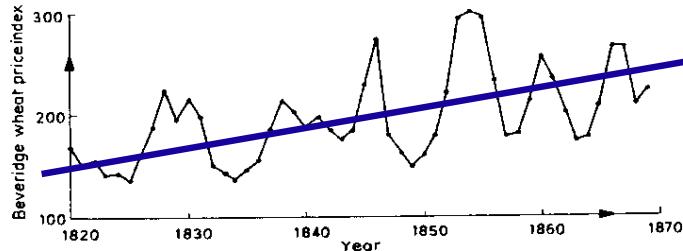


Figure 1.1 Part of the Beveridge wheat price index series.

$$x[n] = \text{trend}[n] + \text{seasonal}[n] + \text{random}[n]$$



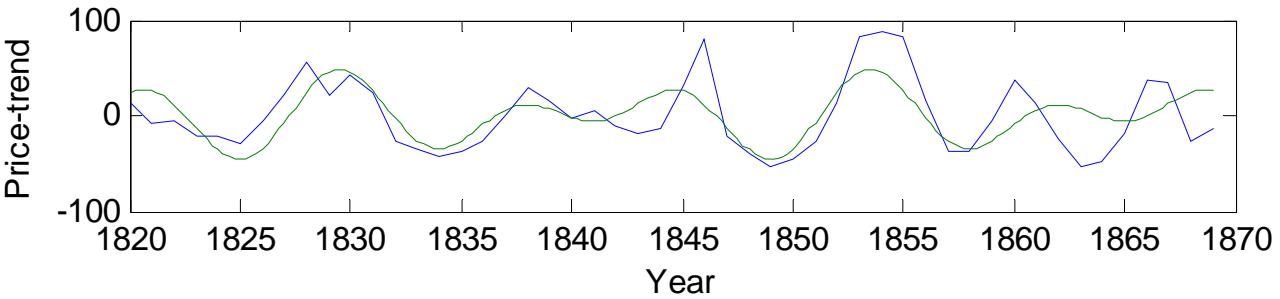
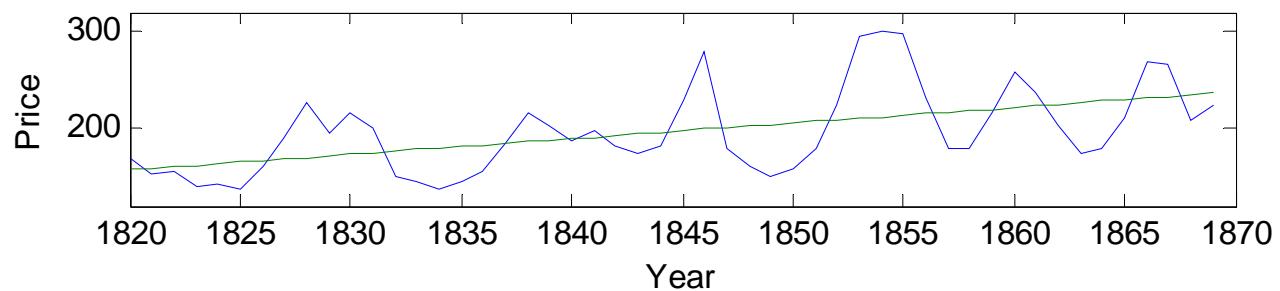
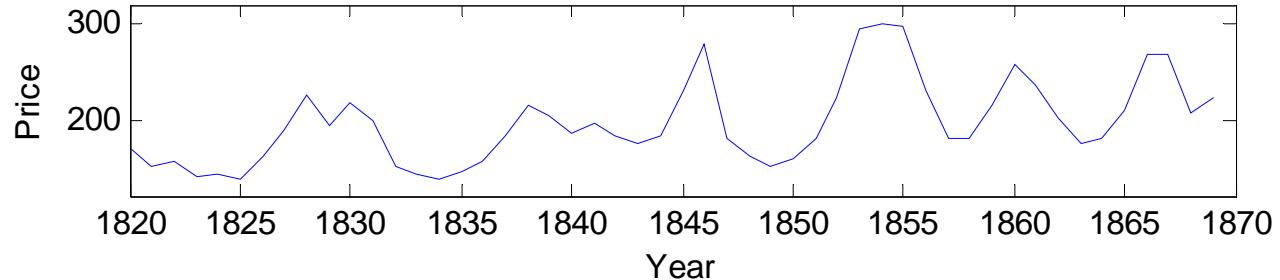
$$\text{trend}[n] = -2777 + 1.6n$$



$$\begin{aligned} y[n] &= x[n] - \text{trend}[n] \\ &= \text{seasonal}[n] + \text{random}[n] \\ &= 26.5 \cos\left(\frac{2\pi}{8}n + \frac{2\pi}{3}\right) + 22.5 \cos\left(\frac{2\pi}{12}n + \pi\right) + \text{random}[n] \end{aligned}$$

↑
Period=8 years

↑
Period=12 years



6. Spectral representation of stationary processes

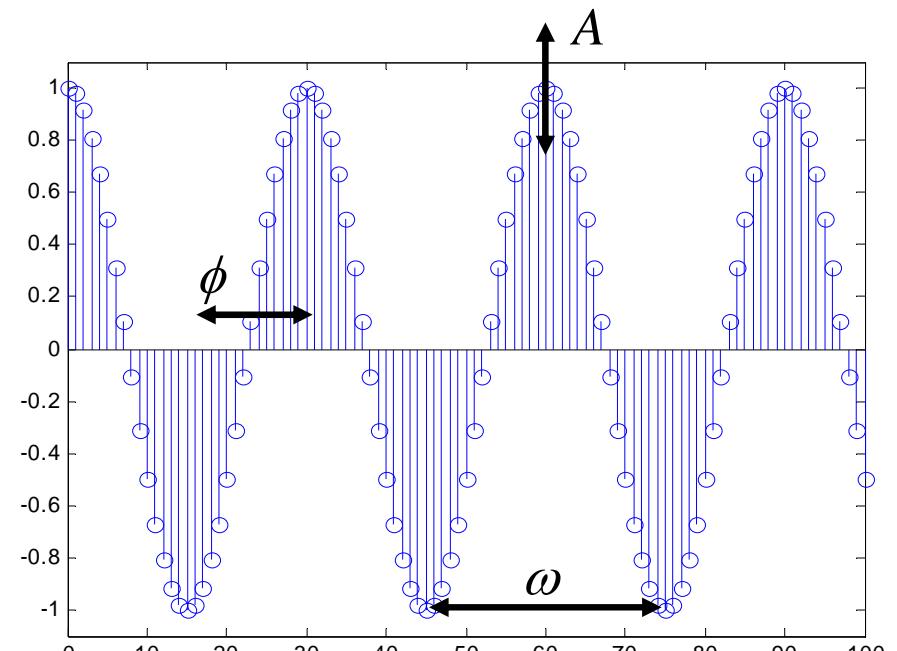
Harmonic components

$$x[n] = A \cos(\omega n + \phi) + \text{random}[n]$$

↑ Amplitude ↑ Frequency (rad/sample) ↓ Phase (rad)

$$x[n] = A \cos(2\pi f n T_s + \phi) + \text{random}[n]$$

↓ Frequency (Hertz)



$$\text{seasonal}[n] = \text{seasonal}[n + N] \Rightarrow N = \frac{2\pi k}{\omega} = \frac{k}{f T_s} = \frac{f_s}{f} k \quad k = 1, 2, 3, \dots$$

Example: $\text{seasonal}[n] = 26.5 \cos\left(\frac{2\pi}{8}n + \frac{2\pi}{3}\right) + 22.5 \cos\left(\frac{2\pi}{12}n + \pi\right)$

Period=8 years Period=12 years

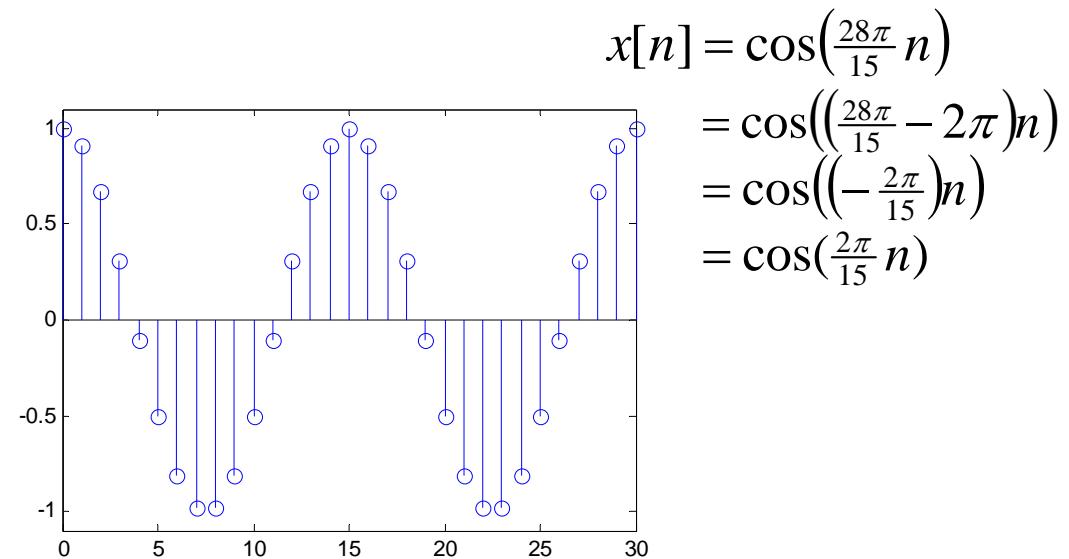
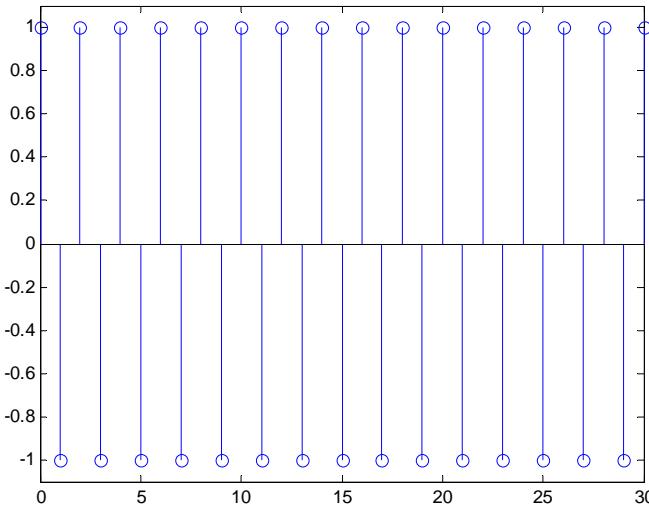
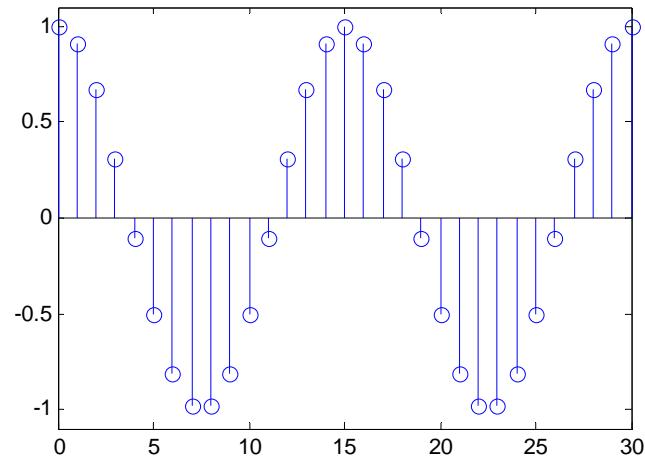
Period=lcm(N₁,N₂)=24 years

6. Spectral representation of stationary processes

Harmonic components

$$x[n] = A \cos(\pi n + \phi)$$

$$x[n] = \cos\left(\frac{2\pi}{15}n\right)$$



$$\begin{aligned}x[n] &= \cos\left(\frac{28\pi}{15}n\right) \\&= \cos\left(\left(\frac{28\pi}{15} - 2\pi\right)n\right) \\&= \cos\left(\left(-\frac{2\pi}{15}\right)n\right) \\&= \cos\left(\frac{2\pi}{15}n\right)\end{aligned}$$

6. Spectral representation of stationary processes

Harmonic representation

$$x[n] = 26.5 \cos\left(\frac{2\pi}{8}n + \frac{2\pi}{3}\right) + 22.5 \cos\left(\frac{2\pi}{12}n + \pi\right)$$

↓

$$x[n] = \sum_{k=1}^K A_k \cos(\omega_k n + \phi_k) \longrightarrow x[n] = \int_0^\pi A(\omega) \cos(\omega n + \phi(\omega)) d\omega$$

$A(\omega) \cos(\omega n + \phi(\omega)) = \frac{1}{2} (A(\omega)e^{j\phi(\omega)}e^{j\omega n} + A(\omega)e^{-j\phi(\omega)}e^{-j\omega n})$
 $X(\omega) = \pi A(\omega) e^{j\phi(\omega)}$
 $X(-\omega) = X^*(\omega)$

Fourier transform pair

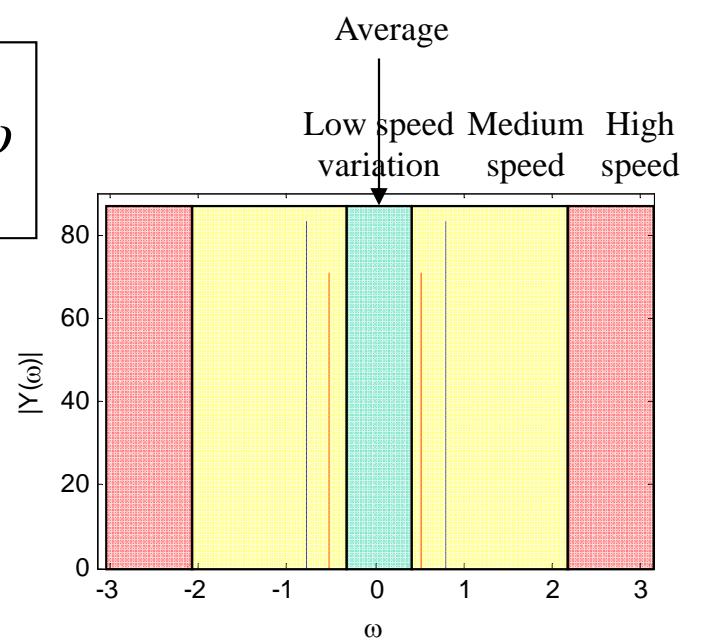
$$X(\omega) = \sum_{n=-\infty}^{\infty} x[n] e^{-j\omega n} \longleftrightarrow x[n] = \frac{1}{2\pi} \int_{-\pi}^{\pi} X(\omega) e^{j\omega n} d\omega$$

$$X(\omega) = FT\{x[n]\} \longleftrightarrow x[n] = FT^{-1}\{X(\omega)\}$$

Example: $Y(\omega) = FT\{26.5 \cos\left(\frac{2\pi}{8}n + \frac{2\pi}{3}\right) + 22.5 \cos\left(\frac{2\pi}{12}n + \pi\right)\}$

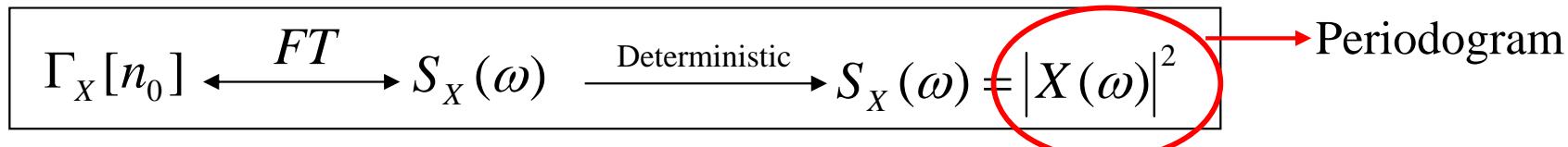
$$= 26.5\pi e^{j\frac{2\pi}{3}} \delta(\omega + \frac{2\pi}{8}) + 26.5\pi e^{j\frac{2\pi}{3}} \delta(\omega - \frac{2\pi}{8})$$

$$+ 22.5\pi e^{j\pi} \delta(\omega + \frac{2\pi}{12}) + 22.5\pi e^{j\pi} \delta(\omega - \frac{2\pi}{12})$$



6. Spectral representation of stationary processes

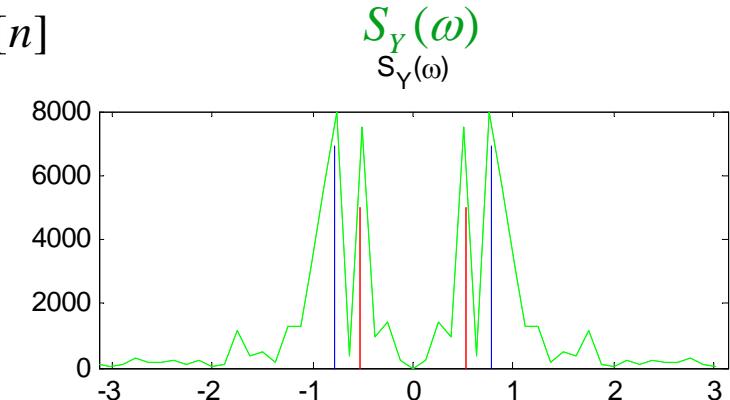
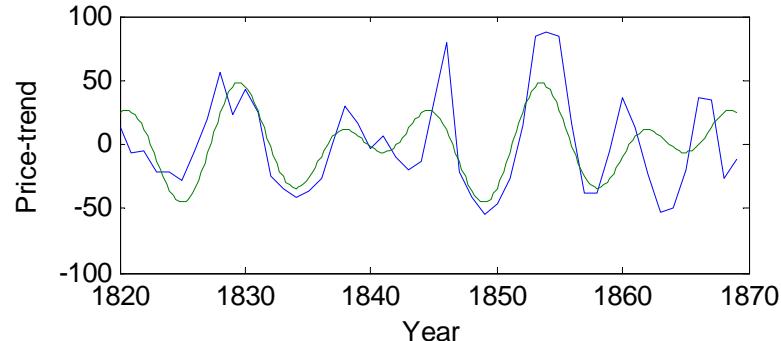
Power spectral density (PSD)



$$\text{Example: } \Gamma_W[n_0] = \sigma_W^2 \delta[n_0] \xleftarrow{FT} S_W(\omega) = \sigma_W^2$$

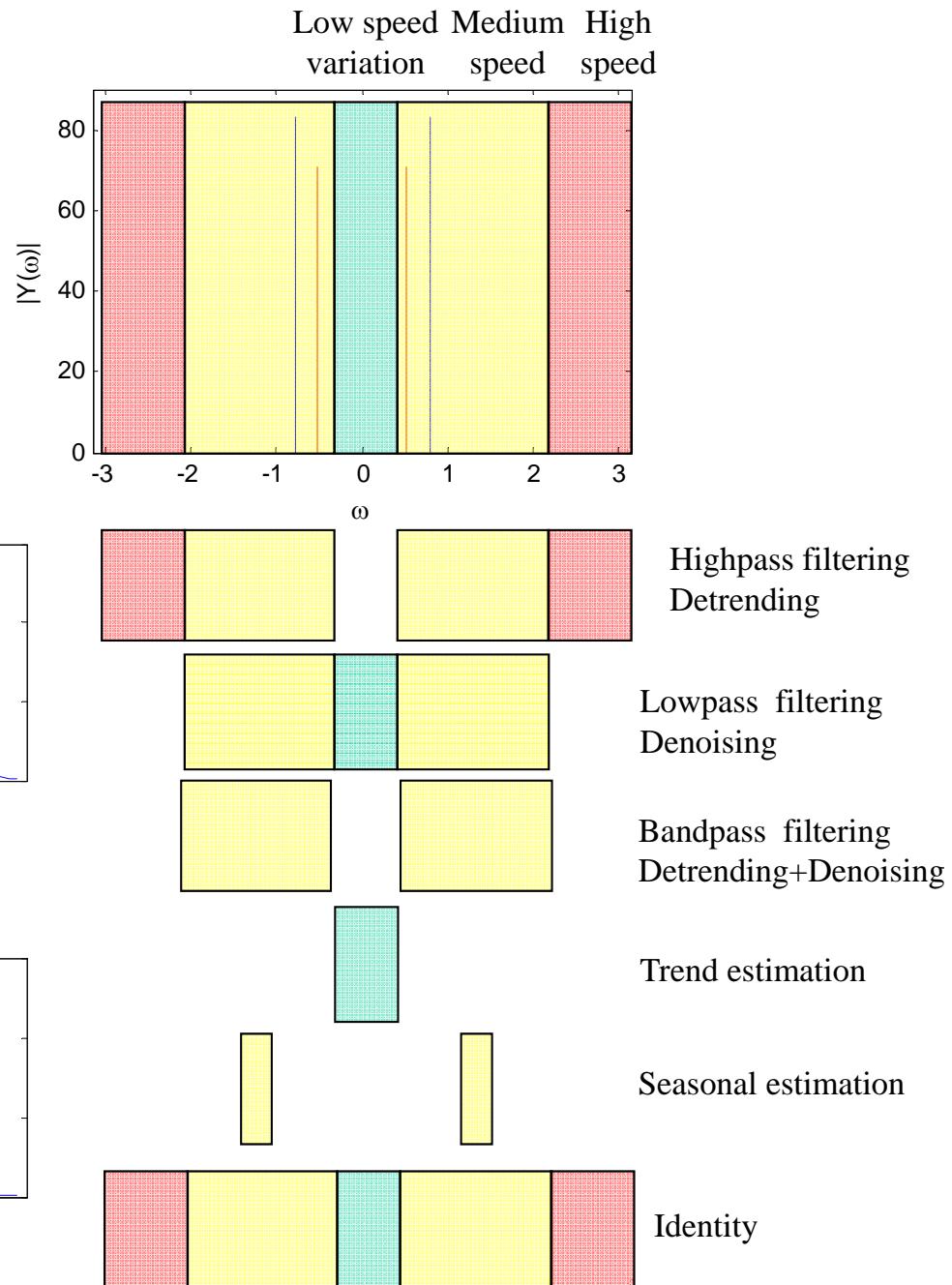
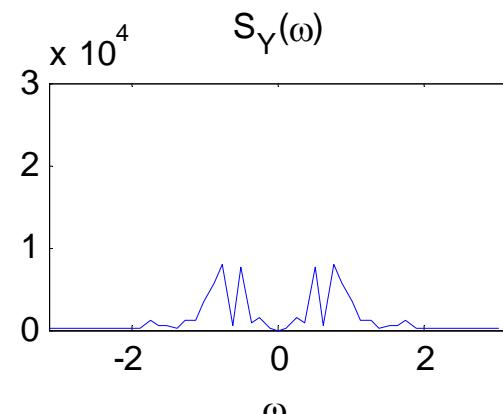
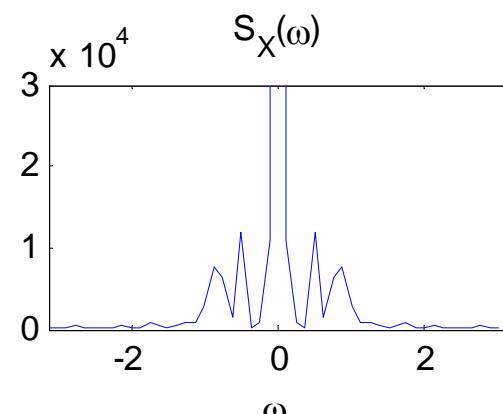
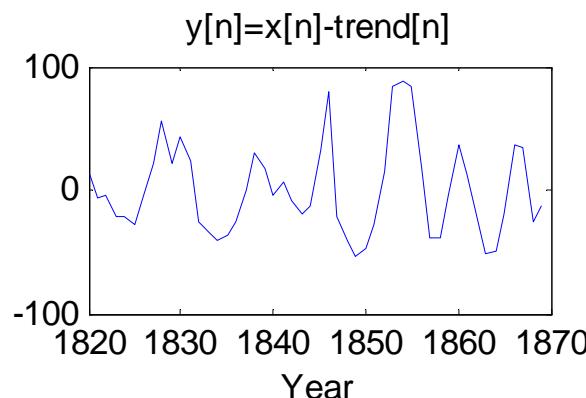
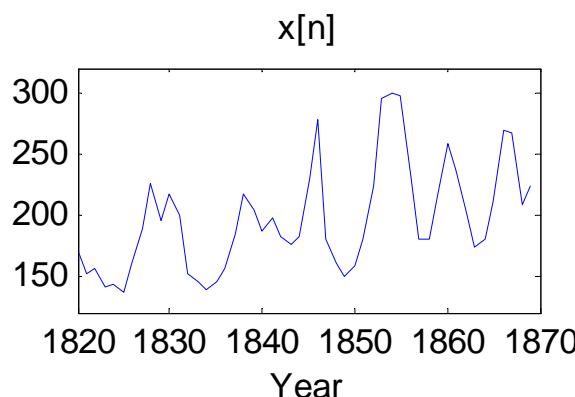
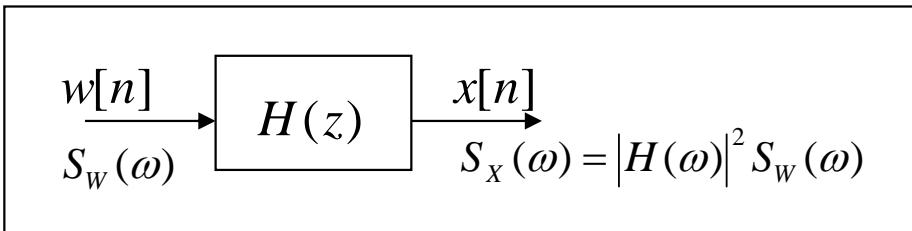
Example:

$$y[n] = 26.5 \cos\left(\frac{2\pi}{8}n + \frac{2\pi}{3}\right) + 22.5 \cos\left(\frac{2\pi}{12}n + \pi\right) + w[n]$$

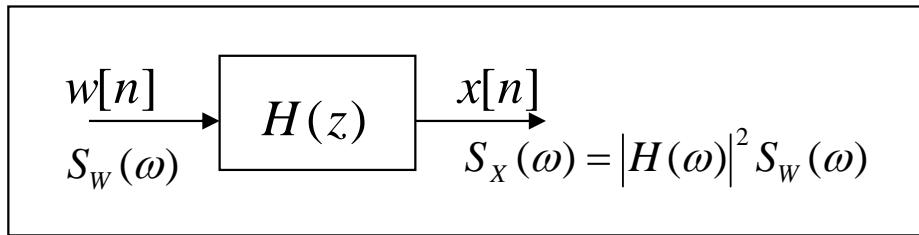


$$\begin{aligned} \hat{S}_Y(\omega) &= (26.5\pi)^2 \delta(\omega + \frac{\omega}{8}) + (26.5\pi)^2 \delta(\omega - \frac{\omega}{8}) \\ &\quad + (22.5\pi)^2 \delta(\omega + \frac{2\pi}{12}) + (22.5\pi)^2 \delta(\omega - \frac{2\pi}{12}) \end{aligned}$$

7. Detrending and filtering



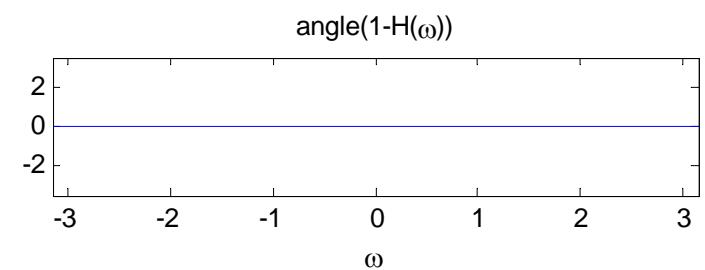
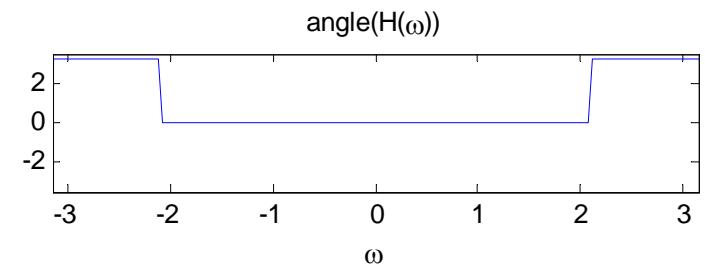
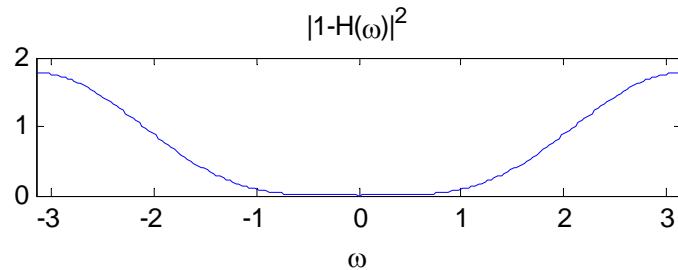
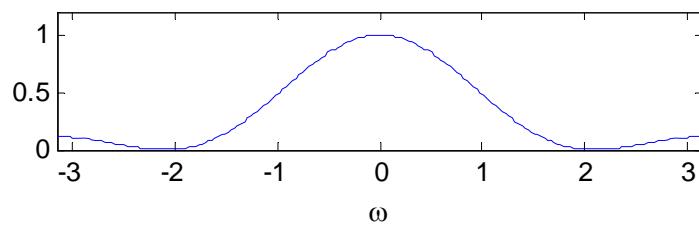
7. Detrending and filtering



$$H(\omega) = H(z) \Big|_{z=e^{j\omega}} = H(e^{j\omega})$$

Example: Smoother

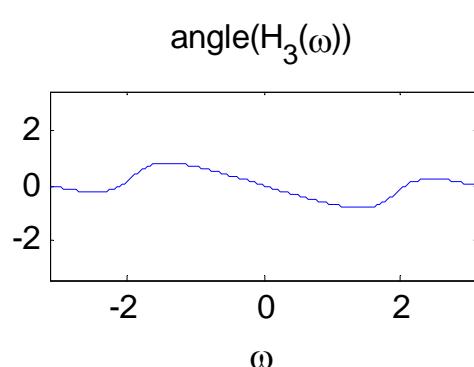
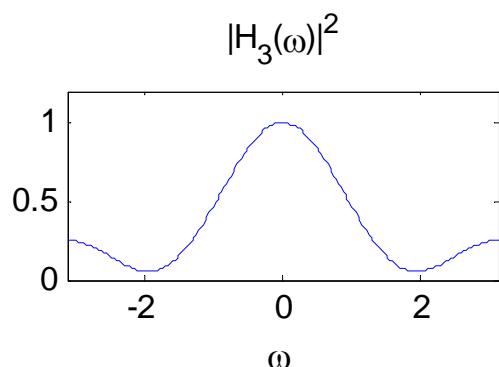
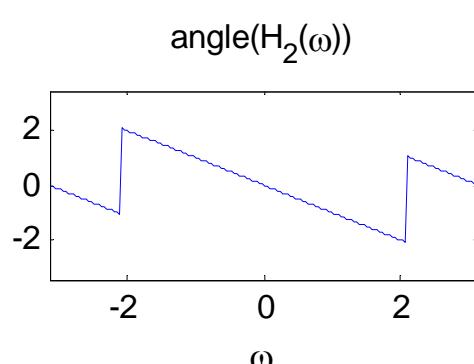
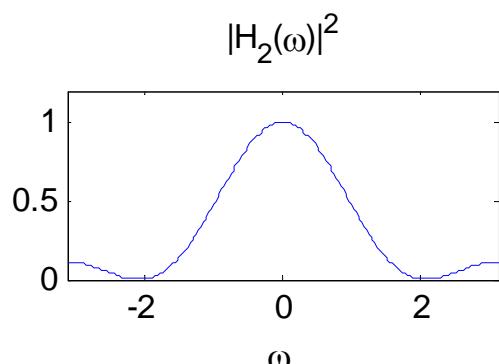
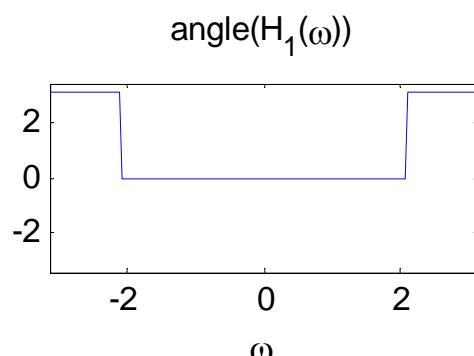
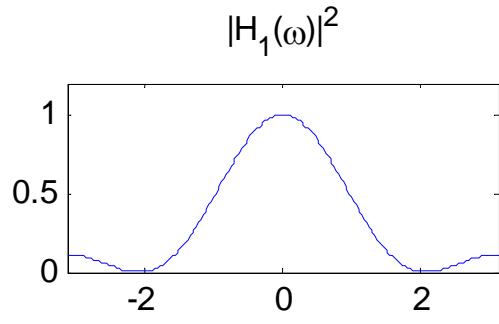
$$y[n] = \frac{x[n-1] + x[n] + x[n+1]}{3} \longleftrightarrow H(z) = \frac{1}{3}(z^{-1} + 1 + z) \longleftrightarrow H(\omega) = \frac{1}{3}(e^{-j\omega} + 1 + e^{j\omega})$$



$$y[n] = x[n] - \frac{x[n-1] + x[n] + x[n+1]}{3} \longleftrightarrow H(z) = 1 - \frac{1}{3}(z^{-1} + 1 + z) \longleftrightarrow H(\omega) = 1 - \frac{1}{3}(e^{-j\omega} + 1 + e^{j\omega})$$

7. Detrending and filtering

Example: Smoothers



$$y[n] = \frac{x[n-1] + x[n] + x[n+1]}{3}$$

$$H_1(z) = \frac{1}{3} (z^{-1} + 1 + z)$$

$$H_1(\omega) = \frac{1}{3} (e^{-j\omega} + 1 + e^{j\omega})$$

$$y[n] = \frac{x[n] + x[n-1] + x[n-2]}{3}$$

$$H_2(z) = \frac{1}{3} (1 + z^{-1} + z^{-2})$$

$$H_2(\omega) = \frac{1}{3} (1 + e^{-j\omega} + e^{-j2\omega})$$

$$y[n] = \frac{1}{2} x[n] + \frac{1}{4} x[n-1] + \frac{1}{4} x[n-2]$$

$$H_3(z) = \frac{1}{2} + \frac{1}{4} z^{-1} + \frac{1}{4} z^{-2}$$

$$H_3(\omega) = \frac{1}{2} + \frac{1}{4} e^{-j\omega} + \frac{1}{4} e^{-j2\omega}$$

7. Detrending and filtering

Example: Yearly season estimation

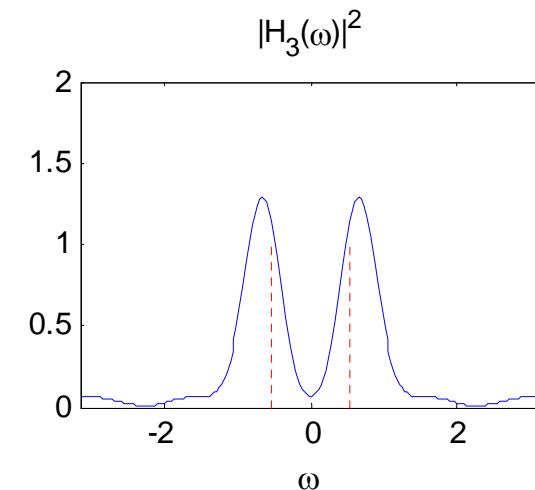
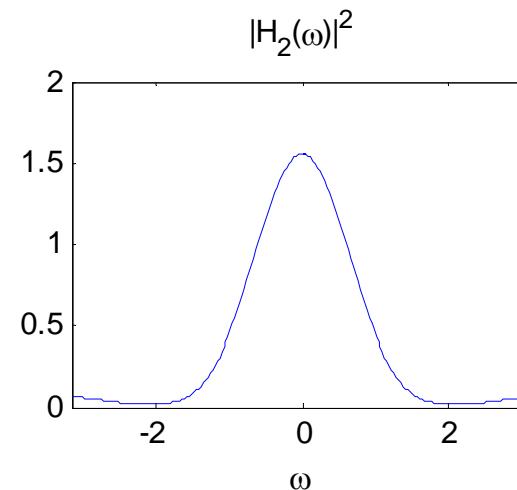
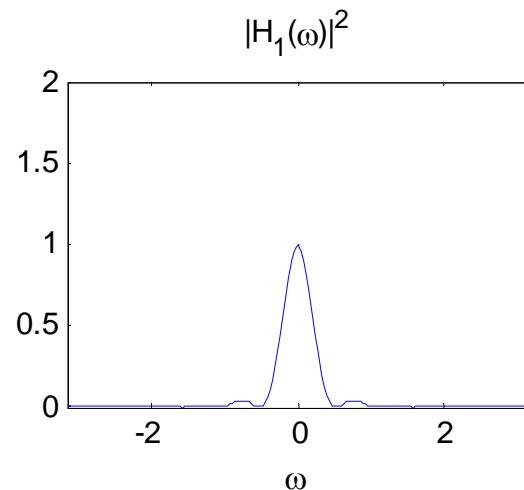
$$S_{est}[n] = \left(\sum_{k=-2}^2 \frac{1}{2^{|k|+1}} x[n-k] \right) - m_{est}[n]$$

$m_{est}[n] = \frac{1}{12} \left(\frac{1}{2} x[n-6] + x[n-5] + x[n-4] + \dots \right)$

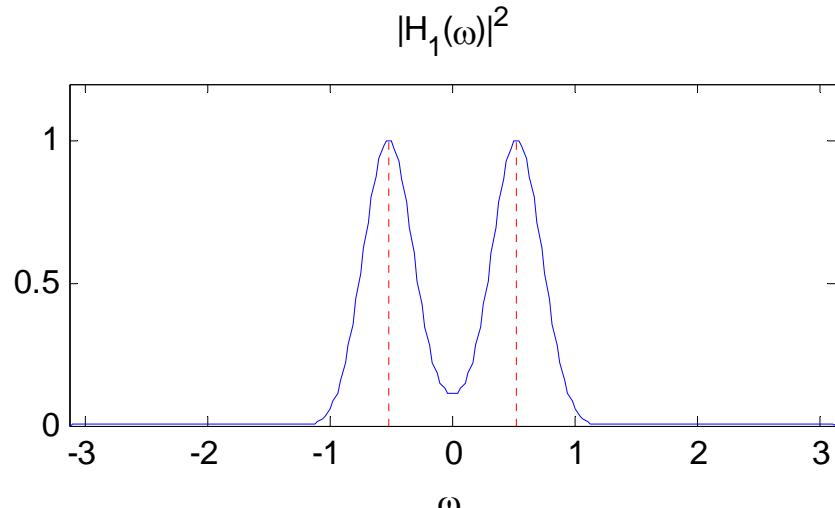
$$H_1(z) = \frac{M_{est}(z)}{X(z)} = \frac{1}{12} \left(\frac{1}{2} z^{-6} + z^{-5} + z^{-4} + z^{-3} + z^{-2} + z^{-1} + 1 + z + z^2 + z^3 + z^4 + z^5 + \frac{1}{2} z^6 \right)$$

$$H_2(z) = \frac{Y'(z)}{X(z)} = \frac{1}{8} z^{-2} + \frac{1}{4} z^{-1} + \frac{1}{2} + \frac{1}{4} z + \frac{1}{8} z^2$$

$$H(z) = H_2(z) - H_1(z)$$



7. Detrending and filtering



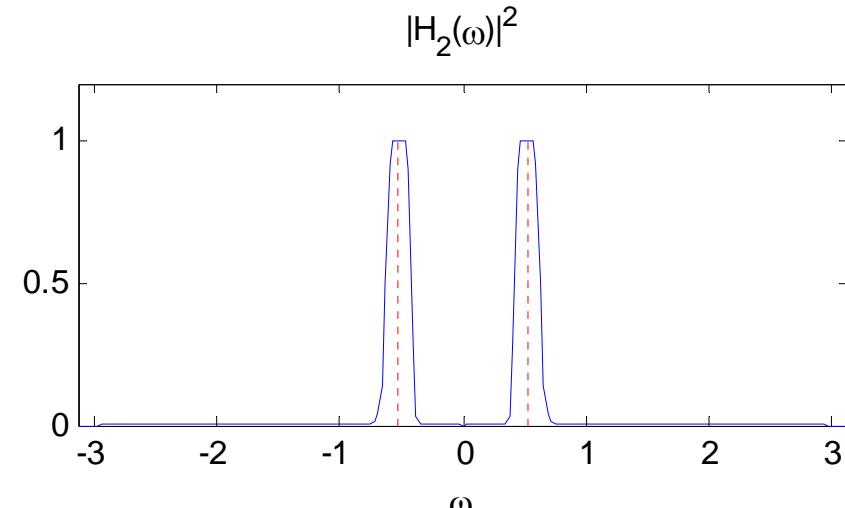
```
fc=2*pi/12;
```

ω

```
ef=2*pi/60;
```

```
b1=fir1(18,[fc-ef fc+ef]/pi);
```

```
y1[n]=-0.0101(x[n-8]+x[n+8])-0.0313(x[n-7]+x[n+7])-0.0611(x[n-6]+x[n+6])
-0.0802(x[n-5]+x[n+5])-0.0634(x[n-4]+x[n+4])+0.0934(x[n-2]+x[n+2])
+0.1772(x[n-1]+x[n+1])+0.2108x[n]
```



```
fc=2*pi/12;
```

```
ef=2*pi/60;
```

```
[b1,a1]=butter(4,[fc-ef fc+ef]/pi);
```

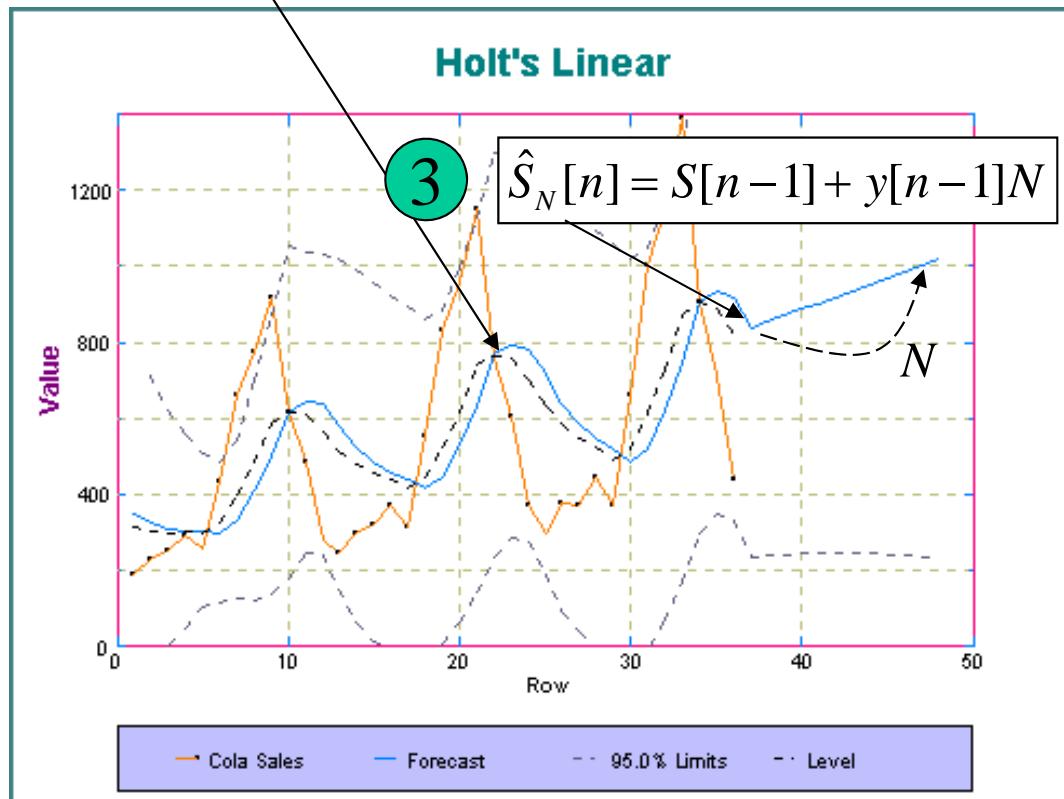
```
y2[n]=0.000093(x[n]+x[n-8])-0.00037(x[n-1]+x[n-6])+0.00056x[n-4]
-6.5y[n-1]+19.3y[n-2]-34y[n-3]+39y[n-4]-29.7y[n-5]+14.6y[n-6]-4.3y[n-7]+0.58y[n-8]
```

7. Detrending and filtering

Example: Holt's linear exponential smoothing

1 $S[n] = ax[n] + (1-a)(x[n-1] + y[n-1])$

2 $y[n] = b(S[n] - S[n-1]) + (1-b)y[n-1]$



$$Y(z) = b(S(z) - S(z)z^{-1}) + (1-b)Y(z)z^{-1} = f(S)$$
$$= f(X)$$

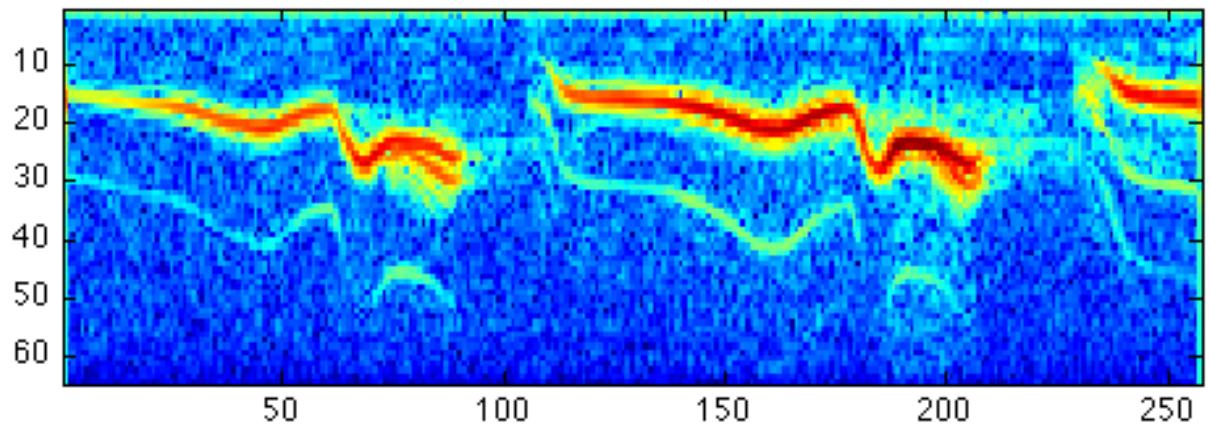
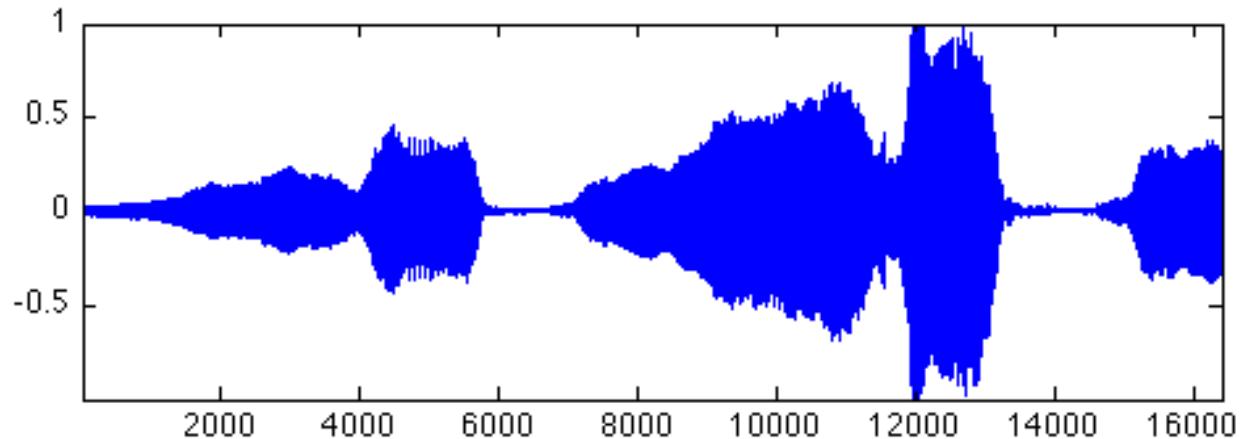
$$S(z) = aX(z) + (1-a)(X(z)z^{-1} + Y(z)z^{-1})$$
$$= f(X, Y) = f(X)$$

$$\hat{S}_N(z) = S(z)z^{-1} + NY(z)z^{-1} = f(X)$$

8. Non-stationary processes

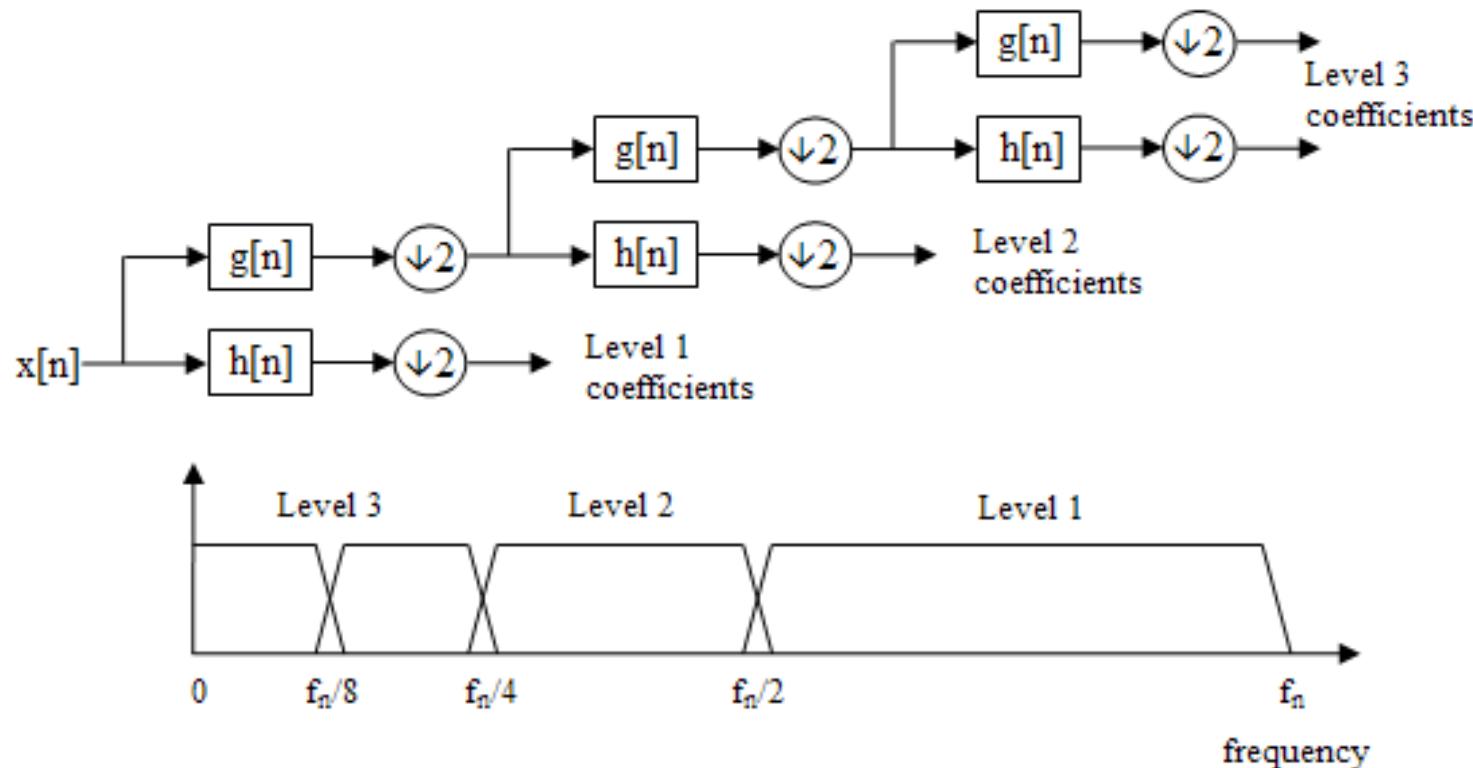
Short-time Fourier Transform

$$X[m, \omega] = \sum_{n=-\infty}^{\infty} x[n+m]w[n]e^{-j\omega n}$$



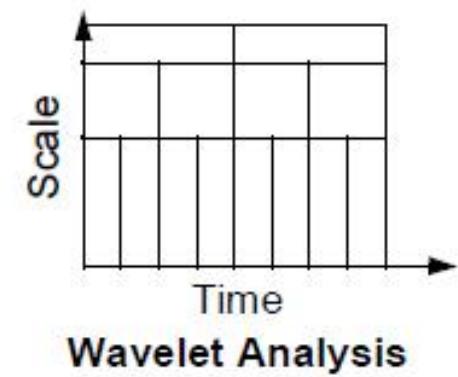
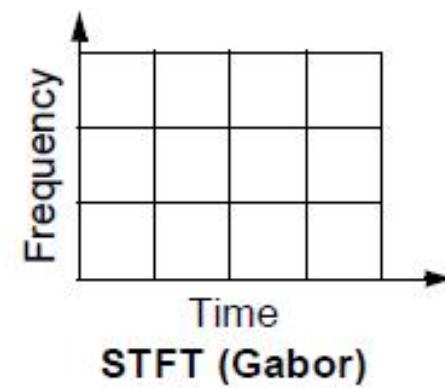
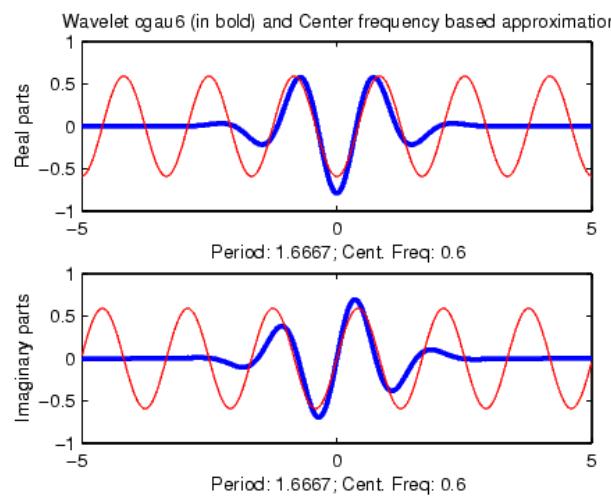
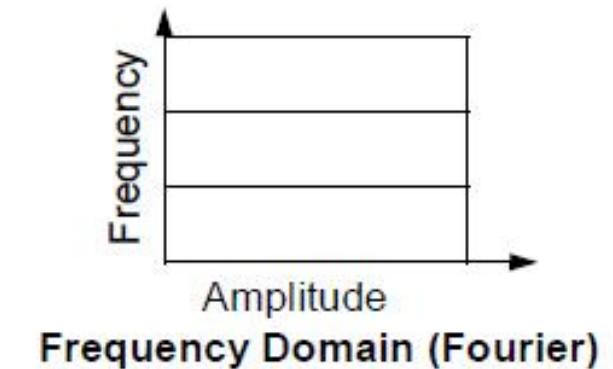
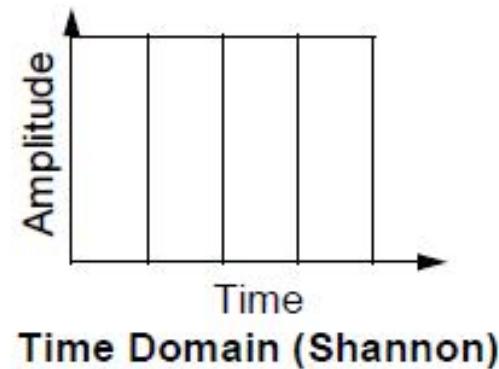
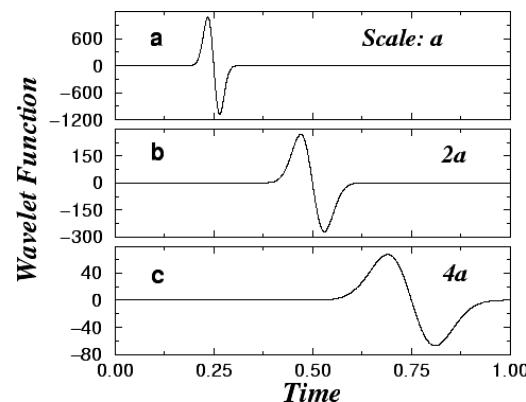
8. Non-stationary processes

Wavelet transform



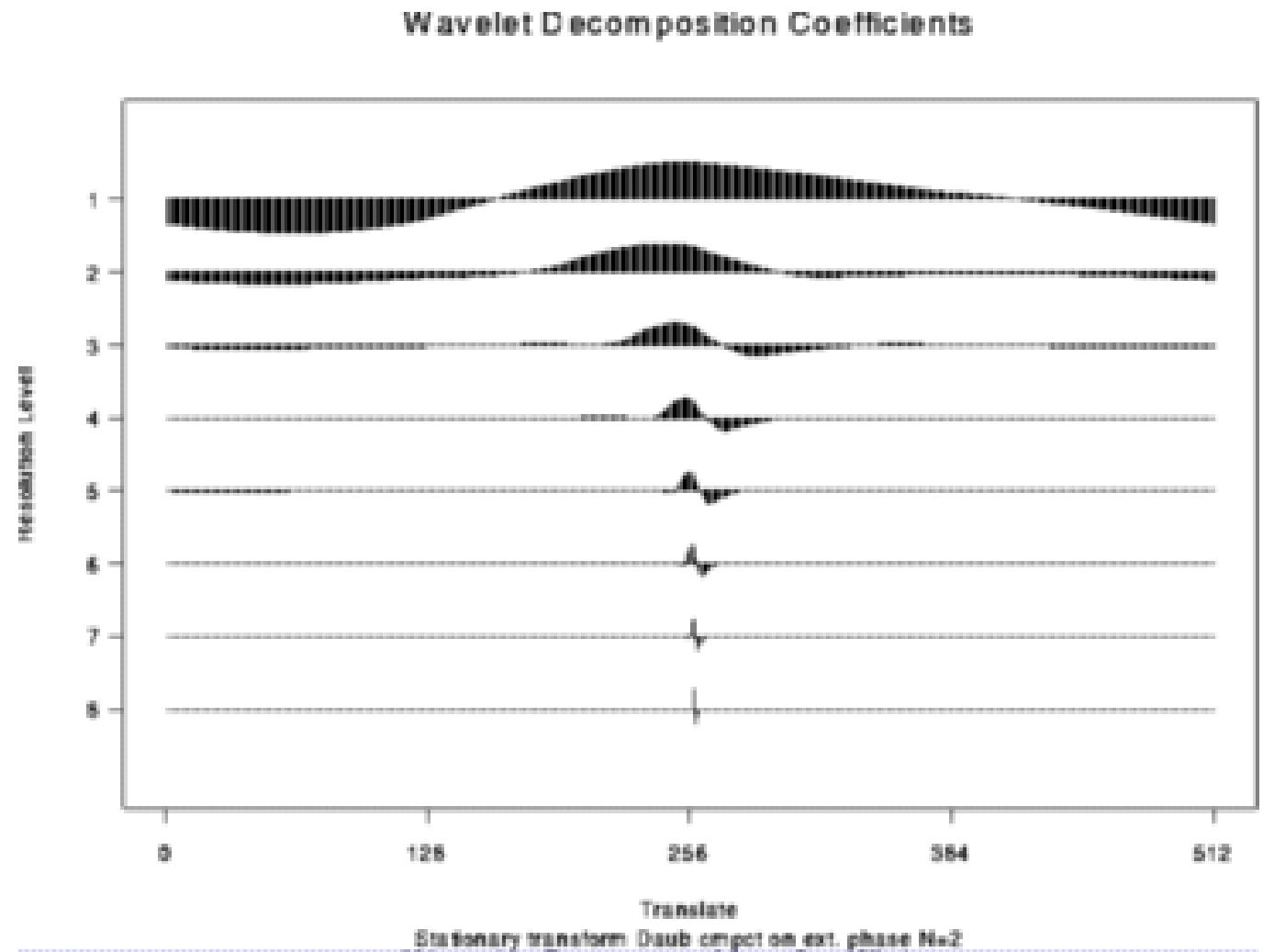
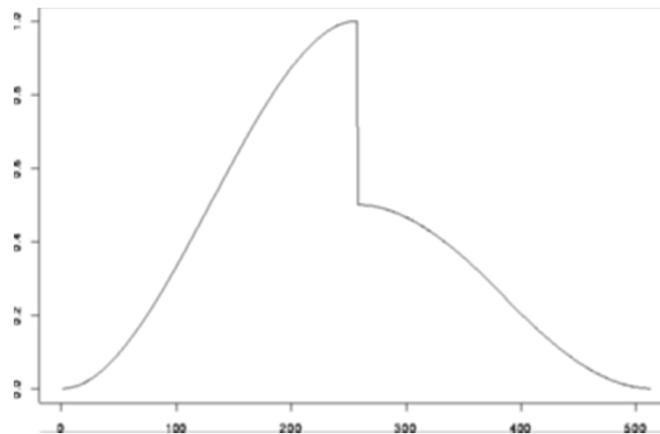
8. Non-stationary processes

Wavelet transform



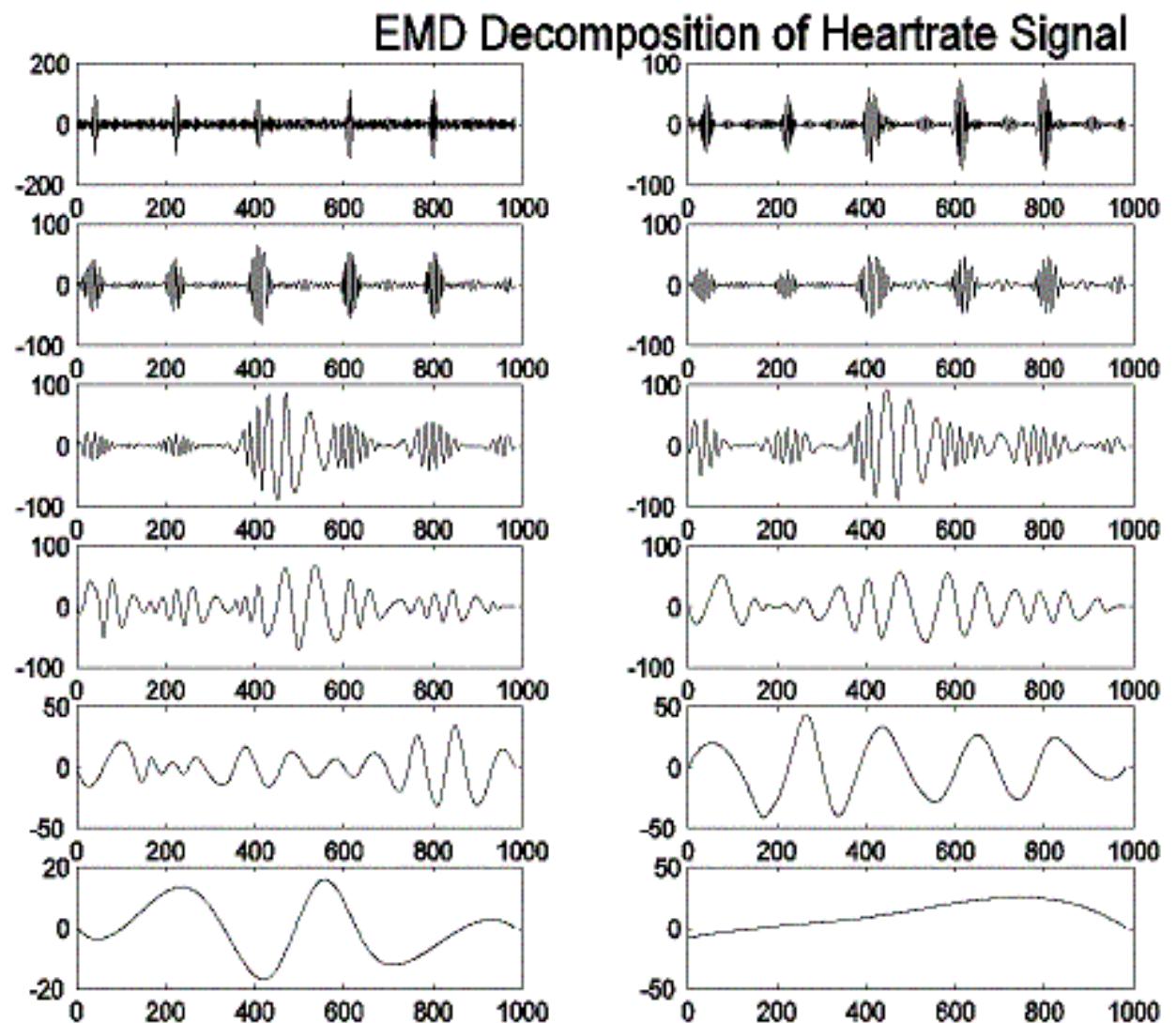
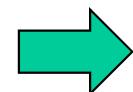
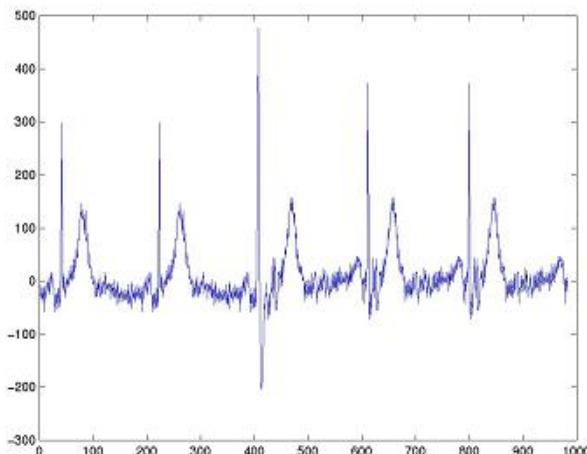
8. Non-stationary processes

Wavelet transform



8. Non-stationary processes

Empirical Mode Decomposition



Session outline

1. Goal
2. Linear and non-linear regression
3. Polynomial fitting
4. Cubic spline fitting
5. A short introduction to system analysis
6. Spectral representation of stationary processes
7. Detrending and filtering
8. Non-stationary processes

Bibliography

- C. Chatfield. *The analysis of time series: an introduction*. Chapman & Hall, CRC, 1996.
- D.S.G. Pollock. *A handbook of time-series analysis, signal processing and dynamics*. Academics Press, 1999.
- J. D. Hamilton. *Time series analysis*. Princeton Univ. Press, 1994.



CEU
*Universidad
San Pablo*

Time Series Analysis

Session III: Probability models for time series

Carlos Óscar Sánchez Sorzano, Ph.D.
Madrid,

Session outline

1. Goal
2. A short introduction to system analysis
3. Moving Average processes (MA)
4. Autoregressive processes (AR)
5. Autoregressive, Moving Average (ARMA)
6. Autoregressive, Integrated, Moving Average (ARIMA, FARIMA)
7. Seasonal, Autoregressive, Integrated, Moving Average (SARIMA)
8. Known external inputs: System identification
9. A family of models
10. Nonlinear models
11. Parameter estimation
12. Order selection
13. Model checking
14. Self-similarity, Fractal dimension, and Chaos theory

1. Goal

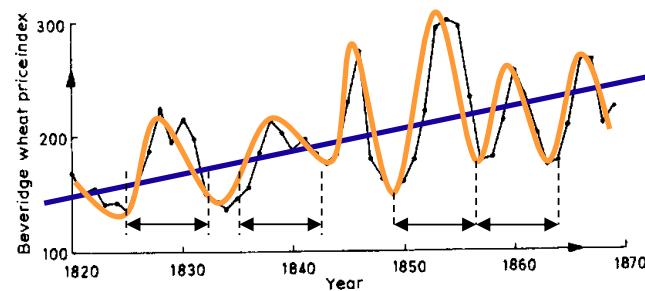
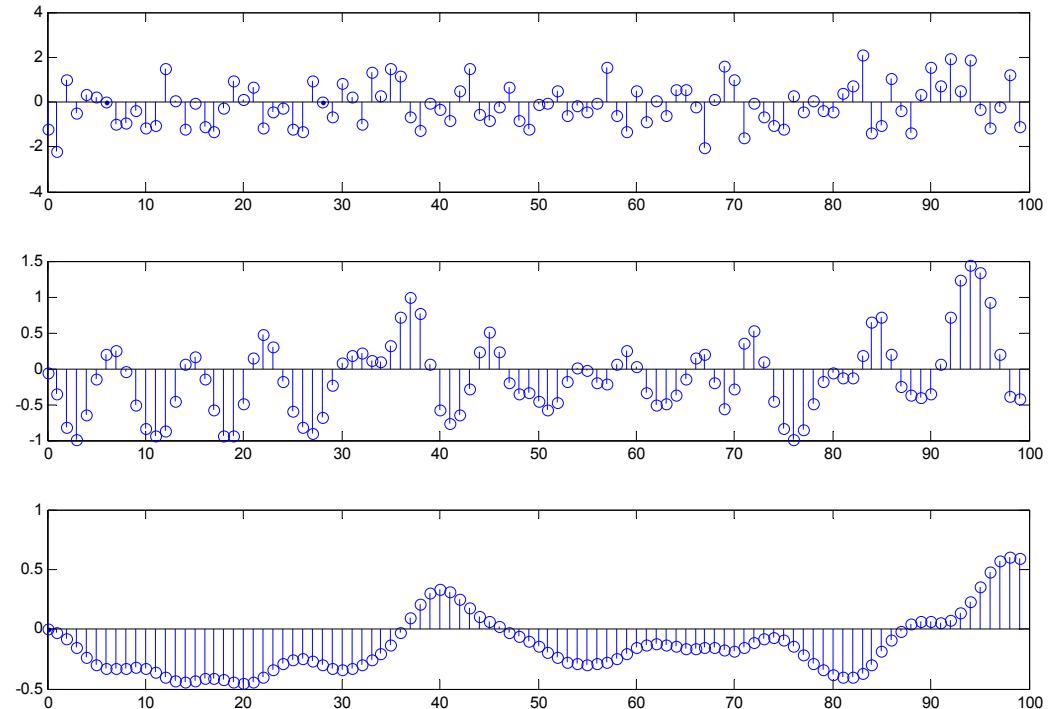


Figure 1.1 Part of the Beveridge wheat price index series.



$$x[n] = \text{trend}[n] + \text{periodic}[n] + \text{random}[n]$$

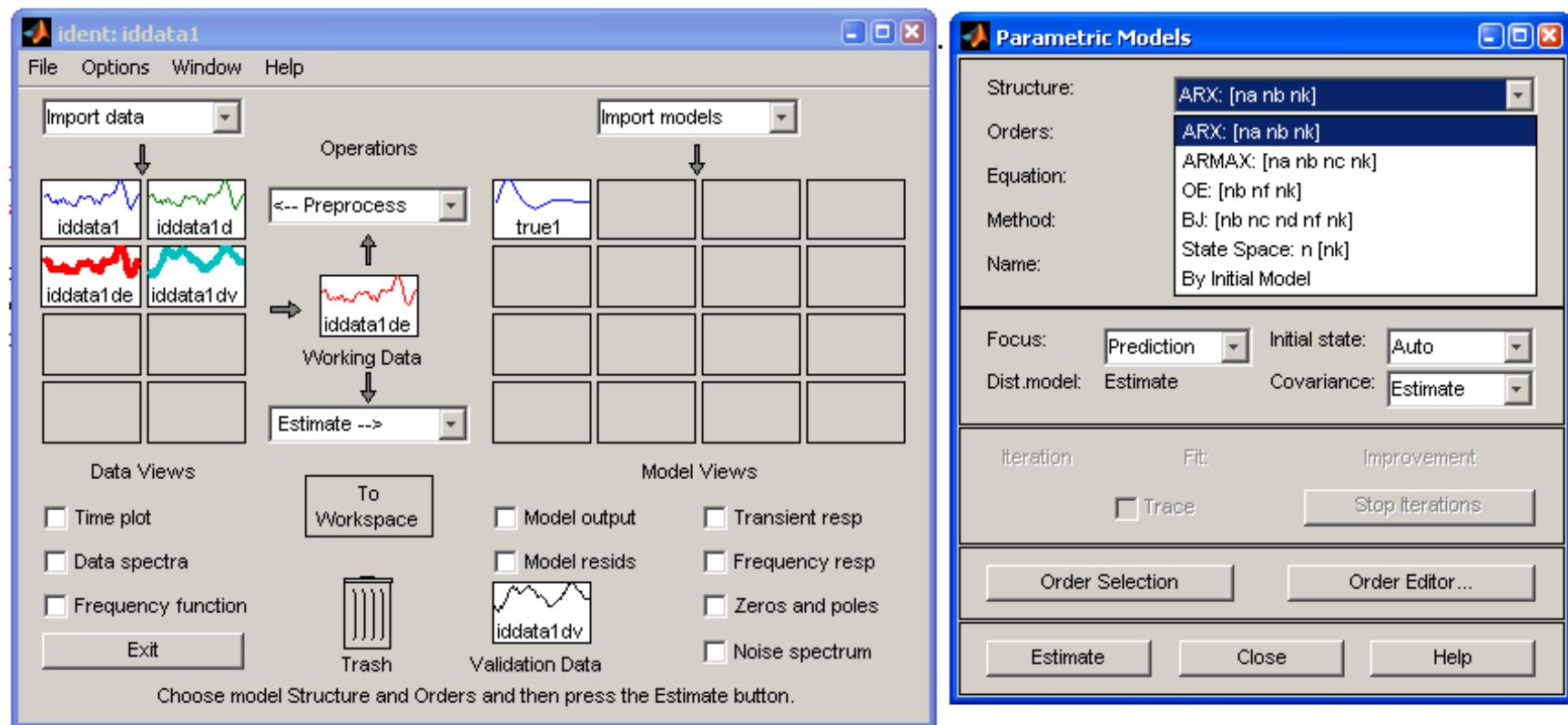


$$\text{random}[n] = 0.9\text{random}[n] + \text{completelyRandom}[n]$$

$$\text{random}[n] = \text{completelyRandom}[n] + 0.9\text{completelyRandom}[n - 1]$$

Explained by statistical models (AR, MA, ...)

1. Goal



2. A short introduction to system analysis



Difference equation

Transfer function

$$\sum_{k=0}^N a_k y[n-k] = \sum_{k=0}^M b_k x[n-k] \longleftrightarrow H(z) = \frac{Y(z)}{X(z)} = \frac{\sum_{k=0}^M b_k z^{-k}}{\sum_{k=0}^N a_k z^{-k}} \quad z \in C$$

$$\sum_{k=0}^N a_k Y(z) z^{-k} = \sum_{k=0}^M b_k X(z) z^{-k}$$

Example: $y[n] + y[n-1] = x[n] - 0.5x[n-1]$

$$\begin{aligned} & y[n] + y[n-1] = x[n] - 0.5x[n-1] \\ & \updownarrow \\ & Y(z) + Y(z)z^{-1} = X(z) - 0.5X(z)z^{-1} \\ & \updownarrow \\ & H(z) = \frac{Y(z)}{X(z)} = \frac{1 - 0.5z^{-1}}{1 + z^{-1}} \end{aligned}$$

$$x[n + n_0] \xrightarrow{ZT} z^{n_0} X(z)$$

2. A short introduction to system analysis

Poles/Zeros

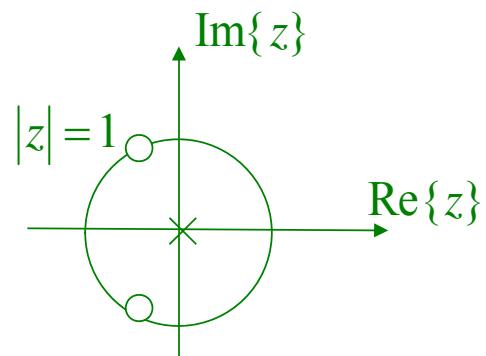
z_0 is a pole of $H(z)$ iff $|H(z_0)| = \infty$

z_0 is a zero of $H(z)$ iff $H(z_0) = 0$

Example: $y[n] = \frac{x[n-1] + x[n] + x[n+1]}{3}$

$$H(z) = \frac{1}{3} z^{-1} + \frac{1}{3} + \frac{1}{3} z$$

$$\text{Poles: } z = 0, \infty \quad \text{Zeros: } z = -\frac{1}{2} \pm j \frac{\sqrt{3}}{2}$$



Stability of LTI systems

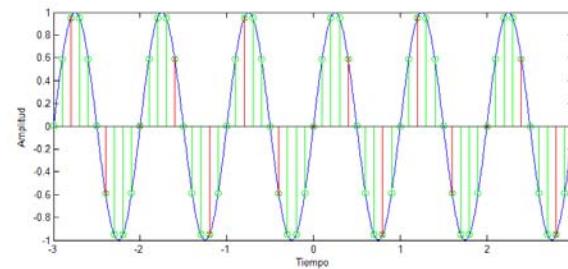
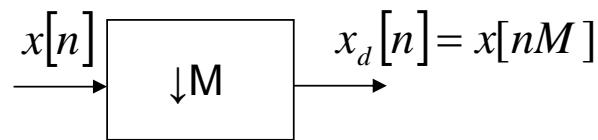
A causal system is stable iff all its poles are inside the unit circle $|z| < 1$

Invertibility of LTI systems

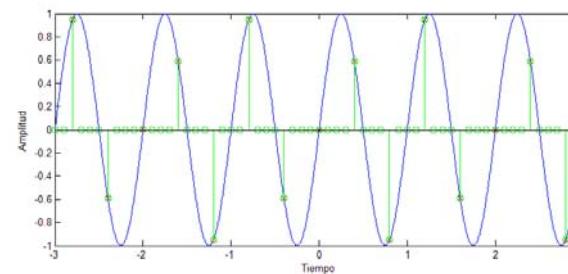
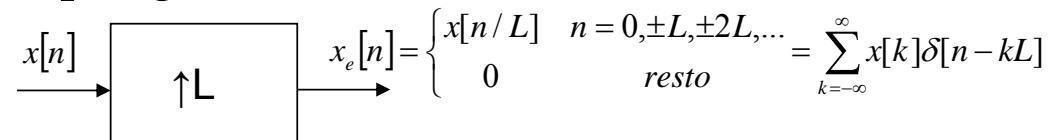
The transfer function of the inverse system of a LTI system whose transfer function is $H(z)$ is $\frac{1}{H(z)}$. Therefore, the zeros of one system are the poles of its inverse, and viceversa.

2. A short introduction to system analysis

Downsampling



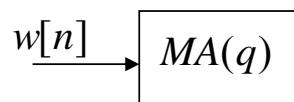
Upsampling



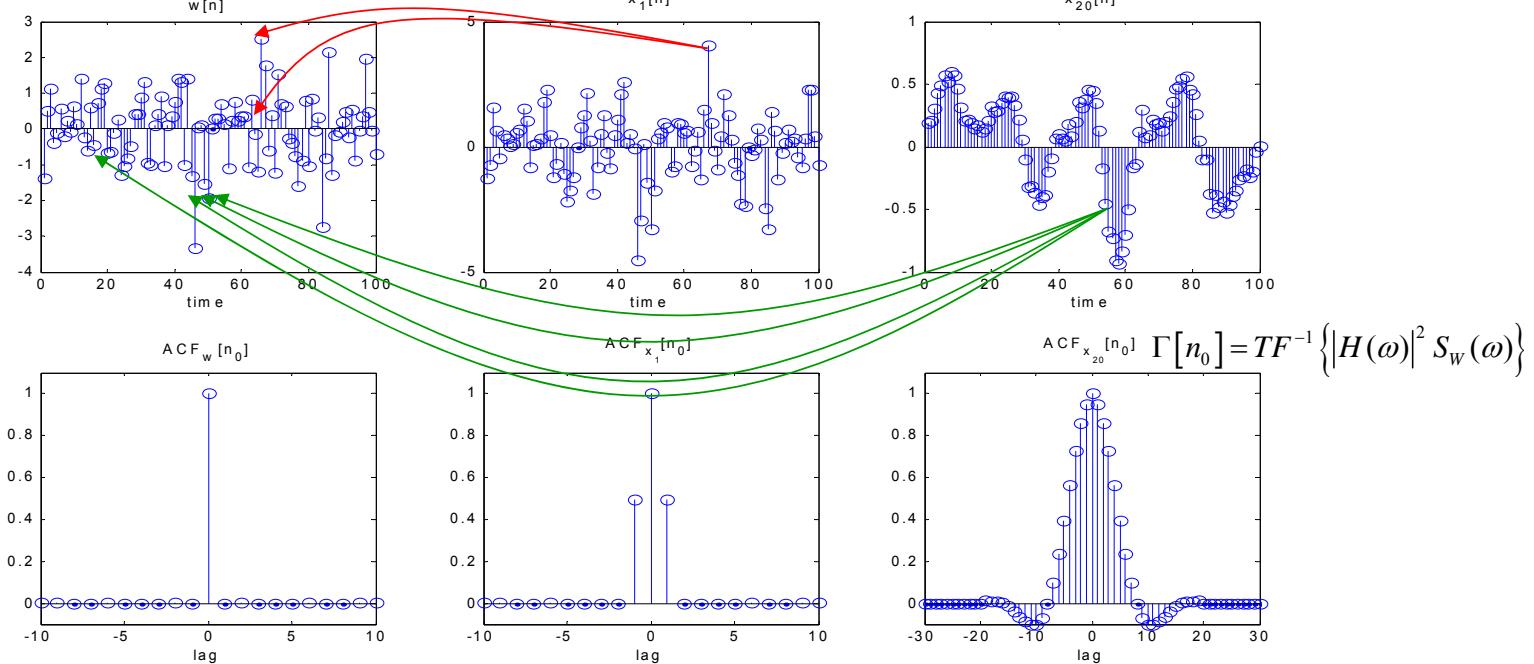
3. Moving average processes: MA(q)

Definition

$$H(z) = b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_q z^{-q} = B(z) \longrightarrow \text{LTI, with memory, invertible, causal, stable}$$



$$x[n] = b_0 w[n] + b_1 w[n-1] + \dots + b_q w[n-q]$$



3. Moving average processes: MA(q)

Statistical properties

$$\begin{aligned}
 w[n] &\xrightarrow{\quad} MA(q) \xrightarrow{\quad} x[n] = \sum_{k=0}^q b_k w[n-k] \\
 N(0, \sigma_w^2) &\xrightarrow{\quad} N(0, \sigma_w^2 \sum_{k=0}^q b_k^2) \\
 \Gamma_w[n_0] = \sigma_w^2 \delta[n_0] &\xrightarrow{\quad} \Gamma_x[n_0] = \begin{cases} 0 & q < n_0 \\ \sigma_w^2 \sum_{k=0}^{q-n_0} b_k b_{k+n_0} & 0 \leq n_0 \leq q \\ \Gamma_x(-n_0) & n_0 < 0 \end{cases} \xrightarrow{\quad} \text{It has limited support!!}
 \end{aligned}$$

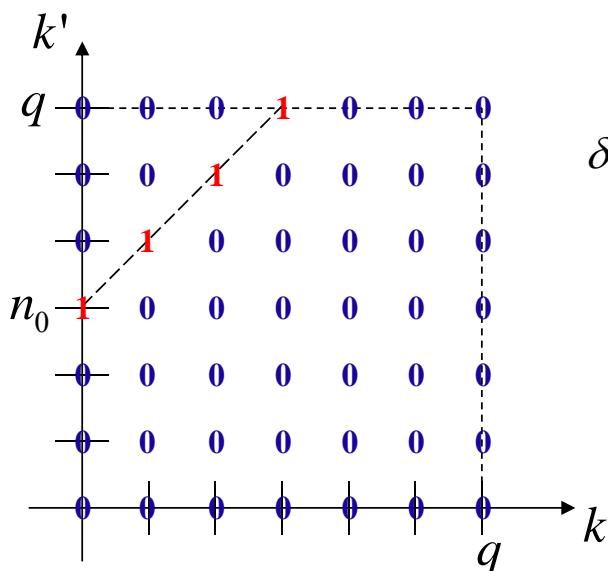
Proof

$$\begin{aligned}
 \Gamma_x[n_0] &= E\{x[n]x[n+n_0]\} = E\left\{\left(\sum_{k=0}^q b_k w[n-k]\right)\left(\sum_{k'=0}^q b_{k'} w[n+n_0-k']\right)\right\} = \sum_{k=0}^q \sum_{k'=0}^q b_k b_{k'} E\{w[n-k]w[n+n_0-k']\} \\
 &= \sum_{k=0}^q \sum_{k'=0}^q b_k b_{k'} E\{w[n']w[n'+n_0-k'+k]\} = \sum_{k=0}^q \sum_{k'=0}^q b_k b_{k'} \sigma_w^2 \delta[n_0 - (k' - k)]
 \end{aligned}$$

3. Moving average processes: MA(q)

Statistical properties

Proof (contd.)



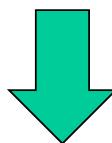
$$\Gamma_X[n_0] = \dots = \sum_{k=0}^q \sum_{k'=0}^q b_k b_{k'} \sigma_w^2 \delta[n_0 - (k' - k)] = \begin{cases} 0 & q < n_0 \\ \sigma_w^2 \sum_{k=0}^{q-n_0} b_k b_{k+n_0} & 0 \leq n_0 \leq q \\ \Gamma_X(-n_0) & n_0 < 0 \end{cases}$$

$$\delta[n_0 - (k' - k)] = \begin{cases} 1 & n_0 - (k' - k) = 0 \longrightarrow k' = k + n_0 \\ 0 & n_0 - (k' - k) \neq 0 \end{cases}$$

3. Moving average processes: MA(q)

(Brute force) determination of the MA parameters

$$\Gamma_x[n_0] = \begin{cases} 0 & q < n_0 \\ \sigma_w^2 \sum_{k=0}^{q-n_0} b_k b_{k+n_0} & 0 \leq n_0 \leq q \\ \Gamma_x(-n_0) & n_0 < 0 \end{cases}$$



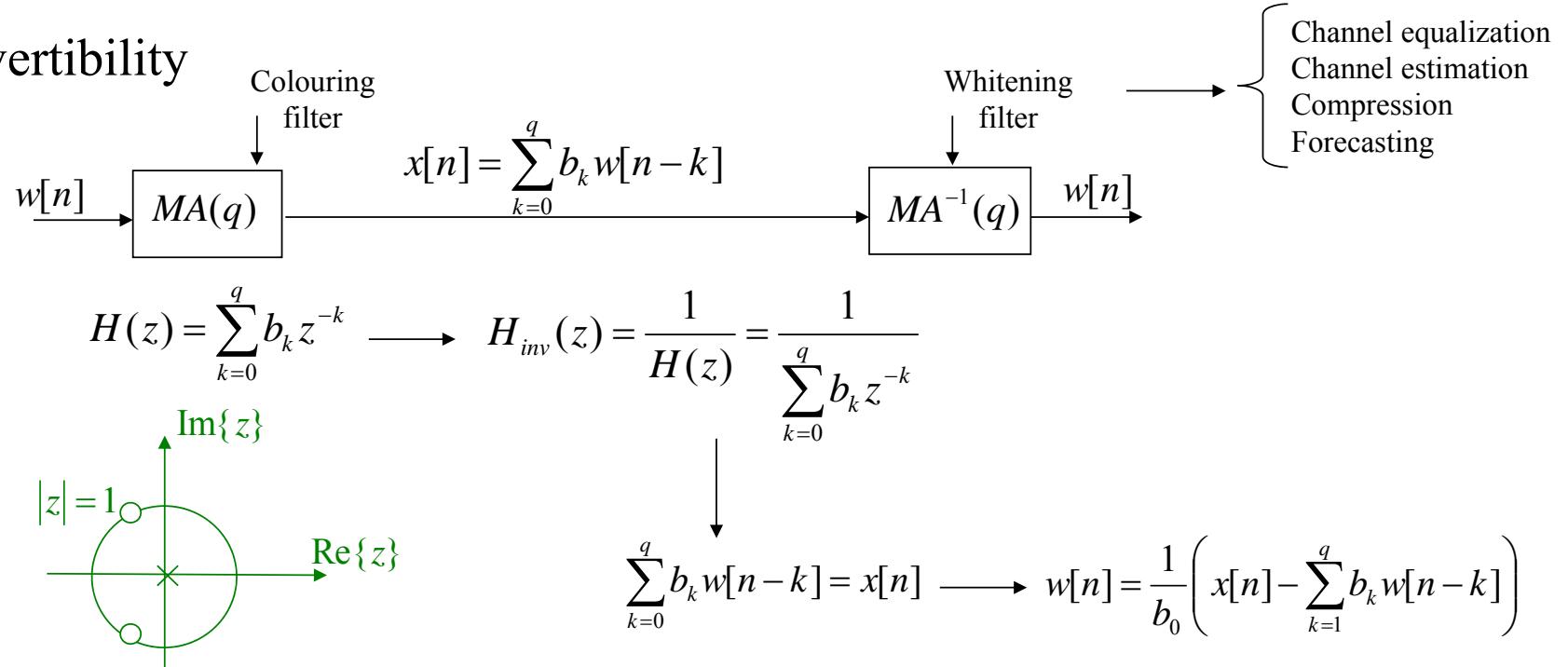
$$r_x[0] = \sigma_w^2 (b_0^2 + b_1^2 + b_2^2)$$

$$r_x[1] = \sigma_w^2 (b_0 b_1 + b_1 b_2)$$

$$r_x[2] = \sigma_w^2 (b_0 b_2)$$

3. Moving average processes: MA(q)

Invertibility



Example: $y[n] = \frac{x[n] + x[n-1] + x[n-2]}{3}$ does not have a stable, causal inverse

$y[n] = x[n] + 0.9x[n-1]$ has a stable, causal inverse

3. Moving average processes: generalizations

Model not restricted to be causal

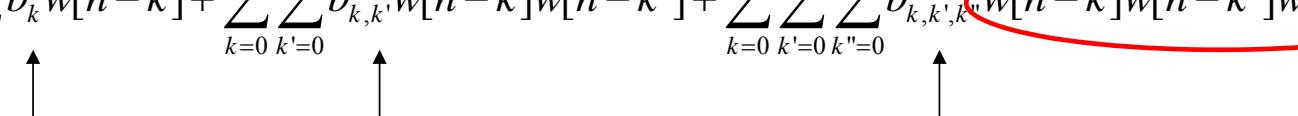
$$x[n] = b_{q_0} w[n - q_0] + b_{q_0+1} w[n - q_0 + 1] + \dots + b_1 w[n - 1] + b_0 w[n] \quad \text{Causal component}$$

$+ b_{-1} w[n + 1] + \dots + b_{q_F} w[n + q_F]$ Anticausal component

Model not restricted to be linear

$$1) \quad x[n] = \sum_{k=0}^q b_k w[n-k] + \sum_{k=0}^q \sum_{k'=0}^{q'} b_{k,k'} w[n-k]w[n-k']$$

Quadratic component

$$x[n] = \sum_{k=0}^q b_k w[n-k] + \sum_{k=0}^q \sum_{k'=0}^{q'} b_{k,k'} w[n-k]w[n-k'] + \sum_{k=0}^q \sum_{k'=0}^{q'} \sum_{k''=0}^{q''} b_{k,k',k''} w[n-k]w[n-k']w[n-k'']$$


Volterra Kernels

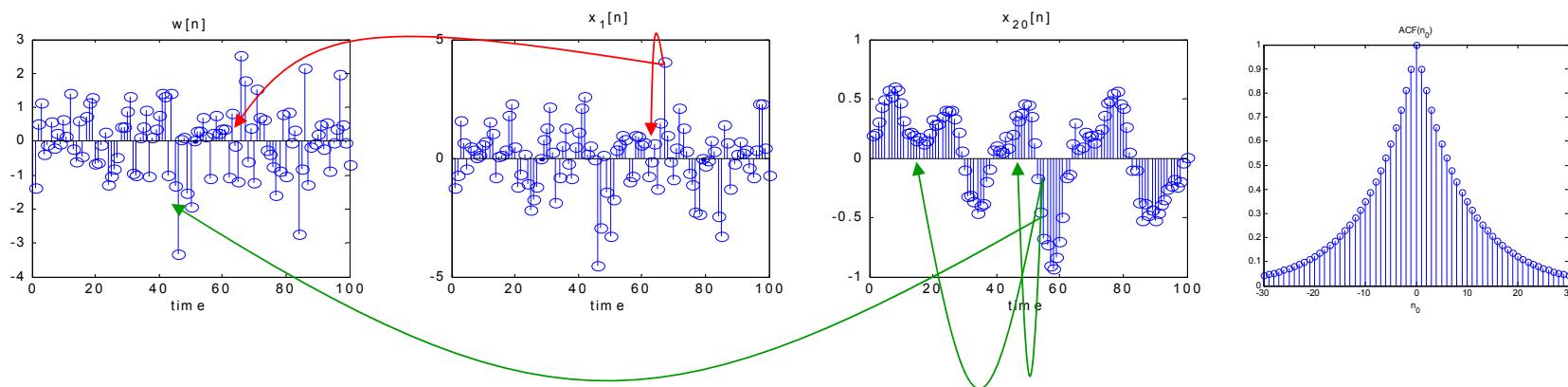
$$2) \quad x[n] = \sum_{k=0}^q b_k f(w[n-k])$$

4. Autoregressive processes: AR(p)

Definition

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p}} = \frac{1}{A(z)} \quad \begin{matrix} \longrightarrow \\ \uparrow \downarrow \\ \end{matrix} \quad \text{LTI, with memory, invertible, causal, stable}$$

$$w[n] \rightarrow \boxed{AR(p)} \rightarrow x[n] = w[n] + a_1 x[n-1] + \dots + a_p x[n-p]$$



4. Autoregressive processes: AR(p)

Definition

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p}} = \frac{1}{A(z)} \quad \xrightarrow{\text{LTI, with memory, invertible, causal, stable}}$$

↑

$$w[n] \xrightarrow{\boxed{AR(p)}} x[n] = w[n] + a_1 x[n-1] + \dots + a_p x[n-p]$$

Statistical properties

$$N(0, \sigma_w^2) \longrightarrow N(0, \Gamma_x[0])$$

$$\Gamma_x[n_0] = \sigma_w^2 \delta[n_0] + \sum_{k=1}^p a_k \Gamma_x[n_0 - k] \quad \text{Yule-Walker equations}$$

$$\text{Whose solution is } \Gamma_x[n_0] = \sum_{k=1}^p A_k z_k^{|n_0|} \quad z_k \equiv \text{Poles of } H(z)$$

Relationship to MA processes

$$H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p}} = 1 + \sum_{k=1}^{\infty} b_k z^{-k}$$

↑
Laurent series

4. Autoregressive processes: AR(p)

Determination of the constants A_k

$$\Gamma_x[n_0] = \sum_{k=1}^p A_k z_k^{|n_0|} \longrightarrow r_x[n_0] = \sum_{k=1}^p A'_k z_k^{|n_0|}$$

$$\Gamma_x[n_0] = \sigma_w^2 \delta[n_0] + \sum_{k=1}^p a_k \Gamma_x[n_0 - k] \longrightarrow r_x[n_0] = \sum_{k=1}^p a_k r_x[n_0 - k] \quad n_0 > 0$$

Example:

$$x[n] = w[n] + a_1 x[n-1] + a_2 x[n-2] \longrightarrow H(z) = \frac{1}{1 - a_1 z^{-1} - a_2 z^{-2}}$$

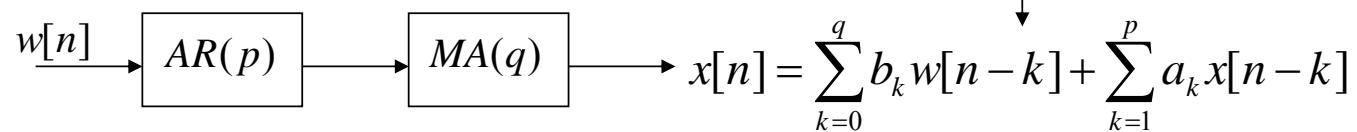
Poles: $z_1, z_2 = \frac{a_1 \pm \sqrt{a_1^2 + 4a_2}}{2} \longrightarrow |z_i| < 1 \Rightarrow a_2 > -1, a_1 + a_2 < 1, a_1 - a_2 > -1$
 $z_i \in R \Rightarrow a_1^2 + 4a_2 > 0$

$$r_x[n_0] = A'_1 z_1^{|n_0|} + A'_2 z_2^{|n_0|} \longrightarrow \begin{cases} r_x[0] = A'_1 + A'_2 = 1 \\ r_x[1] = A'_1 z_1 + A'_2 z_2 = a_1 r_x[0] + a_2 r_x[-1] \end{cases}$$

5. Autoregressive, Moving average: ARMA(p,q)

Definition

$$H(z) = H_{MA(q)}(z)H_{AR(p)}(z) = \frac{b_0 + b_1 z^{-1} + b_2 z^{-2} + \dots + b_q z^{-q}}{1 - a_1 z^{-1} - a_2 z^{-2} - \dots - a_p z^{-p}} = \frac{B(z)}{A(z)}$$



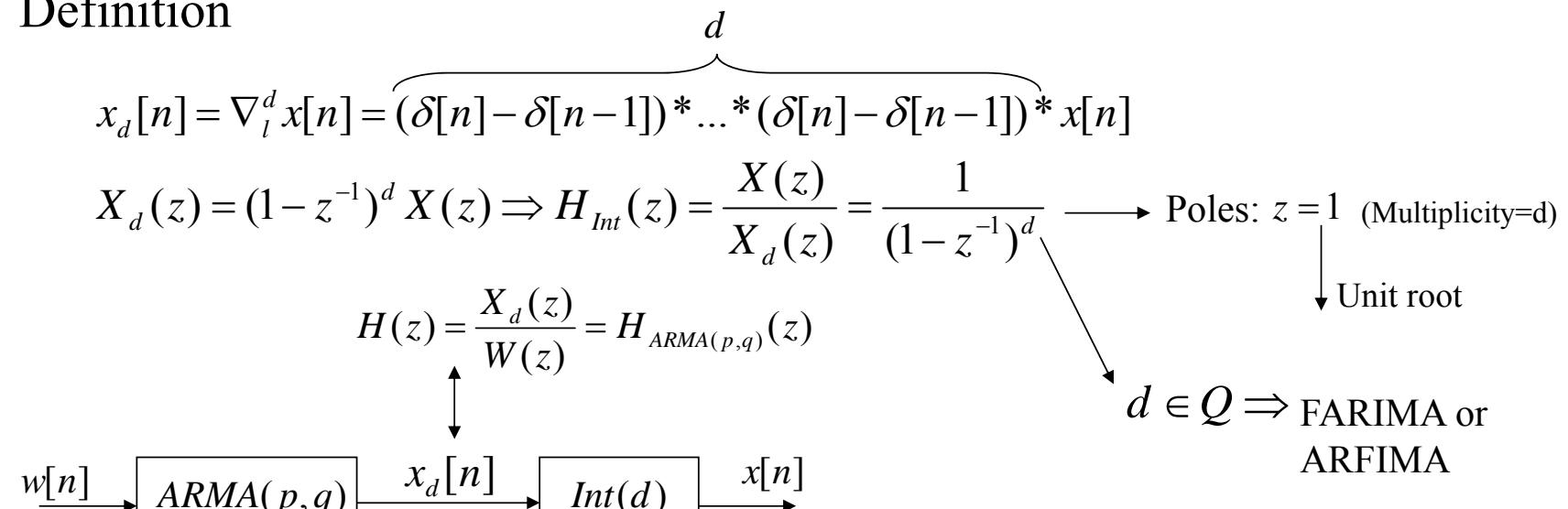
Statistical properties

$$N(0, \sigma_w^2) \longrightarrow N(0, \Gamma_x[0])$$

$$\Gamma_x[n_0] = \sigma_w^2 \sum_{k=0}^q b_k h[k - n_0] - \sum_{k=1}^p a_k \Gamma_x[n_0 - k]$$

6. Autoregressive, Integrated, Moving Average: ARIMA(p,d,q)

Definition



$$H_{ARIMA(p,d,q)}(z) = \frac{X(z)}{W(z)} = \frac{X_d(z)}{W(z)} \frac{X(z)}{X_d(z)} = H_{ARMA(p,q)}(z) H_{Int(d)}(z)$$

Example for $d=1$: $(x[n] - x[n-1]) = \sum_{k=0}^q b_k w[n-k] + \sum_{k=1}^p a_k (x[n-k] - x[n-k-1])$

7. Seasonal ARIMA: SARIMA(p,d,q)x(P,D,Q)_s (Box-Jenkins model)

Definition

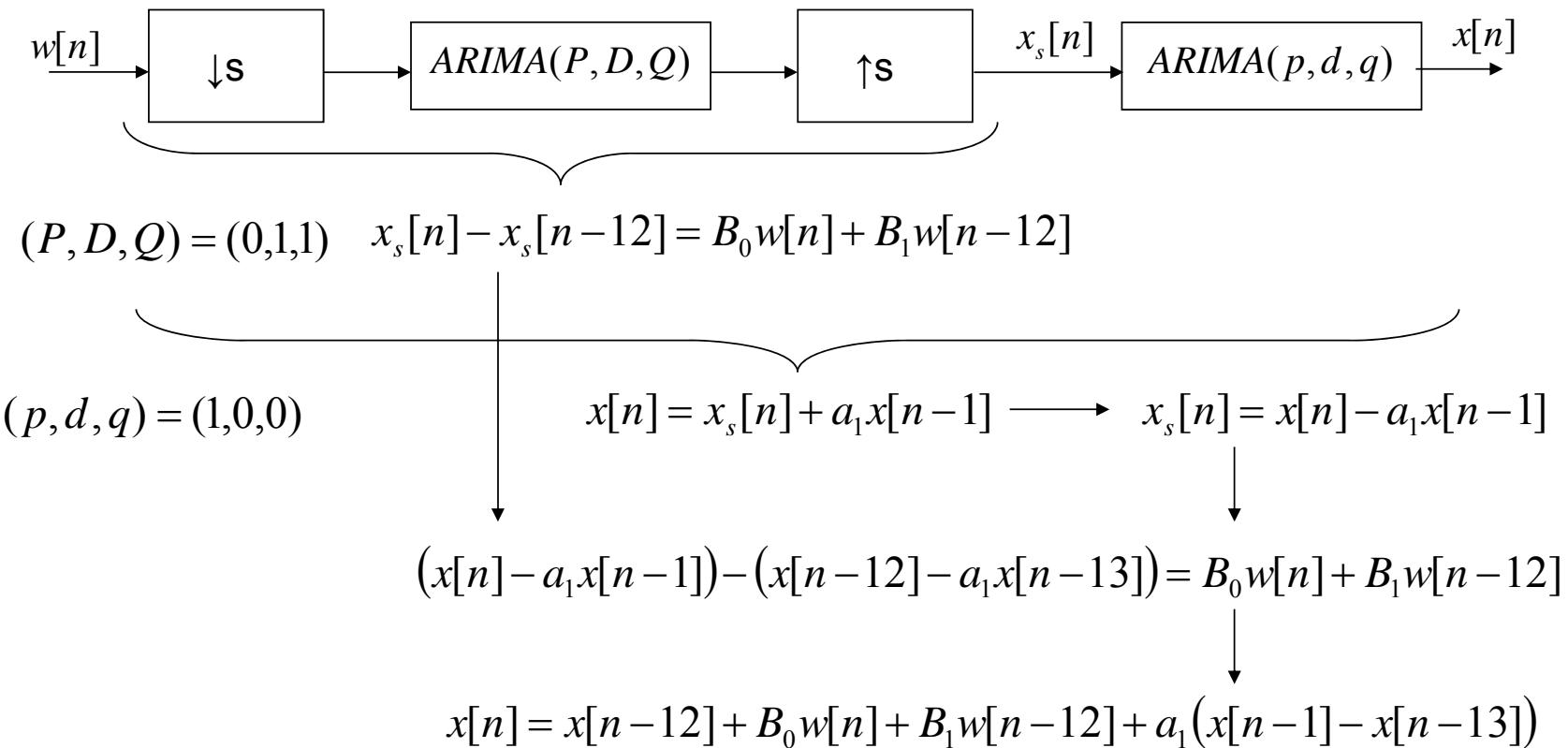
$$\begin{aligned}
 d = 1 \quad & (x[n] - x[n-1]) = \sum_{k=0}^q b_k x_s[n-k] + \sum_{k=1}^p a_k (x[n-k] - x[n-k-1]) \\
 D = 1 \quad & (x_s[n] - x_s[n-s]) = \sum_{k=0}^Q B_k w[n-ks] + \sum_{k=1}^P A_k (x_s[n-ks] - x_s[n-(k-1)s]) \\
 w[n] \rightarrow & \boxed{\downarrow s} \rightarrow \boxed{ARIMA(P, D, Q)} \rightarrow \boxed{\uparrow s} \rightarrow \boxed{ARIMA(p, d, q)} \rightarrow x[n]
 \end{aligned}$$

$$\frac{X_s(z)}{W(z)} = H_{ARIMA(P,D,Q)}(z^s) \quad \frac{X(z)}{X_s(z)} = H_{ARIMA(p,d,q)}(z)$$

$$H_{SARIMA(p,d,q)\times(P,D,Q)_s}(z) = \frac{X(z)}{W(z)} = H_{ARIMA(P,D,Q)}(z^s)H_{ARIMA(p,d,q)}(z)$$

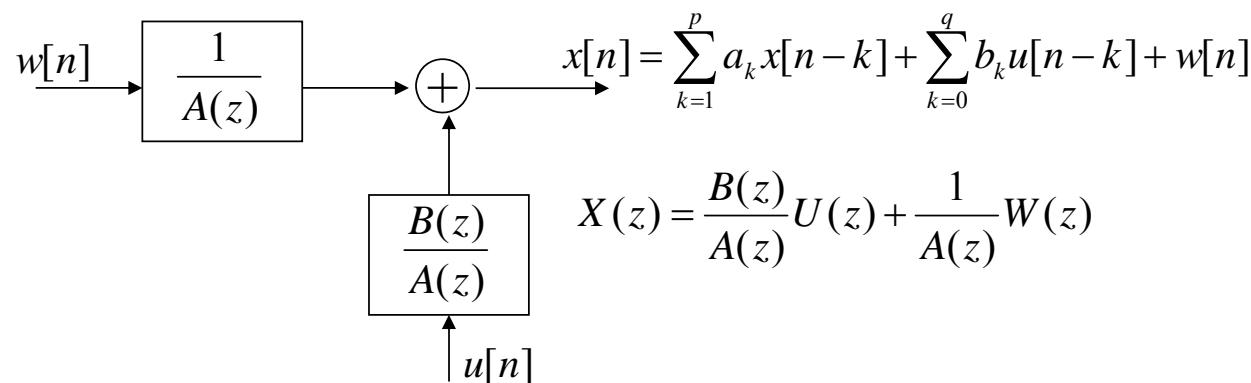
7. Seasonal ARIMA: SARIMA(p,d,q)x(P,D,Q)_s

Example: SARIMA(1,0,0)x(0,1,1)₁₂

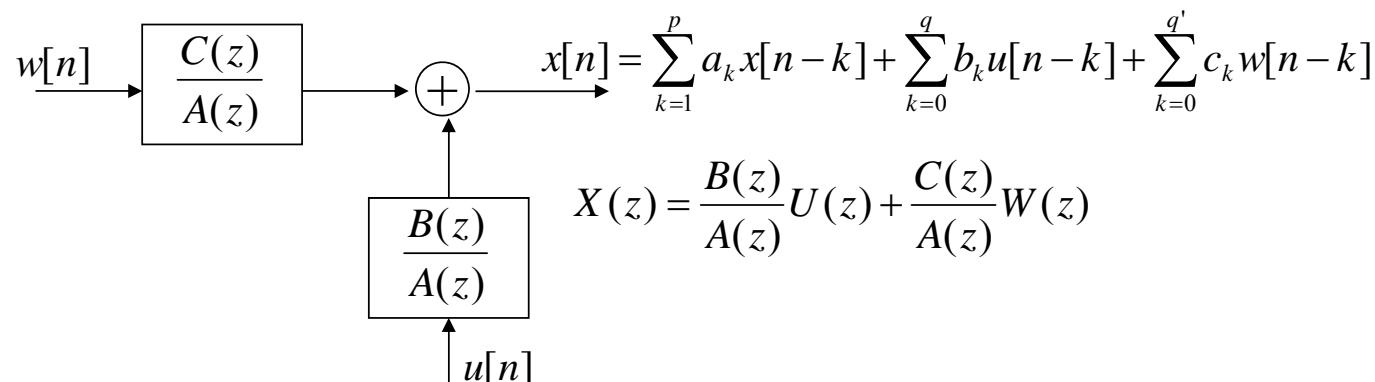


8. Known external inputs: System identification

ARX

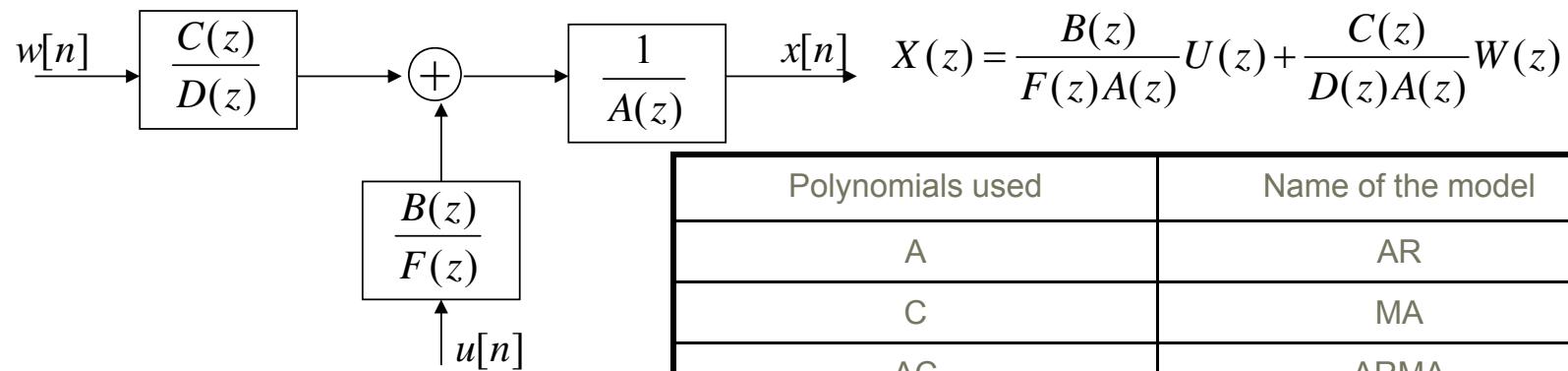


ARMAX



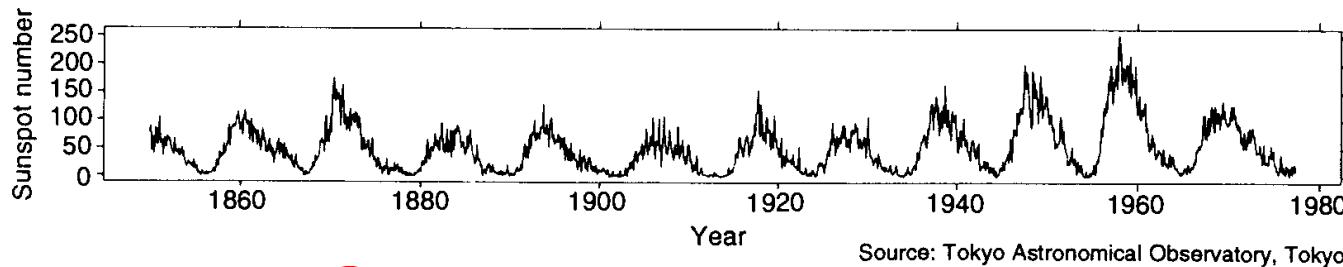
9. A family of models

General model



Polynomials used	Name of the model
A	AR
C	MA
AC	ARMA
ACD	ARIMA
AB	ARX
ABC	ARMAX
ABD	ARARX
ABCD	ARARMAX
BFCD	Box-Jenkins

10. Nonlinear models



Nonlinear AR: $x[n] = f(x[n-1], x[n-2], \dots, x[n-p]) + w[n]$

Time-varying AR: $x[n] = \sum_{k=1}^p a_k[n]x[n-k] + w[n]$

Smooth transition: $x[n] = \left(\sum_{k=1}^p a_k x[n-k] \right) p[n] + \left(\sum_{k=1}^p a_k x[n-k] \right) (1 - p[n]) + w[n]$

Random coeff. AR: $x[n] = \sum_{k=1}^p (a_k + \varepsilon[n])x[n-k] + w[n]$

Bilinear models $x[n] = \sum_{k=1}^p a_k x[n-k] + \sum_{k=1}^q b_k x[n-k]w[n-k-M] + w[n]$

10. Nonlinear models

Threshold AR (TAR):
$$x[n] = \begin{cases} \sum_{k=1}^p a_k^{(1)} x[n-k] + w[n] & x[n-d] \leq t \\ \sum_{k=1}^p a_k^{(2)} x[n-k] + w[n] & x[n-d] > t \end{cases}$$

Smooth TAR (STAR):
$$x[n] = \sum_{k=1}^p a_k^{(1)} x[n-k] + \left(\sum_{k=1}^p a_k^{(2)} x[n-k] \right) S(x[n-d]) + w[n]$$

Heterocedastic model: $x[n] = \sigma[n] \eta[n] \xrightarrow{N(0,1)}$ Random walk

ARCH $\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^p a_k x^2[n-k]$

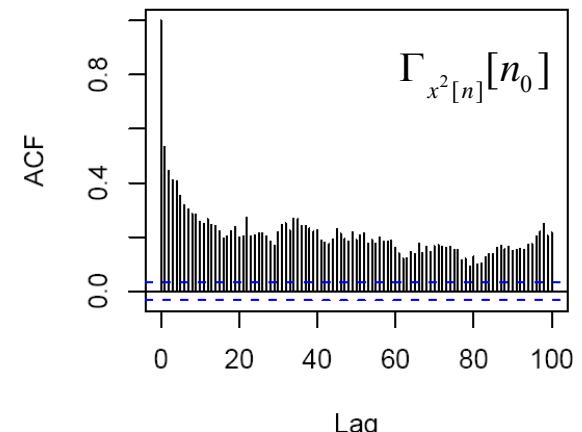
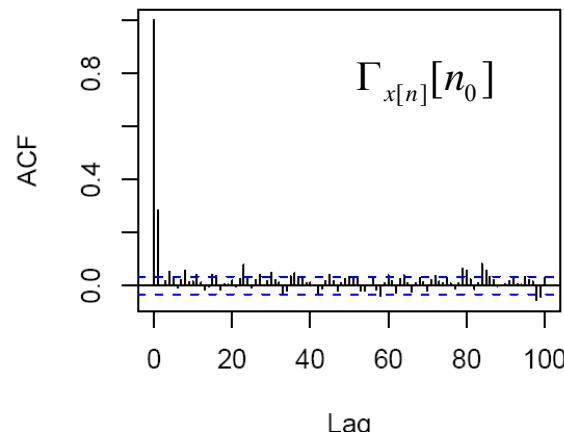
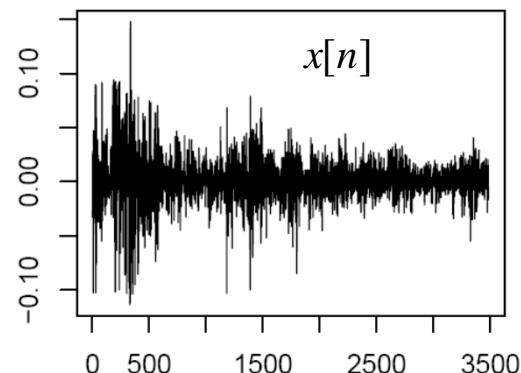
GARCH $\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^p a_k x^2[n-k] + \sum_{k=1}^q b_k \sigma^2[n-k]$

(Neural networks)

(Chaos)

10. Nonlinear models

GARCH $\sigma^2[n] = \sigma_0^2 + \sum_{k=1}^p a_k x^2[n-k] + \sum_{k=1}^q b_k \sigma^2[n-k]$



Properties

The model is unique and stationary if $\sum_{k=1}^p a_k + \sum_{k=1}^q b_k < 1$

Zero mean $E\{x[n]\} = 0$

Lack of correlation $\Gamma_{x[n]}[n_0] = \frac{\sigma_0^2}{1 - \sum_{k=1}^p a_k + b_k} \delta[n_0]$

10. Nonlinear models

$$\text{GARCH} \quad \sigma^2[n] = \sigma_0^2 + \sum_{k=1}^p a_k x^2[n-k] + \sum_{k=1}^q b_k \sigma^2[n-k]$$

Estimation through Maximum Likelihood

Forecasting

$$\hat{x}^2[n+h] = \sigma_0^2 + \sum_{k=1}^{\min\{p,q\}} (a_k + b_k) \hat{x}^2[n+h-k] - \sum_{k=1}^q b_k z[n+h-k]$$
$$\begin{aligned} \hat{x}^2[n+h-1] &= \sigma_0^2 + \dots \\ \hat{x}^2[n+h-2] &= \sigma_0^2 + \dots \\ &\vdots \\ x^2[n] & \\ x^2[n-1] & \end{aligned} \left. \begin{array}{l} \\ \\ \dots \\ \left. \begin{array}{c} \text{Observed} \end{array} \right. \end{array} \right\}$$
$$\begin{aligned} z[n+h-1] &= 0 \\ z[n+h-2] &= 0 \\ &\vdots \\ z[n] &= x^2[n] - \sigma^2[n] \\ z[n-1] &= x^2[n-1] - \sigma^2[n-1] \end{aligned} \right\}$$

GARCH(1,1)

$$\hat{x}^2[n+1] = \sigma_0^2 + a_1 x^2[n] + b_1 \sigma^2[n]$$

$$\hat{x}^2[n+h] = \sigma_0^2 \sum_{k=0}^{h-1} (a_1 + b_1)^k + a_1 (a_1 + b_1) x^2[n] + b_1 (a_1 + b_1)^{h-1} \sigma^2[n]$$

10. Nonlinear models

Extensions of GARCH

Exponential GARCH (EGARCH)

$$\log \sigma^2[n] = \log \sigma_0^2 + \sum_{k=1}^p a_k \log x^2[n-k] + \sum_{k=1}^q b_k \log \sigma^2[n-k]$$

Integrated GARCH (IGARCH)

$$\sum_{k=1}^p a_k + \sum_{k=1}^q b_k < 1 \longrightarrow \sum_{k=1}^p a_k + \sum_{k=1}^q b_k = 1$$

GARCH IGARCH

11. Parameter estimation

$$\text{AR(1)} \quad x[n] = a_1 x[n-1] + w[n]$$

Assume that we observe $(x[1], x[2], \dots, x[N])$

Maximum Likelihood Estimates (MLE)

$$x[n] = a_1 x[n-1] + w[n]$$

$N(0, \sigma_w^2)$ ↑

$$\Gamma_w[n_0] = \sigma_w^2 \delta[n_0]$$

→ $X_n = a_1 X_{n-1} + W_n = a_1(a_1 X_{n-2} + W_{n-1}) + W_n = \dots$

$= W_n + a_1 W_{n-1} + a_1^2 W_{n-2} + a_1^3 W_{n-3} + \dots$

$E\{X_n\} = 0$

$$X_1 | \theta \rightarrow N\left(0, \frac{\sigma_w^2}{1-a_1^2}\right) \quad \theta = \{a_1, \sigma_w^2\} \quad E\{X_n\} = E\{W_n + a_1 W_{n-1} + a_1^2 W_{n-2} + a_1^3 W_{n-3} + \dots\} = \sigma_w^2 (1 + a_1^2 + a_1^4 + a_1^6 + \dots) = \frac{\sigma_w^2}{1-a_1^2}$$

$$X_2 = a_1 X_1 + W_2 \rightarrow W_2 = X_2 - a_1 X_1 \rightarrow X_2 | X_1, \theta \sim N\left(a_1 x_1, \frac{\sigma_w^2}{1-a_1^2}\right)$$



11. Parameter estimation

Maximum Likelihood Estimates (MLE)

$$\begin{aligned}
 X_1 | \theta &\rightarrow N\left(0, \frac{\sigma_w^2}{1-a_1^2}\right) \\
 X_2 | X_1, \theta &\rightarrow N\left(a_1 x_1, \frac{\sigma_w^2}{1-a_1^2}\right) \\
 X_3 | X_2, X_1, \theta &\rightarrow N\left(a_1 x_2, \frac{\sigma_w^2}{1-a_1^2}\right) \\
 &\dots
 \end{aligned}
 \quad \left. \right\} \rightarrow$$

$$\begin{aligned}
 f_{X_1 X_2 \dots X_N | \theta}(x_1, x_2, \dots, x_N) &= \\
 &= f_{X_1 | \theta}(x_1) f_{X_2 | X_1, \theta}(x_2) f_{X_3 | X_2, \theta}(x_3) \dots f_{X_N | X_{N-1}, \theta}(x_N)
 \end{aligned}$$

$$L(\theta) = \log f_{X_1 X_2 \dots X_N | \theta}(x_1, x_2, \dots, x_N) = -\frac{1}{2} \log(2\pi) - \frac{1}{2} \log \frac{\sigma_w^2}{1-a_1^2} - \frac{x_1^2}{2 \frac{\sigma_w^2}{1-a_1^2}} - \frac{N-1}{2} \log(2\pi\sigma_w^2) - \frac{1}{2} \sum_{n=2}^N \left(\frac{x_n - a_1 x_{n-1}}{\sigma_w^2} \right)^2$$

$$\hat{a}_1, \hat{\sigma}_w^2 = \arg \max_{a_1} L(\theta) \Rightarrow \frac{\partial L(\theta)}{\partial a_1} = 0 = \frac{\partial L(\theta)}{\partial \sigma_w^2}$$

→ Confidence intervals

→ Numerical, iterative solution

11. Parameter estimation

$$\left. \begin{array}{l} x[n] = a_1 x[n-1] + w[n] \\ \hat{x}[n] = a_1 x[n-1] \end{array} \right\} x[n] = \hat{x}[n] + w[n]$$

Least Squares Estimates (LSE)

$$w[n] = x[n] - \hat{x}[n]$$

$$E\{w[n]\} = 0$$

$$\sigma_w^2 = E\{w^2[n]\} = \Gamma_x[0] + 2a_1\Gamma_x[1] + a_1^2\Gamma_x[0]$$

$$\boxed{\frac{\partial \sigma_w^2}{\partial a_1} = 0 = 2\Gamma_x[1] + 2a_1\Gamma_x[0]} \Rightarrow a_1 = \boxed{-\frac{\Gamma_x[1]}{\Gamma_x[0]}} \Rightarrow \boxed{\sigma_w^2 = \Gamma_x[0] \left(\frac{\Gamma_x^2[0]}{\Gamma_x^2[1]} - 1 \right)}$$

12. Order selection

If I have to fit a model ARMA(p,q), what are the p and q values I have to supply?

- ACF/PACF analysis
- Akaike Information Criterion

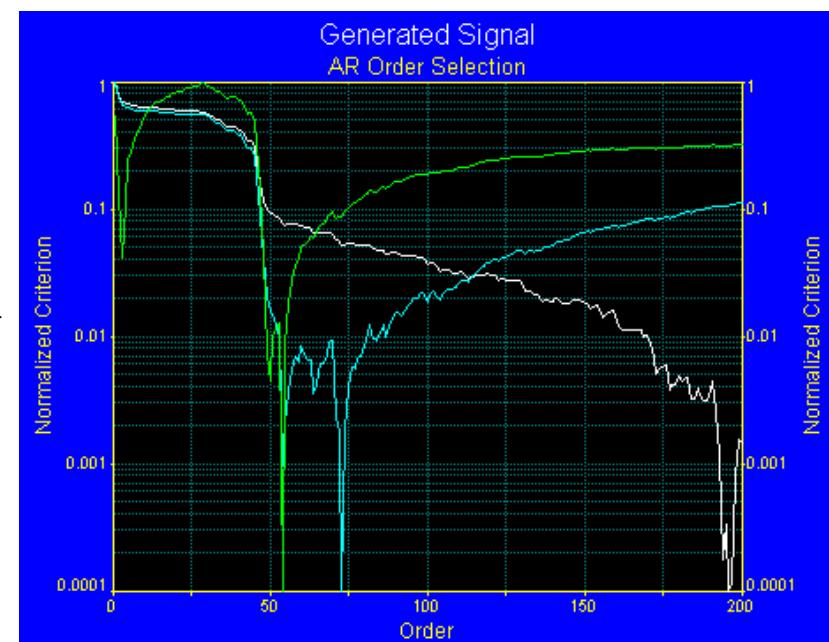
$$AIC(p, q) = \log \sigma_w^2 + (p + q) \frac{2}{N}$$

- Bayesian Information Criterion

$$BIC(p, q) = \log \sigma_w^2 + (p + q) \frac{\log N}{N}$$

- Final Prediction Error

$$FPE(p) = \frac{N + p}{N - p} \sigma_w^2$$



12. Order selection

Partial correlation coefficients (PACF, First-order correlations)

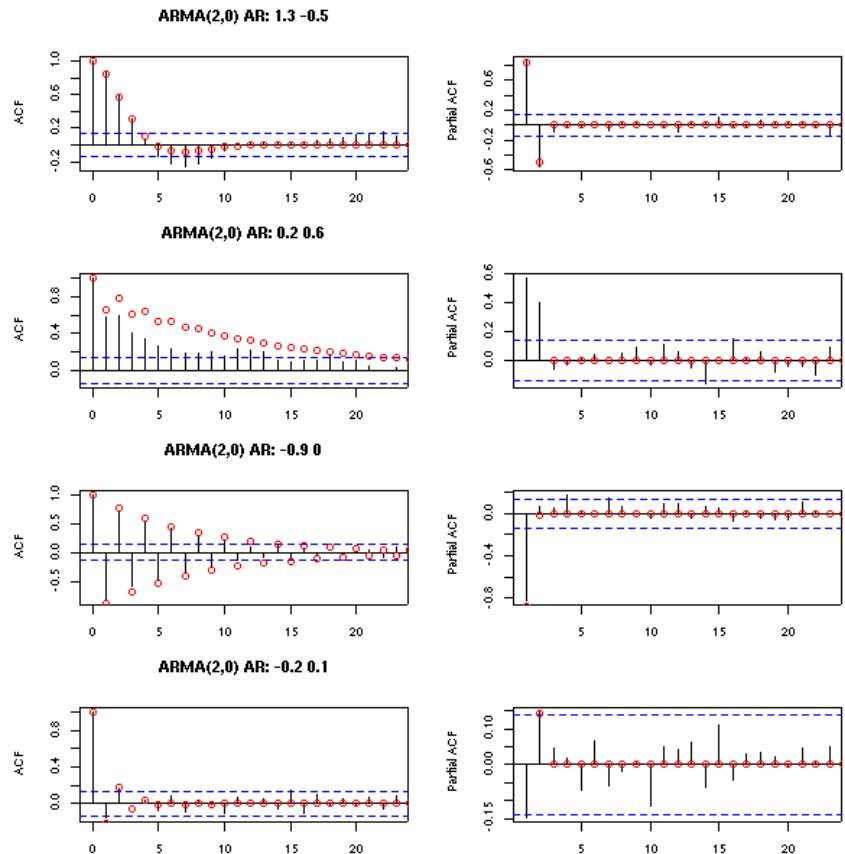
$$x[n] = \phi_{n_0,1}x[n-1] + \phi_{n_0,2}x[n-2] + \dots + \phi_{n_0,n_0}x[n-n_0] + w[n]$$

$$r[n_0] = \sum_{n=1}^{n_0} \phi_{n_0,n} r[n_0 - n]$$

$$\begin{pmatrix} r[0] & r[1] & r[2] & \dots & r[n_0-2] & r[n_0-1] \\ r[1] & r[0] & r[1] & \dots & r[n_0-3] & r[n_0-2] \\ r[2] & r[1] & r[0] & \dots & r[n_0-4] & r[n_0-3] \\ \dots & \dots & \dots & \dots & \dots & \dots \\ r[n_0-2] & r[n_0-3] & r[n_0-4] & \dots & r[0] & r[1] \\ r[n_0-1] & r[n_0-1] & r[n_0-2] & \dots & [1] & r[0] \end{pmatrix} \begin{pmatrix} \phi_{n_0,1} \\ \phi_{n_0,2} \\ \phi_{n_0,3} \\ \dots \\ \phi_{n_0,n_0-1} \\ \phi_{n_0,n_0} \end{pmatrix} = \begin{pmatrix} r[1] \\ r[2] \\ r[3] \\ \dots \\ r[n_0-1] \\ r[n_0] \end{pmatrix}$$

Yule-Walker
equations

12. Order selection



Thumb rule

- ARMA(1,0): ACF: exponential decrease; PACF: one peak
- ARMA(2,0): ACF: exponential decrease or waves; PACF: two peaks
- ARMA(p,0): ACF: unlimited, decaying; PACF: limited
- ARMA(0,1): ACF: one peak; PACF: exponential decrease
- ARMA(0,2): ACF: two peaks; PACF: exponential decrease or waves
- ARMA(0,q): ACF: limited; PACF: unlimited decaying
- ARMA(1,1): ACF&PACF: exponential decrease
- ARMA(p,q): ACF: unlimited; PACF: unlimited

13. Model checking

Residual Analysis

Example: ARMA (1,1)

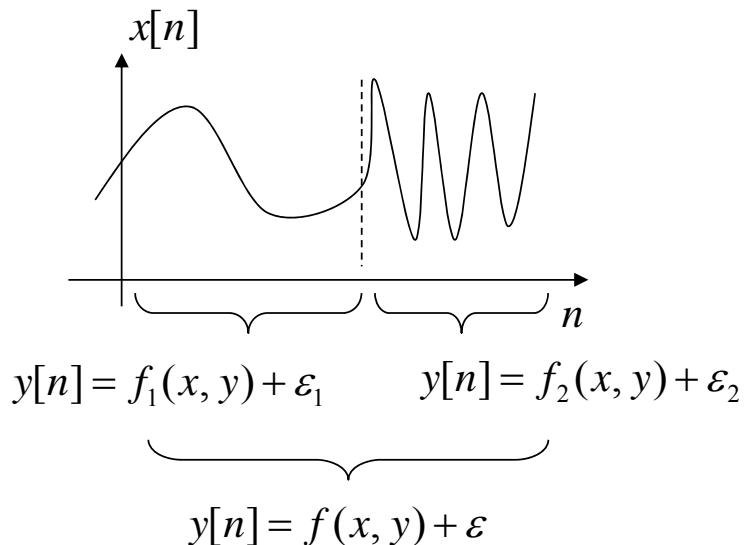
$$x[n] = ax[n-1] + b_0 w[n] + b_1 w[n-1] \Rightarrow \begin{aligned} \hat{w}[n] &= \frac{1}{b_0} (x[n] - ax[n-1] - b_1 \hat{w}[n-1]) \\ \hat{w}[0] &= 0 \end{aligned}$$

Assumptions

1. Gaussianity:
 1. The input random signal $w[n]$ is univariate normal with zero mean
 2. The output signal, $x[n]$ (the time series being studied), is multivariate normal and its covariance structure is fully determined by the model structure and parameters
2. Stationarity: $x[n]$ is stationary once that the necessary operations to produce a stationary signal have been carried out.
3. Residual independency: the input random signal $w[n]$ is independent of all previous samples.

13. Model checking

Structural changes



Chow test:

$$H_0 : f_1 = f_2$$

$$H_1 : f_1 \neq f_2$$

$$S_i = \sum_{n=1}^N (y[n] - \hat{y}_i[n])^2 \leftarrow \text{Sum of squares of the residuals with model } i$$

$$F = \frac{\frac{1}{k}(S - (S_1 + S_2))}{\frac{1}{N_1 + N_2 - 2k}(S_1 + S_2)} \sim F(k, N_1 + N_2 - 2k)$$

↑ ↑
Number of parameters in the model Number of samples in each period

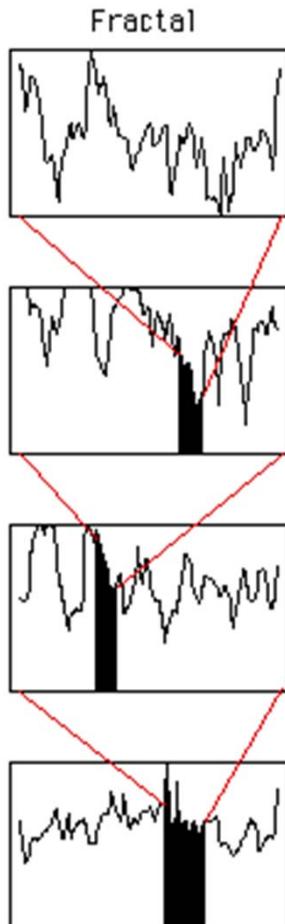
Assumption: variance is the same in both regions
Solution: Robust standard errors

13. Model checking

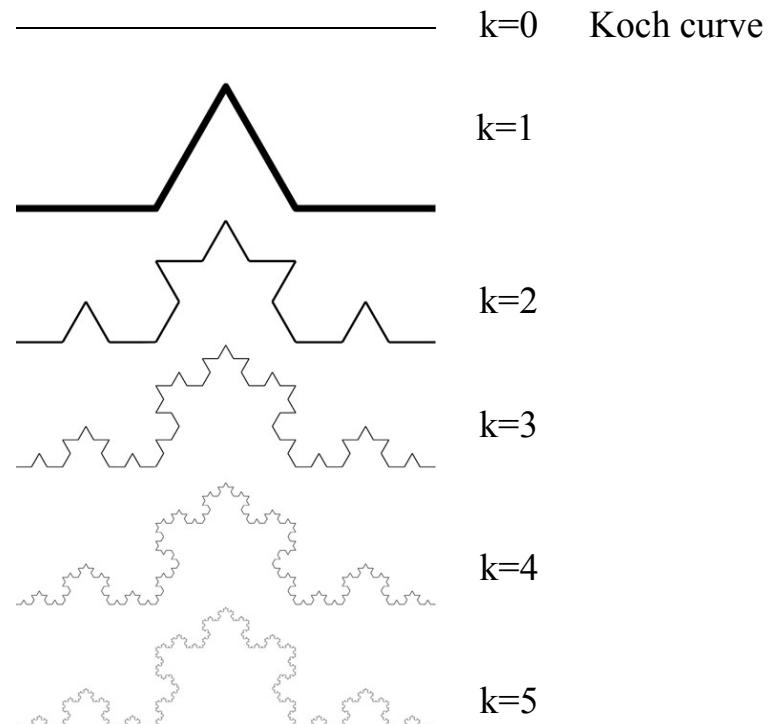
Diagnostic checking

1. Compute and plot the residual error
2. Check that its mean is approximately zero
3. Check for the randomness of the residual, i.e., there are no time intervals where the mean is significantly different from zero (intervals where the residual is systematically positive or negative).
4. Check that the residual autocorrelation is not significantly different from zero for all lags
5. Check that the residual is normally distributed.
6. Check if there are residual outliers.
7. Check the ability of the model to predict future samples

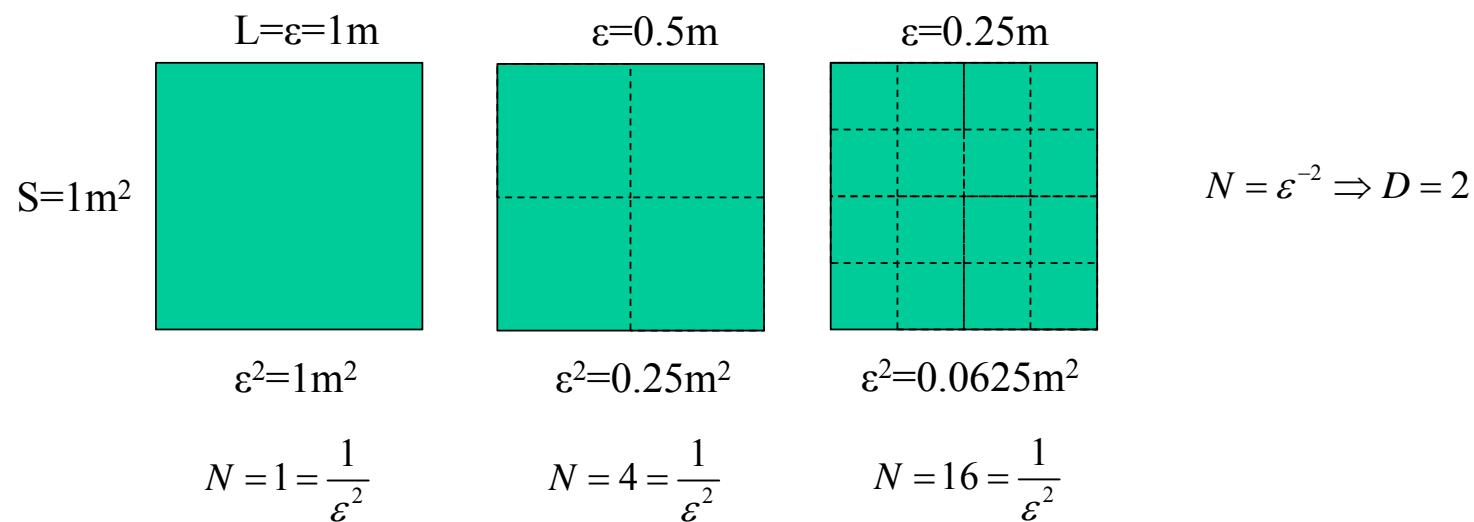
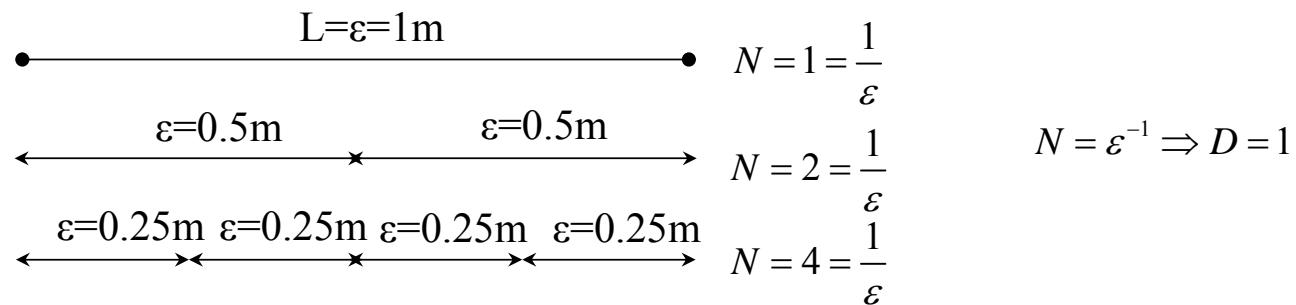
14. Self-similarity, Fractal dimension, Chaos theory



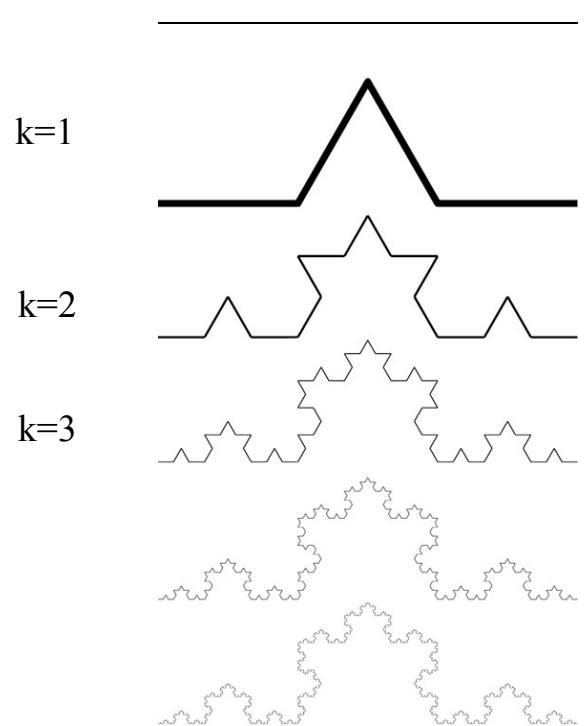
Intuitively a fractal is a curve that is self-similar at all scales



14. Self-similarity, Fractal dimension, Chaos theory



14. Self-similarity, Fractal dimension, Chaos theory



$$\varepsilon = 1m \quad N = \varepsilon^{-D} \Rightarrow 1 = 1^{-D}$$

$$\varepsilon = \frac{1}{3}m \quad N = \varepsilon^{-D} \Rightarrow 4 = \left(\frac{1}{3}\right)^{-D}$$

$$\varepsilon = \frac{1}{3^2}m \quad N = \varepsilon^{-D} \Rightarrow 4^2 = \left(\frac{1}{3^2}\right)^{-D}$$

$$\varepsilon = \frac{1}{3^3}m \quad N = \varepsilon^{-D} \Rightarrow 4^3 = \left(\frac{1}{3^3}\right)^{-D}$$

...

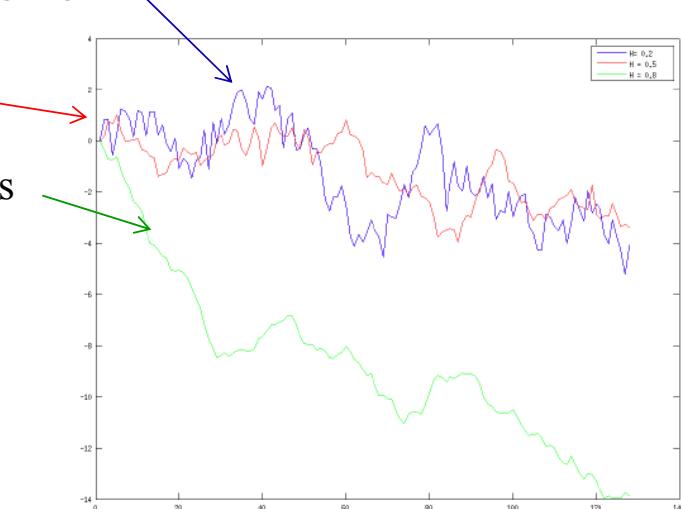
$$N = \varepsilon^{-D} \Rightarrow 4^k = \left(\frac{1}{3^k}\right)^{-D} \Rightarrow D = \frac{\log 4}{\log 3} = 1.26$$

14. Self-similarity, Fractal dimension, Chaos theory

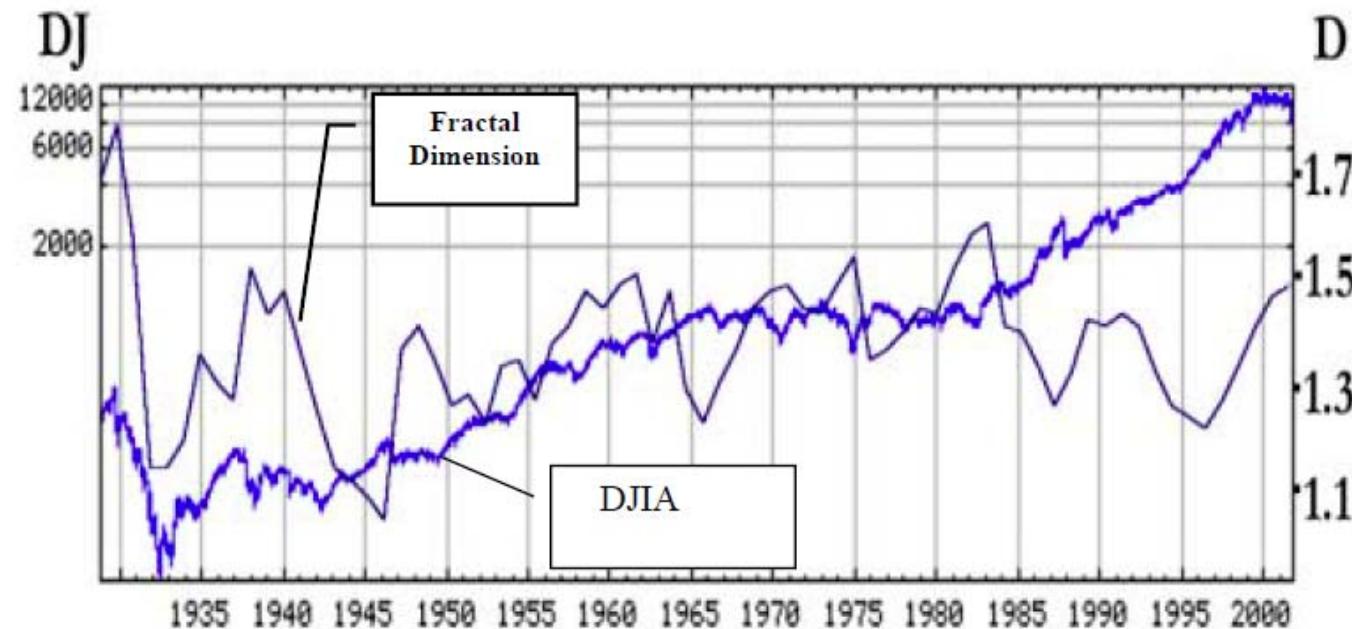
$$S_X(\omega) \propto \omega^{-\alpha} \quad \leftarrow \quad \alpha = 5 - 2D$$
$$H = 2 - D \quad \text{Hurst exponent}$$

$$H \in [0,1] = \begin{cases} 0.2 & \text{Long-range dependent alternating signs} \\ 0.5 & \text{Uncorrelated time series} \\ 0.8 & \text{Long-range dependent same signs} \end{cases}$$

$$\sigma_{MA} = \sqrt{\frac{1}{N_{\max} - n_{\max}} \sum_{n=n_{\max}}^{N_{\max}} (x[n] - x_{MA}[n])} \propto n^H$$



14. Self-similarity, Fractal dimension, Chaos theory



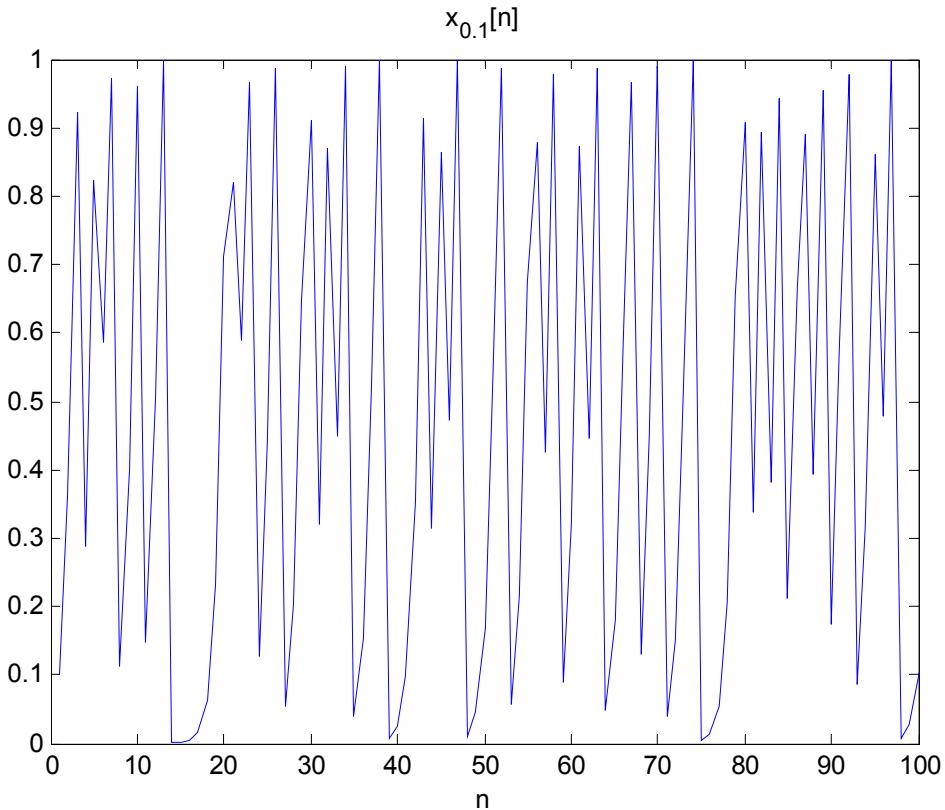
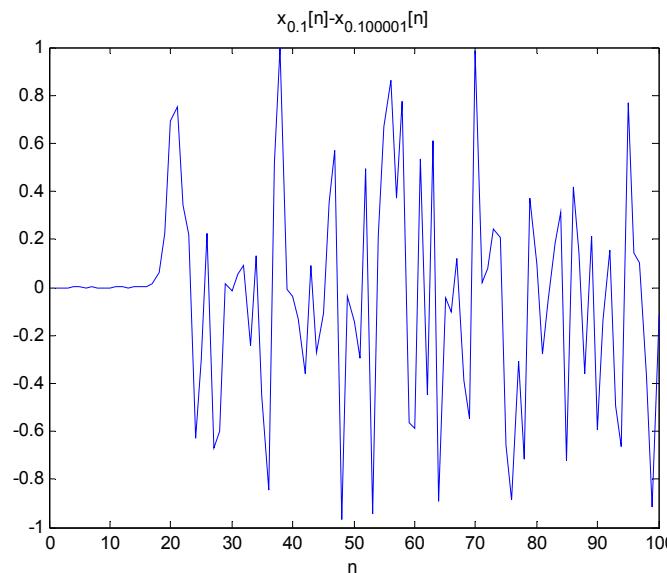
14. Self-similarity, Fractal dimension, Chaos theory

Chaotic systems

$$x[0] = 0.1$$

$$x[n] = 4x[n-1](1 - x[n-1])$$

Logistic equation



14. Self-similarity, Fractal dimension, Chaos theory

Phase space

2-history $\mathbf{x}_2[n] = (x[n], x[n-1]) \in \mathbb{R}^2$

3-history $\mathbf{x}_3[n] = (x[n], x[n-1], x[n-2]) \in \mathbb{R}^3$

h-history $\mathbf{x}_h[n] = (x[n], x[n-1], \dots, x[n-h+1]) \in \mathbb{R}^h$

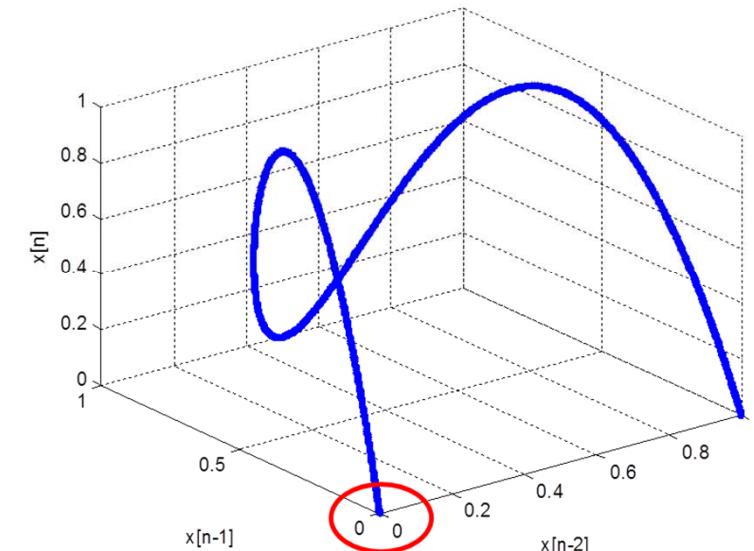
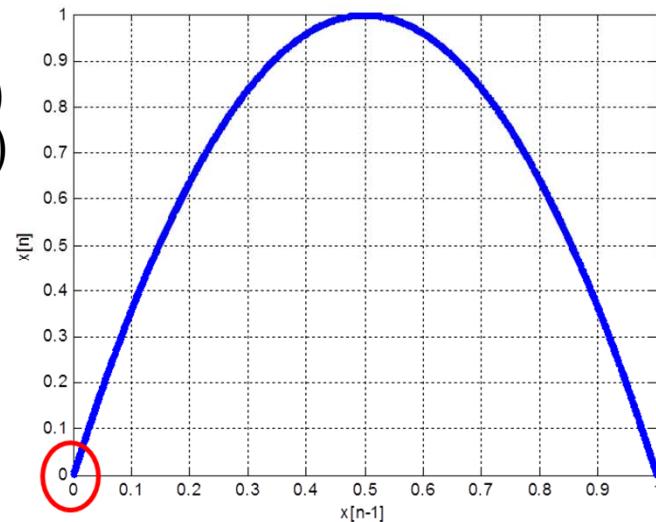
2-Phase space

$$\mathbf{x}_2[1] = (x[1], x[0])$$

$$\mathbf{x}_2[2] = (x[2], x[1])$$

$$\mathbf{x}_2[3] = (x[3], x[2])$$

...

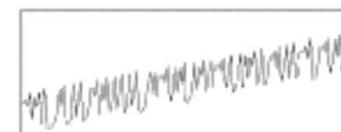
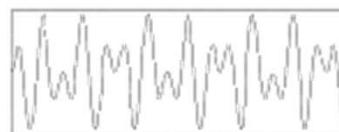
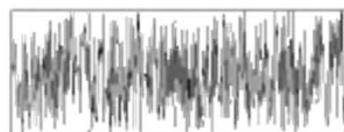


Attractor (Fixed point)

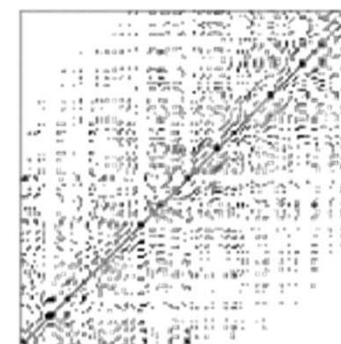
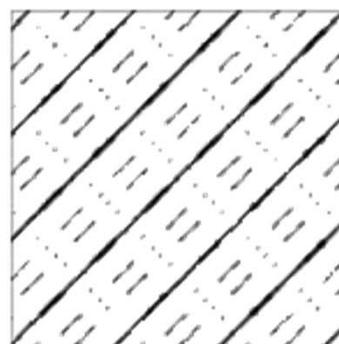
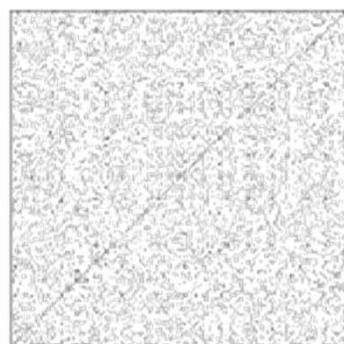
14. Self-similarity, Fractal dimension, Chaos theory

Recurrence plots

$$R_{h,\varepsilon}(n,n') = \begin{cases} 1 & \|\mathbf{x}_h[n] - \mathbf{x}_h[n']\| < \varepsilon \\ 0 & \|\mathbf{x}_h[n] - \mathbf{x}_h[n']\| \geq \varepsilon \end{cases}$$



$$R_{2,\varepsilon}(n,n')$$



White noise

Sinusoidal

Chaotic system
with trend

AR model

14. Self-similarity, Fractal dimension, Chaos theory

Recurrence plots

Observation	Interpretation
Homogeneity	the process is obviously stationary
Fading to the upper left and lower right corners	nonstationarity; the process contains a trend or drift
Disruptions (white bands) occur	nonstationarity; some states are rare or far from the normal; transitions may have occurred
Periodic/ quasi-periodic patterns	cyclicities in the process; the time distance between periodic patterns (e.g. lines) corresponds to the period; long diagonal lines with different distances to each other reveal a quasi-periodic process
Single isolated points	heavy fluctuation in the process; if only single isolated points occur, the process may be an uncorrelated random or even anti-correlated process
Diagonal lines (parallel to the LOI)	the evolution of states is similar at different times; the process could be deterministic; if these diagonal lines occur beside single isolated points, the process could be chaotic (if these diagonal lines are periodic, unstable periodic orbits can be retrieved)
Diagonal lines (orthogonal to the LOI)	the evolution of states is similar at different times but with reverse time; sometimes this is a sign for an insufficient embedding
Vertical and horizontal lines/clusters	some states do not change or change slowly for some time; indication for laminar states
Long bowed line structures	the evolution of states is similar at different epochs but with different velocity; the dynamics of the system could be changing (but note: this is not fully valid for short bowed line structures)



14. Self-similarity, Fractal dimension, Chaos theory

Correlation dimension (Grassberger-Procaccia plots)

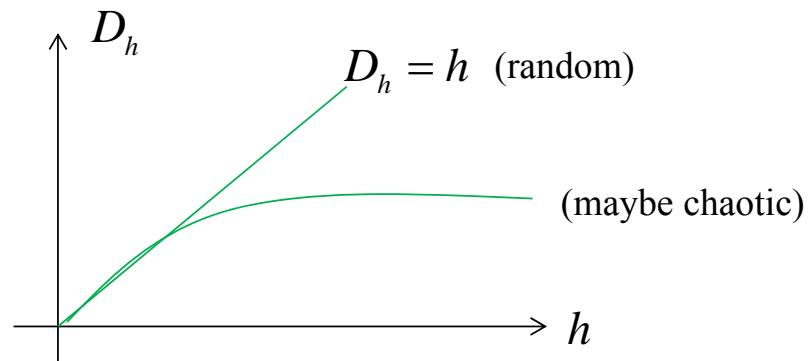
Correlation integral

$$C_h(\varepsilon) = \Pr \left\{ \|\mathbf{x}_h[n] - \mathbf{x}_h[n']\| < \varepsilon \right\} = \lim_{N \rightarrow \infty} \frac{1}{\frac{N(N-1)}{2}} \sum_{\substack{n, n' = 1 \\ n \neq n'}}^N H(\varepsilon - \|\mathbf{x}_h[n] - \mathbf{x}_h[n']\|)$$

↑ Heaviside function $H(x) = \begin{cases} 0 & x \leq 0 \\ 1 & x > 0 \end{cases}$

Correlation dimension

$$D_h = \lim_{\varepsilon \rightarrow 0} \frac{\log(C_h(\varepsilon))}{\log(\varepsilon)}$$



14. Self-similarity, Fractal dimension, Chaos theory

Brock-Dechert-Scheinkman (BDS) Test

H_0 : samples are iid

$$V_{h,\varepsilon} = \frac{C_h(\varepsilon) - (C_1(\varepsilon))^h}{\sqrt{N}} \sim N(0,1)$$

Lyapunov exponent

Consider two time points such that

$$\|\mathbf{x}_h[n] - \mathbf{x}_h[n']\| = \delta_0 \ll 1$$

Consider the distance m samples later

$$\delta_m = \|\mathbf{x}_h[n+m] - \mathbf{x}_h[n'+m]\|$$

The Lyapunov exponent relates these two distances $\delta_m = \delta_0 e^{\lambda m}$

$\lambda > 0 \longrightarrow$ Histories diverge: chaos, cannot be predicted

$\lambda < 0 \longrightarrow$ Histories converge: can be predicted

Maximal Lyapunov exponent

$$\lambda = \lim_{m, \delta_0 \rightarrow \infty} \frac{1}{m} \log \frac{\delta_m}{\delta_0}$$

Bibliography

- C. Chatfield. *The analysis of time series: an introduction*. Chapman & Hall, CRC, 1996.
- D.S.G. Pollock. *A handbook of time-series analysis, signal processing and dynamics*. Academics Press, 1999.
- J. D. Hamilton. *Time series analysis*. Princeton Univ. Press, 1994.



CEU
*Universidad
San Pablo*

Time Series Analysis

Session IV: Forecasting and Data Mining

Carlos Óscar Sánchez Sorzano, Ph.D.
Madrid

Session outline

1. Forecasting
2. Univariate forecasting
3. Intervention modelling
4. State-space modelling
5. Time series data mining
 1. Time series representation
 2. Distance measure
 3. Anomaly/Novelty detection
 4. Classification/Clustering
 5. Indexing
 6. Motif discovery
 7. Rule extraction
 8. Segmentation
 9. Summarization

1. Forecasting

$$x[n+h] ? \longrightarrow \hat{x}[n+h] = g_n[h] = \arg \min_g E\{L(x[n+h], g_n[h])\}$$

Goals:

- Symmetric loss functions:
 - Quadratic loss function: $L(x[n+h], \hat{x}[n+h]) = (x[n+h] - \hat{x}[n+h])^2$
 - Other loss functions: $L(x[n+h], \hat{x}[n+h]) = |x[n+h] - \hat{x}[n+h]|$
- Assymmetric loss functions:

$$L(x[n+h], \hat{x}[n+h]) = \begin{cases} a(x[n+h] - \hat{x}[n+h])^2 & x[n+h] \leq \hat{x}[n+h] \\ b(x[n+h] - \hat{x}[n+h])^2 & x[n+h] > \hat{x}[n+h] \end{cases}$$

Univariate Forecasting: use only samples of the time series to be predicted

→ Solution: $\hat{x}[n+h] = g_n[h] = E\{x[n+h] | x[1], x[2], \dots, x[n]\}$

Multivariate Forecasting: use samples of the time series to be predicted and other “companion” time series.

2. Univariate Forecasting

Trend and seasonal component extrapolation

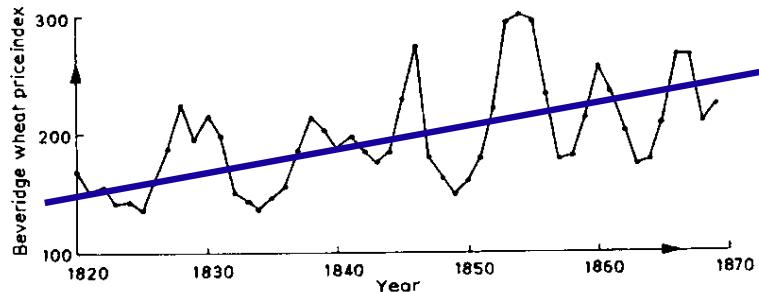
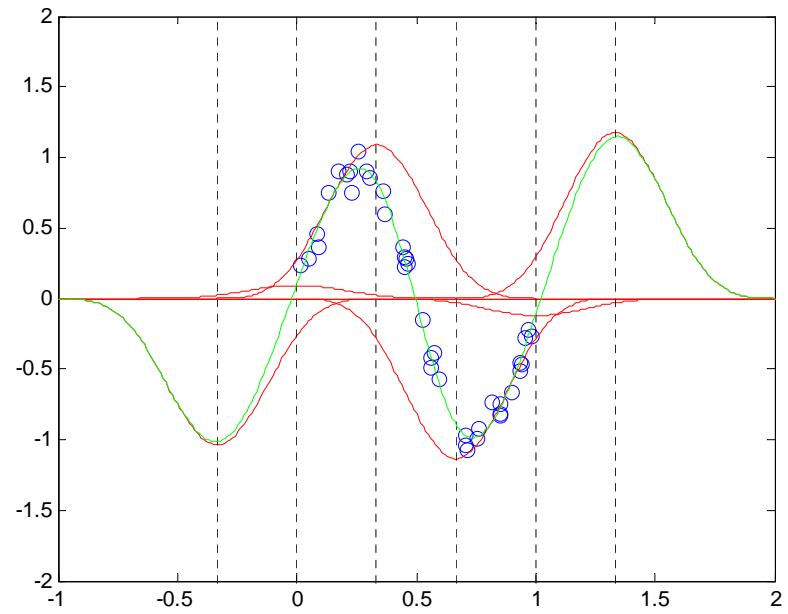
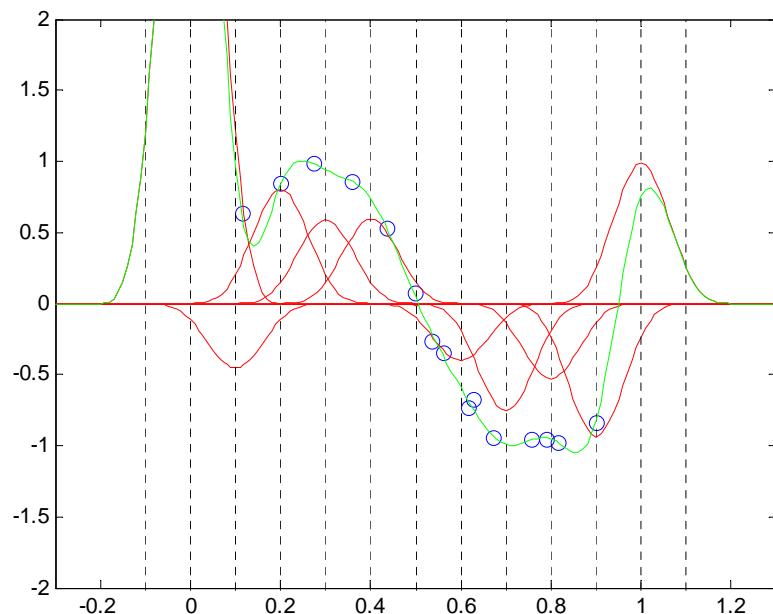


Figure 1.1 Part of the Beveridge wheat price index series.

$$x[n] = \text{trend}[n] + \text{seasonal}[n] + \text{random}[n]$$

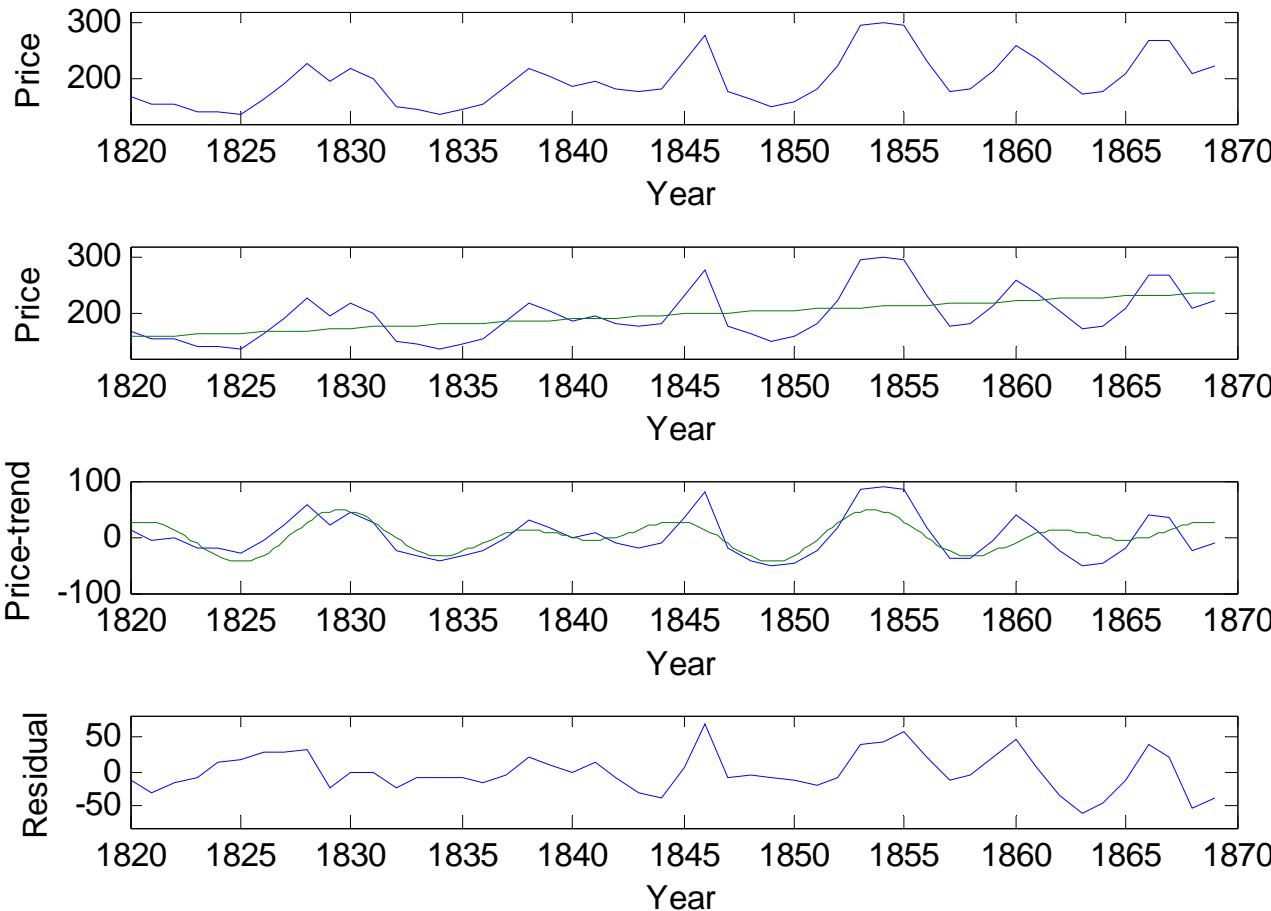
$$\text{trend}[n] = -2777 + 1.6n$$

$$\text{seasonal}[n] = 26.5 \cos\left(\frac{2\pi}{8}n + \frac{2\pi}{3}\right) + 22.5 \cos\left(\frac{2\pi}{12}n + \pi\right)$$



2. Univariate Forecasting

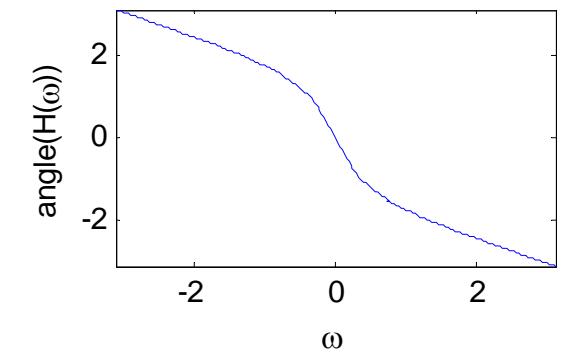
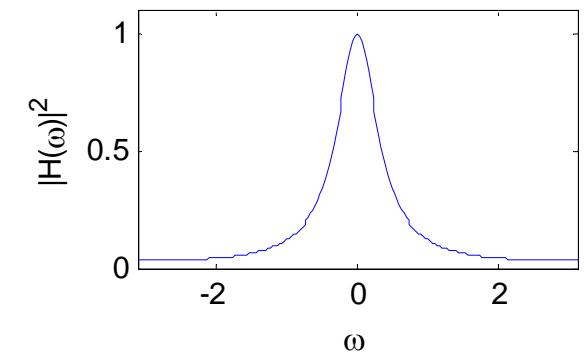
Exponential smoothing



$$\begin{aligned}\hat{x}[n+1] &= \alpha x[n] + (1 - \alpha)\hat{x}[n] \\ &= \alpha(x[n] - \hat{x}[n]) + \hat{x}[n] \\ &= \alpha e[n] + \hat{x}[n]\end{aligned}$$

$$\hat{x}[1] = x[1]$$

$$H(z) = \frac{\hat{X}(z)}{X(z)} = \frac{\alpha}{z + \alpha - 1}$$



2. Univariate forecasting

Model based forecasting (Box-Jenkins procedure)

1. Model identification: examine the data to select an appropriate model structure (AR, ARMA, ARIMA, SARIMA, etc.)
2. Model estimation: estimate the model parameters
3. Diagnostic checking: examine the residuals of the model to check if it is valid
4. Consider alternative models if necessary: if the residual analysis reveals that the selected model is not appropriate
5. Use the model difference equation to predict: as shown in the next two slides.

2. Univariate forecasting

Model based forecasting

Example: ARMA(1,1)

$$\begin{aligned}
 x[n] &= ax[n-1] + w[n] + bw[n-1] \longrightarrow H_W(z) = \frac{X(z)}{W(z)} = \frac{1+bz^{-1}}{1-az^{-1}} \\
 \boxed{\hat{x}[n+1] = ax[n] + w[n+1] + bw[n]} \quad &\longrightarrow z\hat{X}(z) = aX(z) + bW(z) = aX(z) + b\frac{X(z)}{H_W(z)} \\
 \hat{H}_X(z) &= \frac{\hat{X}(z)}{X(z)} = \frac{1}{z} \left(a + \frac{b}{1+bz^{-1}} \right) = \frac{1}{z} \frac{a+b}{1+bz^{-1}} = \frac{a+b}{z+b} \\
 h > 1 \quad &\downarrow \quad \hat{x}[n+1] + b\hat{x}[n] = (a+b)x[n] \quad \boxed{\hat{x}[n+1] = (a+b)x[n] - b\hat{x}[n]} \\
 \boxed{\hat{x}[n+h] = a\hat{x}[n+h-1] + w[n+h] + bw[n+h-1]} &= a^2\hat{x}[n+h-2] = a^3\hat{x}[n+h-3] = \dots = \\
 &= a^{h-1}\hat{x}[n+1] = \boxed{a^{h-1}((a+b)x[n] - b\hat{x}[n])}
 \end{aligned}$$

2. Univariate forecasting

Model based forecasting

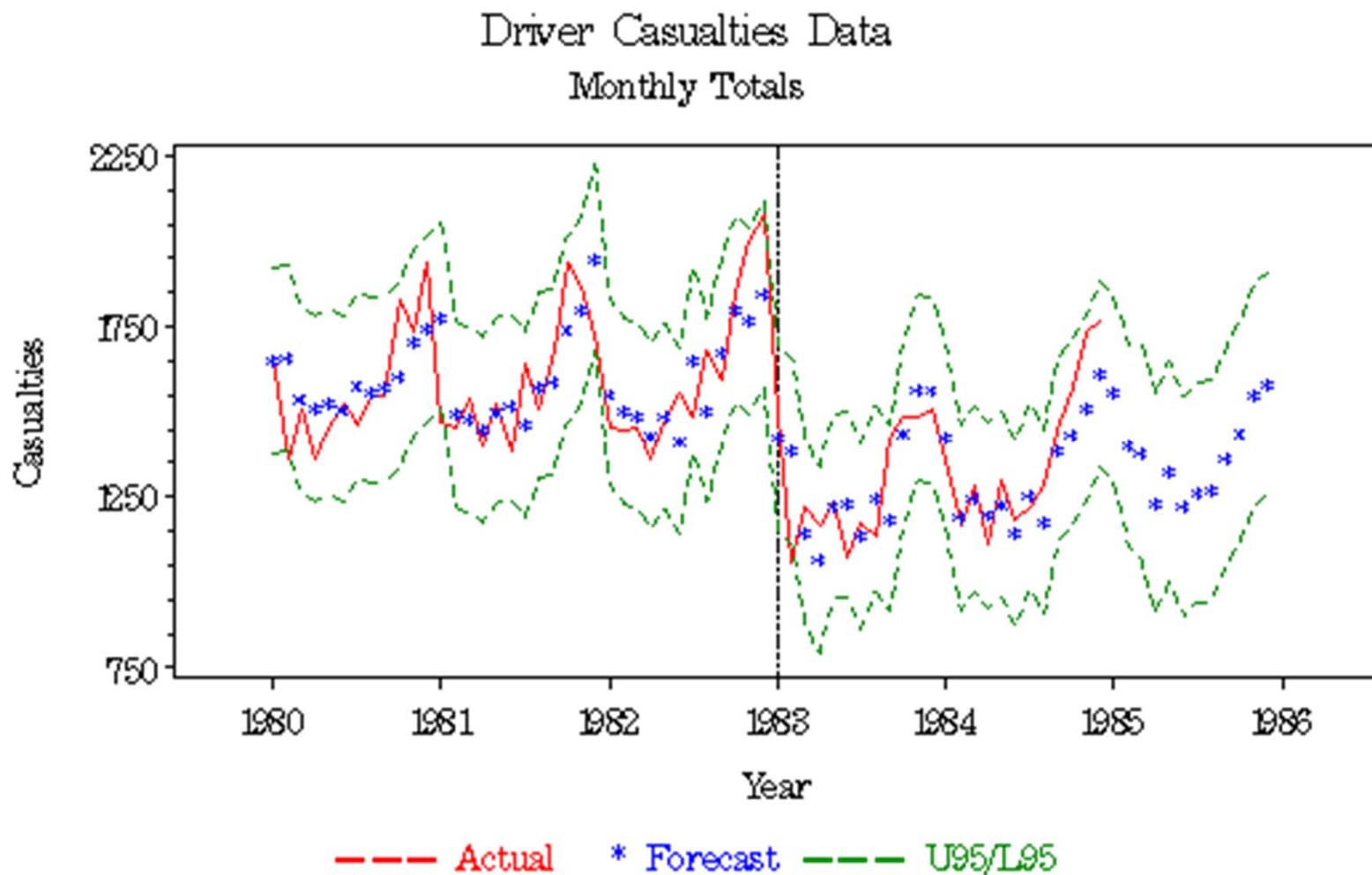
Example: SARIMA(1,0,0)x(0,1,1)₁₂

$$x[n] = x[n-12] + \alpha(x[n-1] - x[n-13]) + w[n] + \beta w[n-12] \longrightarrow H_w(z) = \frac{X(z)}{W(z)} = \frac{1 + \beta z^{-12}}{1 - \alpha z^{-1} - z^{-12} + \alpha z^{-13}}$$
$$\downarrow$$
$$\hat{x}[n+1] = x[n-11] + \alpha(x[n] - x[n-12]) + \cancel{w[n+1]} + \beta w[n-11] \quad \downarrow$$
$$\downarrow$$
$$w[n] = x[n] - \alpha x[n-1] - x[n-12] + \alpha x[n-13] - \beta w[n-12]$$

Eliminating $w[n]$

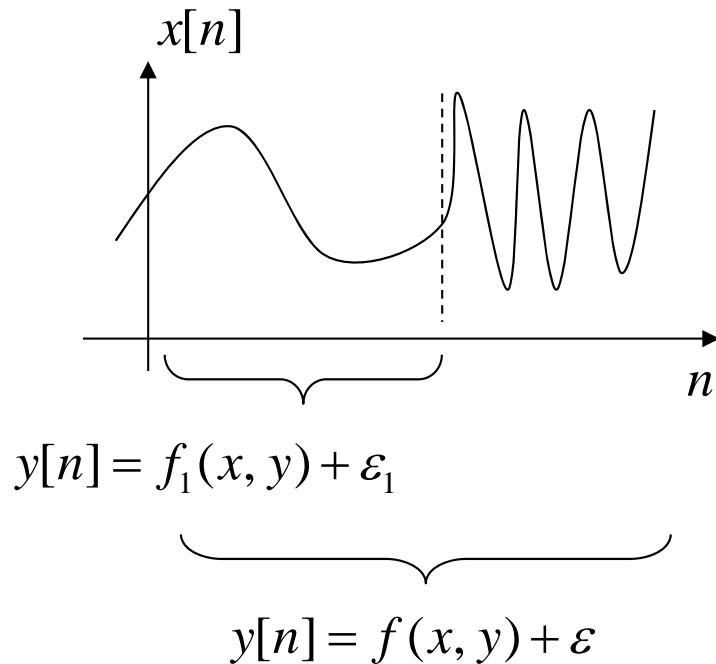
$$\begin{aligned}\hat{x}[n+1] &= (\alpha + \beta)x[n] - \alpha\beta x[n-1] + x[n-11] + (\alpha\beta - \alpha - \beta)x[n-12] + \alpha\beta x[n-13] \\ &\quad + \beta x[n-23] - \alpha\beta x[n-24] - \beta \hat{x}[n-12]\end{aligned}$$

2. Univariate Forecasting



2. Univariate forecasting

Predictive power test



Chow test:

$$H_0 : f_1 = f$$

$$H_1 : f_1 \neq f$$

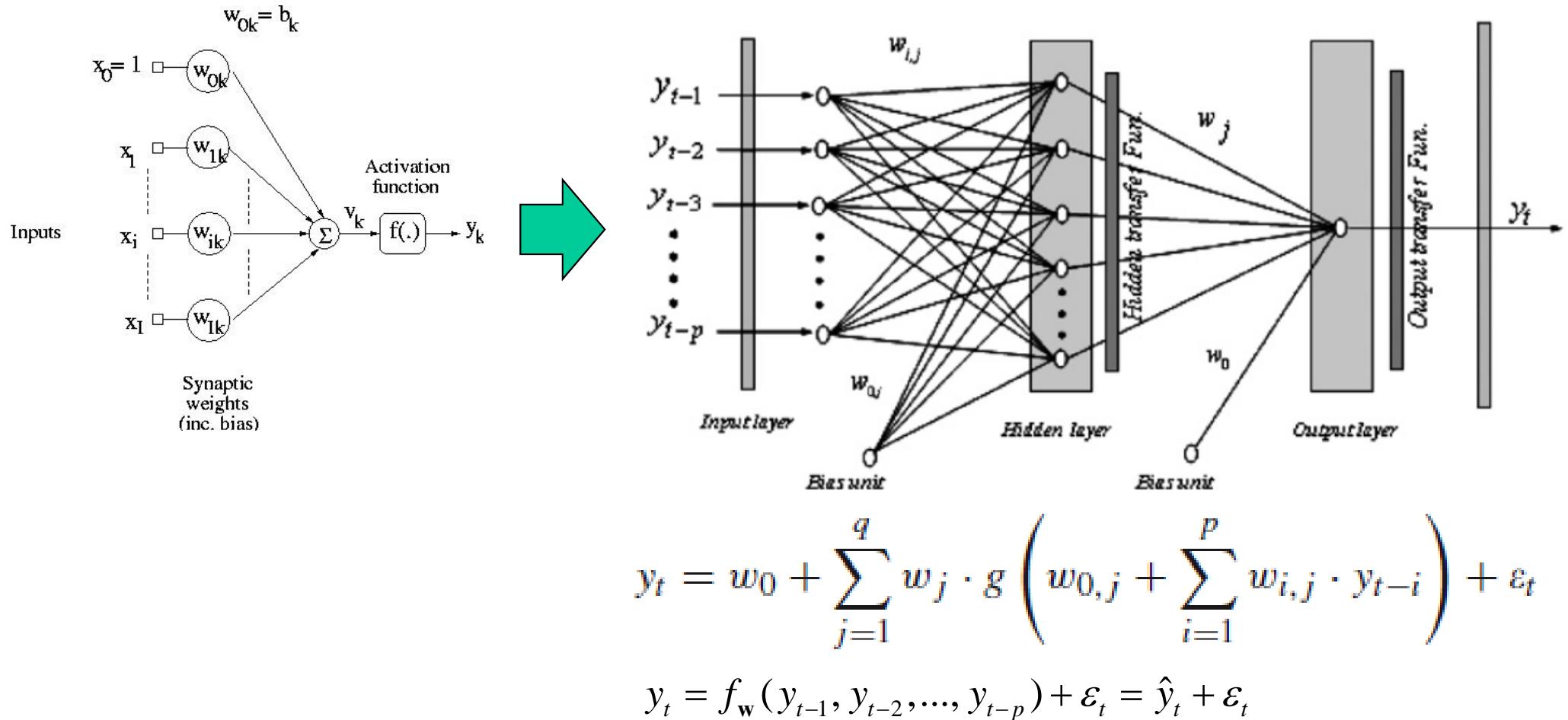
$$F = \frac{\frac{1}{N_2}(S - S_1)}{\frac{1}{N_1 - k} S_1} \sim F(N_2, N_1 - k)$$

Assumption: variance is the same in both regions

Solution: Robust standard errors

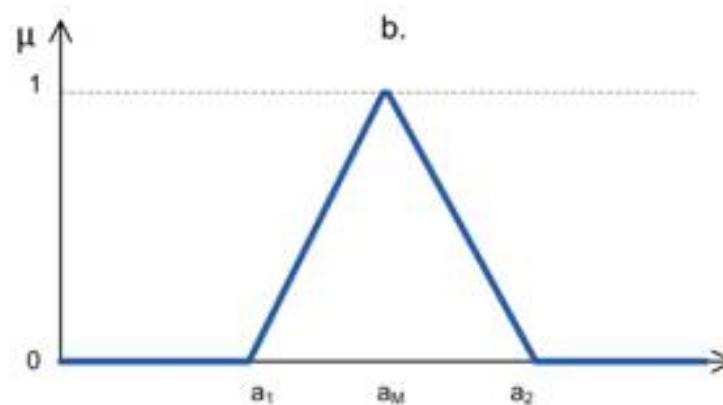
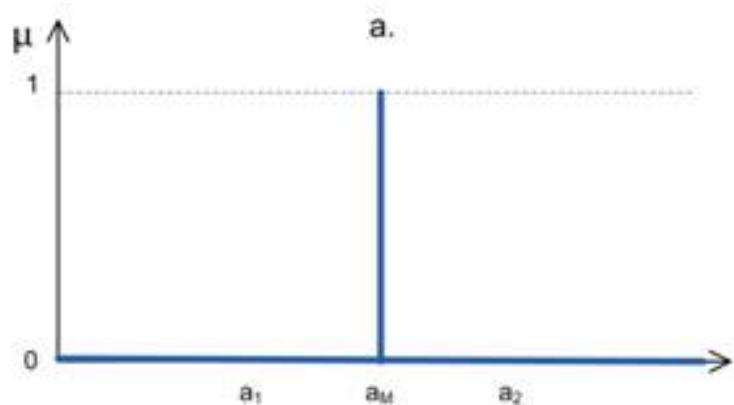
2. Univariate forecasting

Artificial Neural Network (ANN)

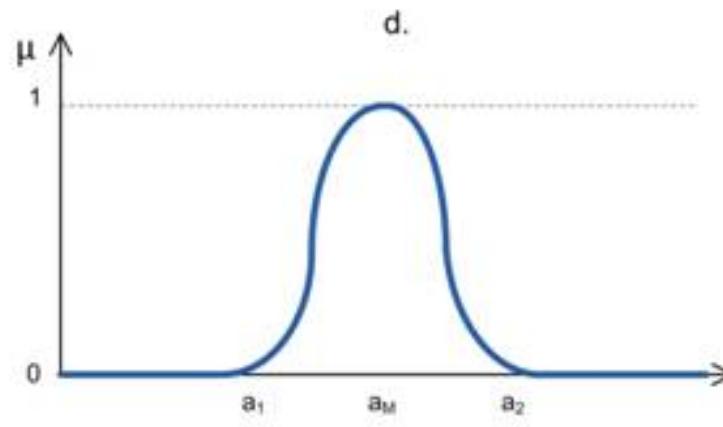
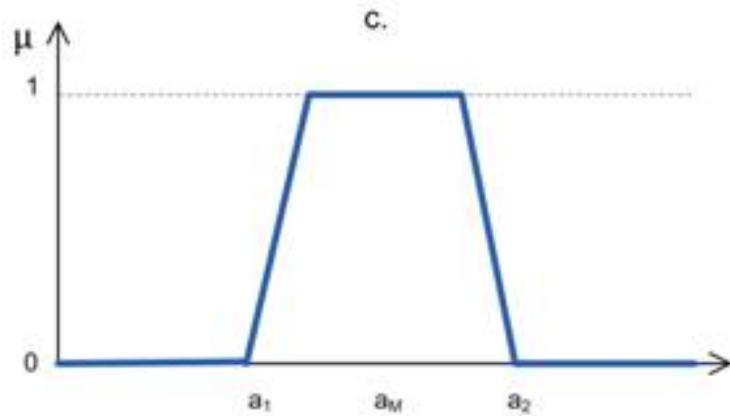


2. Univariate forecasting

Fuzzy Artificial Neural Network (ANN)



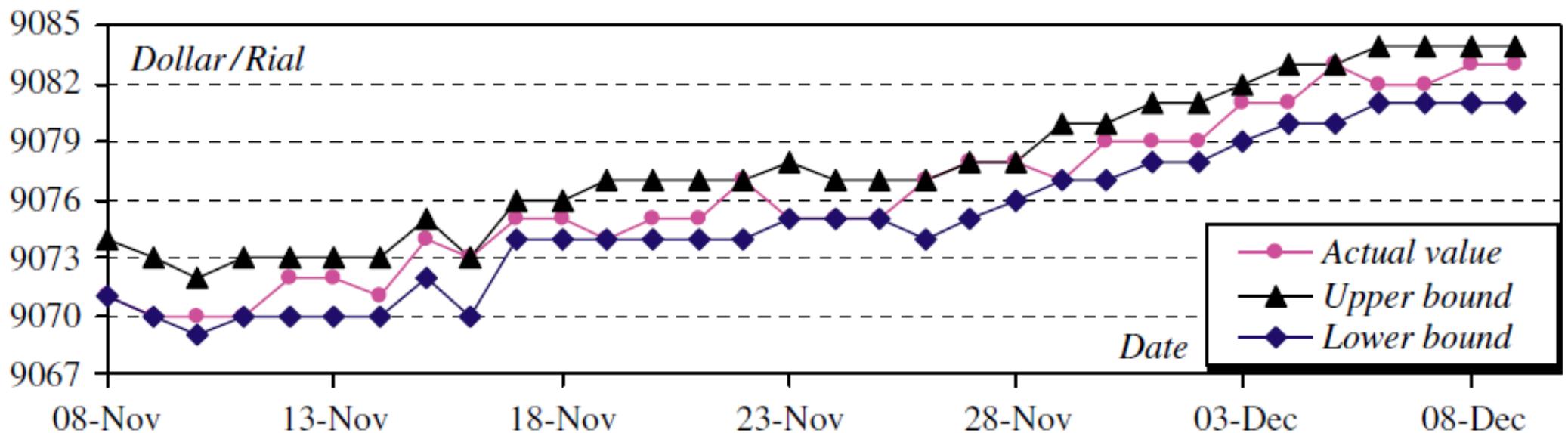
$$\mu_{a_M}(x) : \mathbb{R} \rightarrow [0,1]$$



2. Univariate forecasting

Fuzzy Artificial Neural Network (ANN)

$$y_t = w_0 + \sum_{j=1}^q w_j \cdot g \left(w_{0,j} + \sum_{i=1}^p w_{i,j} \cdot y_{t-i} \right) + \varepsilon_t$$



2. Univariate forecasting

Fuzzy Artificial Neural Network (ANN)

Begin

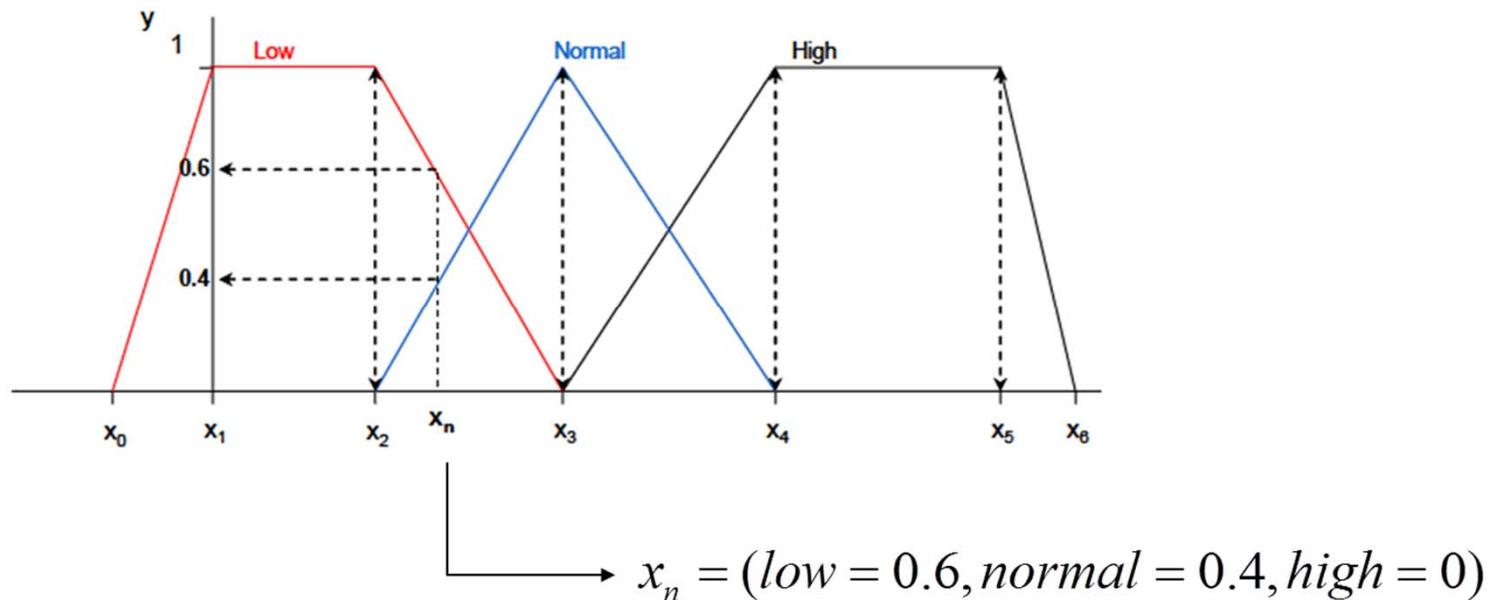
Step 1: Perform interval partitioning for each factor and fuzzify each time series using FCM (Section 3.1)

Step 2: Construct certain transition rules from fuzzy time series using Algorithm CTR (Section 3.2)

Step 3: Forecast in accordance with certain transition rules using Algorithm Forecast (Section 3.3)

Step 4: Defuzzify the forecasting outputs using Eq. (11) (Section 3.3)

End



2. Univariate forecasting

Fuzzy Artificial Neural Network (ANN)

$(A_1, B_2) \rightarrow (A_3, B_3)$
 $(A_4, B_4) (A_1, B_3) \rightarrow (A_7, B_4)$
 $(A_2, B_7) (A_1, B_3) \rightarrow (A_2, B_1)$
 $(A_1, B_4) \rightarrow (A_2, B_2)$
 $(A_2, B_1) \rightarrow (A_1, B_2)$
 $(A_2, B_2) \rightarrow (A_5, B_2)$
 $(A_2, B_6) \rightarrow (A_5, B_3)$
 $(A_4, B_8) \rightarrow (A_6, B_8)$

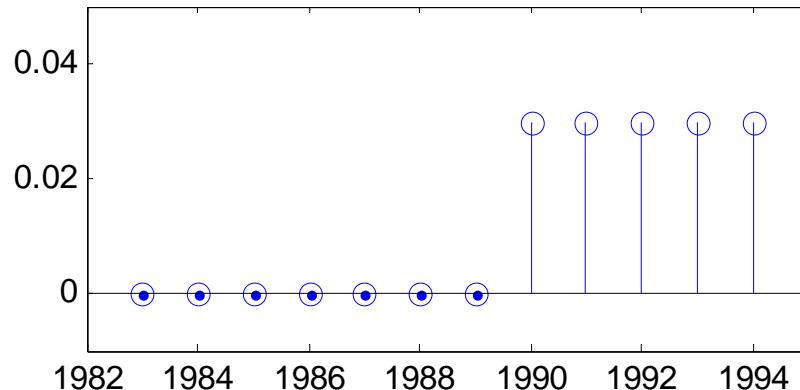
$$F(t - 1) \rightarrow F(t).$$

Day	Temperature	Fuzzified temperature	Cloud density	Fuzzified cloud density	Forecasting state	Defuzzified result	Forecasting error
1	26.1	A_1	36	B_4			
2	27.6	A_2	23	B_2	A_2	27.8	-0.2
3	29.0	A_5	23	B_2	A_5	29	0
4	30.5	A_8	10	B_1	A_8	30.2	0.3
5	30.0	A_8	13	B_1	A_8	30.2	-0.2
6	29.5	A_7	30	B_4	A_7	29.7	-0.2

3. Intervention modelling

What happens in the time series of a price if there is tax raise of 3% in 1990?

$$price[n] = \alpha price[n-1] + w[n] + 0.03u[n-1990]$$

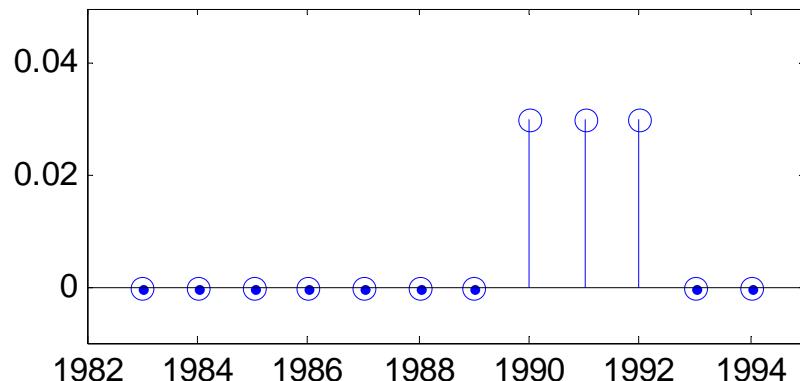


$$u[n] = \begin{cases} 1 & n \geq 0 \\ 0 & n < 0 \end{cases}$$

$$price(z) = \alpha price(z)z^{-1} + w(z) + 0.03 \frac{1}{1-z^{-1}} z^{-1990}$$

What happens in the time series of a price if there is tax raise of 3% between 1990 and 1992?

$$price[n] = \alpha price[n-1] + w[n] + 0.03(u[n-1990] - u[n-1993])$$

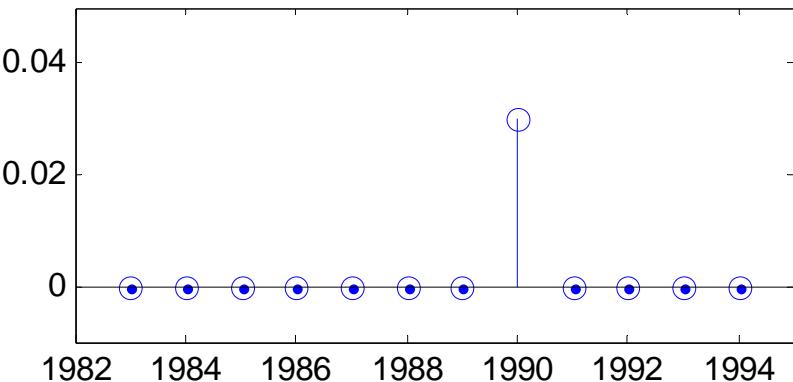


$$price(z) = \alpha price(z)z^{-1} + w(z) + 0.03 \frac{1}{1-z^{-1}} (z^{-1990} - z^{-1993})$$

3. Intervention modelling

What happens in the time series of a price if in 1990 there was an earthquake?

$$price[n] = \alpha price[n-1] + w[n] + \beta \delta[n-1990]$$

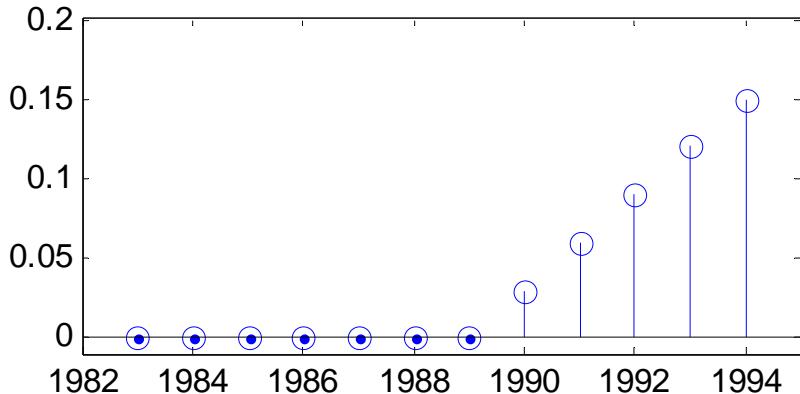


$$\delta[n] = \begin{cases} 1 & n = 0 \\ 0 & \text{otherwise} \end{cases}$$

$$price(z) = \alpha price(z)z^{-1} + w(z) + \beta \cdot 1 \cdot z^{-1990}$$

What happens in the time series of a price if there is a steady tax raise of 3% since 1990?

$$price[n] = \alpha price[n-1] + w[n] + 0.03(n-1989)u[n-1990]$$



$$price(z) = \alpha price(z)z^{-1} + w(z) + 0.03 \frac{1}{1-z^{-1}} \frac{1}{1-z^{-1}} z^{-1990}$$

3. Intervention modelling

In general

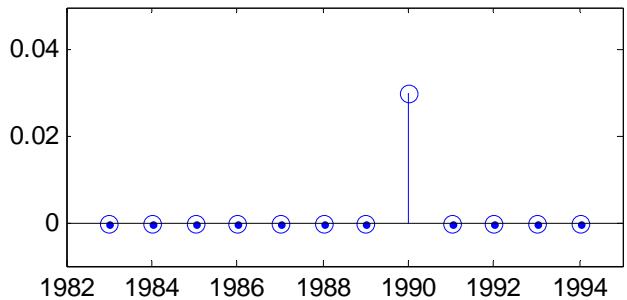
$$x[n] = f(x[n-1], \dots, x[n-q], w[n], w[n-1], \dots, w[n-p]) + f(i[n])$$

$$X(z) = \frac{B(z)}{A(z)}W(z) + \frac{C(z)}{D(z)}I(z)$$



We are back to the system identification problem with external inputs

3. Intervention modelling: outliers revisited



Additive outliers:

$$X(z) = \frac{B(z)}{A(z)} W(z) + I(z)$$

Innovational outliers:

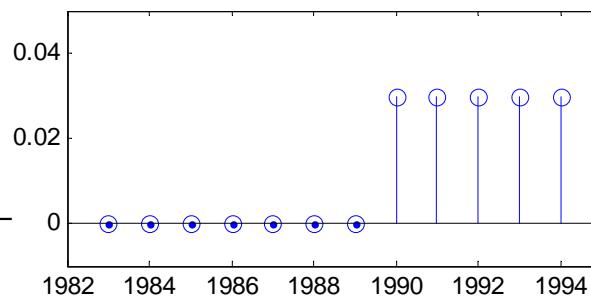
$$X(z) = \frac{B(z)}{A(z)} W(z) + \frac{C(z)}{D(z)} I(z)$$

Time change outliers:

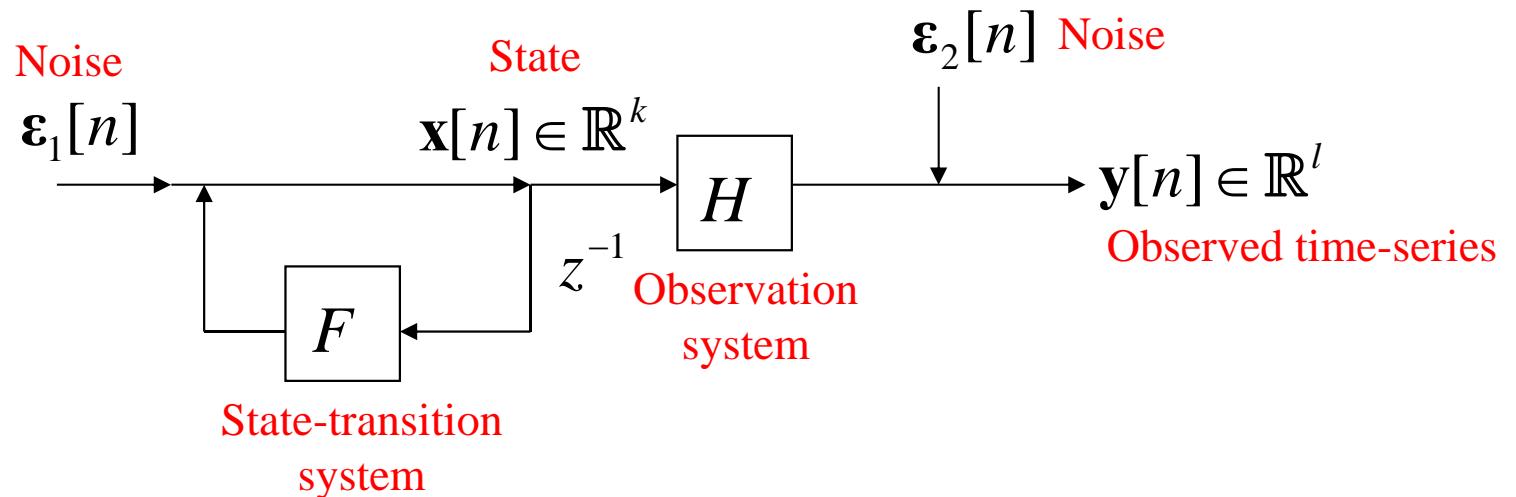
$$X(z) = \frac{B(z)}{A(z)} W(z) + \frac{1}{D(z)} I(z)$$

Level outliers:

$$X(z) = \frac{B(z)}{A(z)} W(z) + I(z)$$



4. State-space modelling



Linear, Gaussian system: $\mathbf{x}[n] = F\mathbf{x}[n-1] + Q\boldsymbol{\varepsilon}_1[n]$ State-transition model

$\mathbf{y}[n] = H\mathbf{x}[n] + \boldsymbol{\varepsilon}_2[n]$ Observation model

$$\mathbf{x}[0] \sim N(\boldsymbol{\mu}_0, \Sigma_0)$$

$$\boldsymbol{\varepsilon}_1[n] \sim N(\mathbf{0}, \Sigma_1)$$

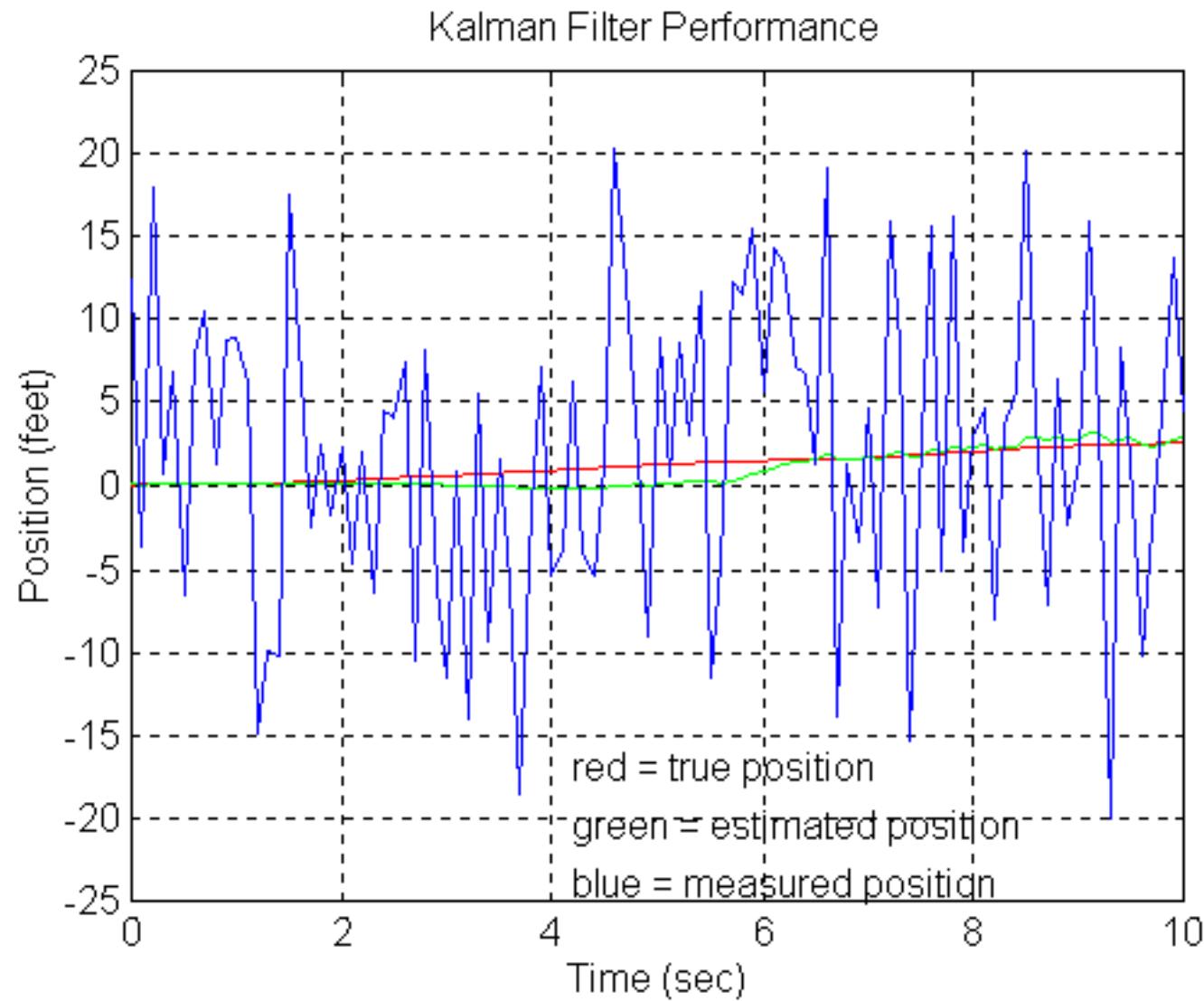
$$\boldsymbol{\varepsilon}_2[n] \sim N(\mathbf{0}, \Sigma_2)$$

Kalman filter: Given $y[n], H, F, \boldsymbol{\mu}_0, \Sigma_0$ estimate $x[n], \Sigma_1, \Sigma_2$

Time series modelling (System Identification):

Given $y[n]$ estimate $x[n], H, F, \boldsymbol{\mu}_0, \Sigma_0, \Sigma_1, \Sigma_2$

4. State-space modelling



4. State-space modelling

Linear, Gaussian system: $\mathbf{x}[n] = F\mathbf{x}[n-1] + Q\boldsymbol{\varepsilon}_1[n]$ State-transition model

$\mathbf{y}[n] = H\mathbf{x}[n] + \boldsymbol{\varepsilon}_2[n]$ Observation model

General system: $\mathbf{x}[n] = f(\mathbf{x}[n-1], \boldsymbol{\varepsilon}_1[n])$ f, h Differentiable

$\mathbf{y}[n] = h(\mathbf{x}[n], \boldsymbol{\varepsilon}_2[n])$

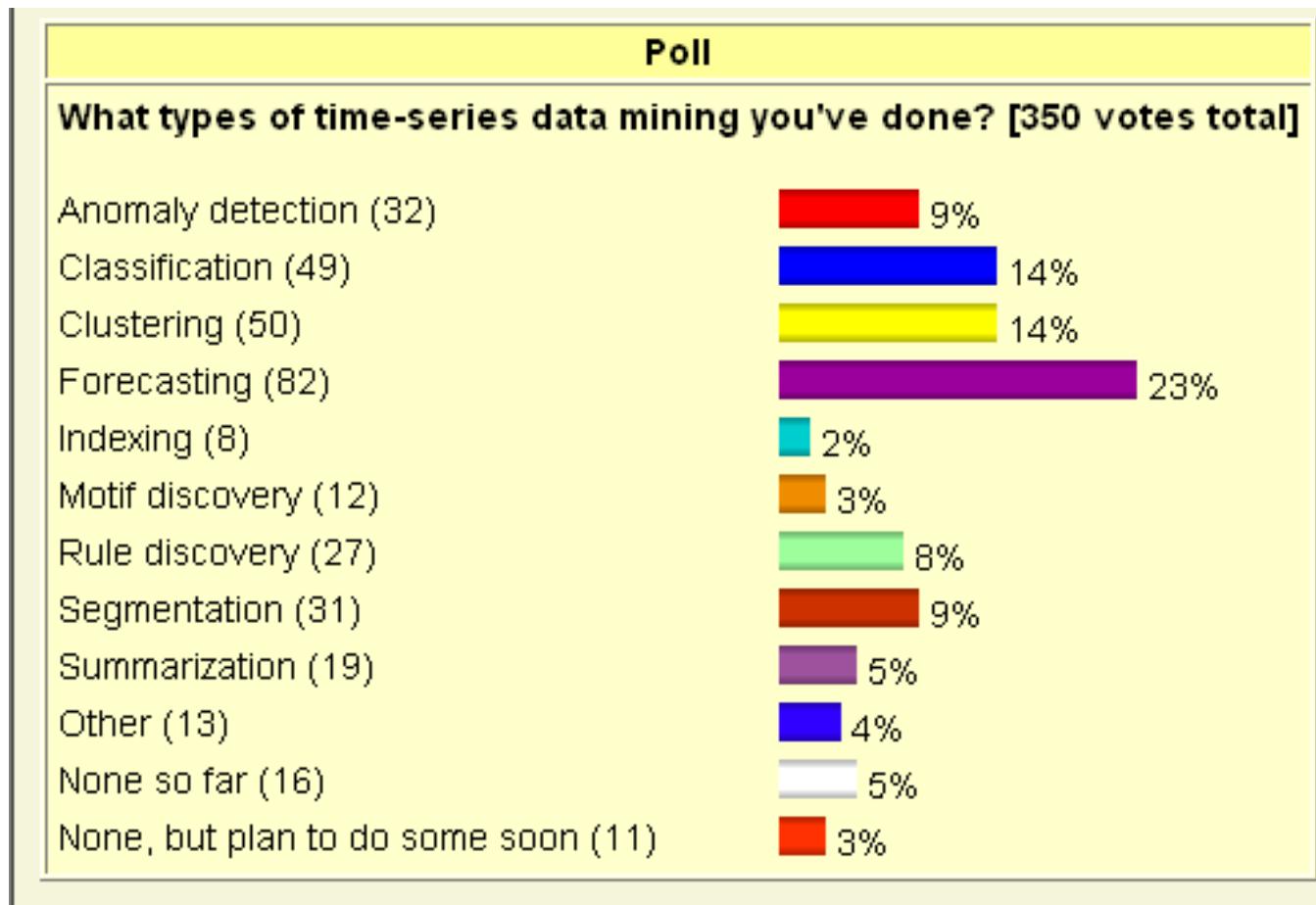
Extended/Unscented Kalman filter: Given $y[n], h, f, \boldsymbol{\mu}_0, \Sigma_0$ estimate $x[n], \Sigma_1, \Sigma_2$

Particle filter:

Given $y[n], h, f, \boldsymbol{\mu}_0, \Sigma_0$ and the probability density functions of $\boldsymbol{\varepsilon}_1, \boldsymbol{\varepsilon}_2$

estimate $\mathbf{x}[n]$ and the free parameters of the noise signals.

5. Time series data mining

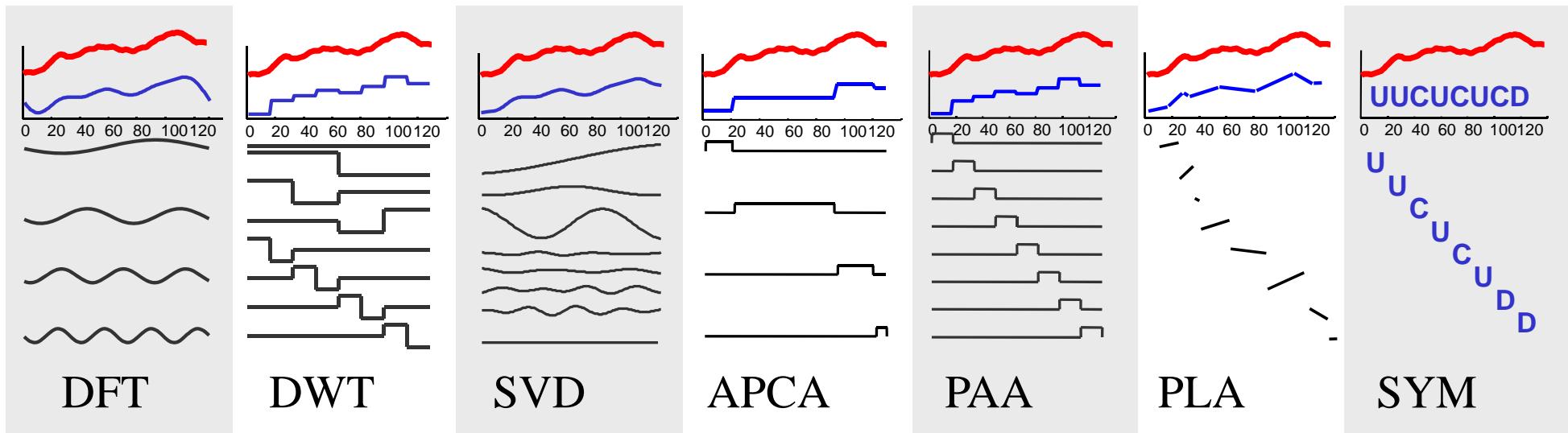
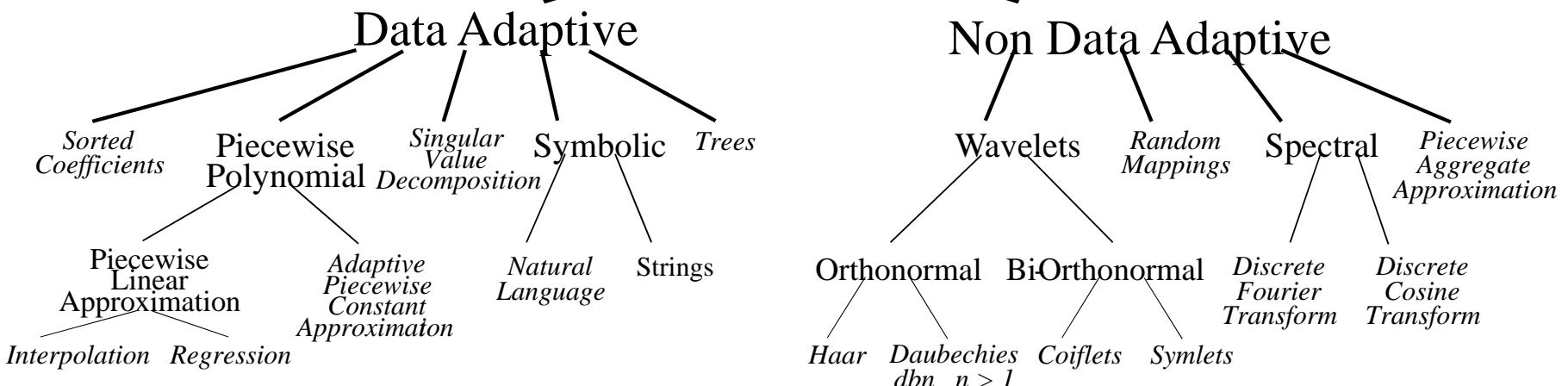


Source: http://www.kdnuggets.com/polls/2004/time_series_data_mining.htm

5. Time series data mining

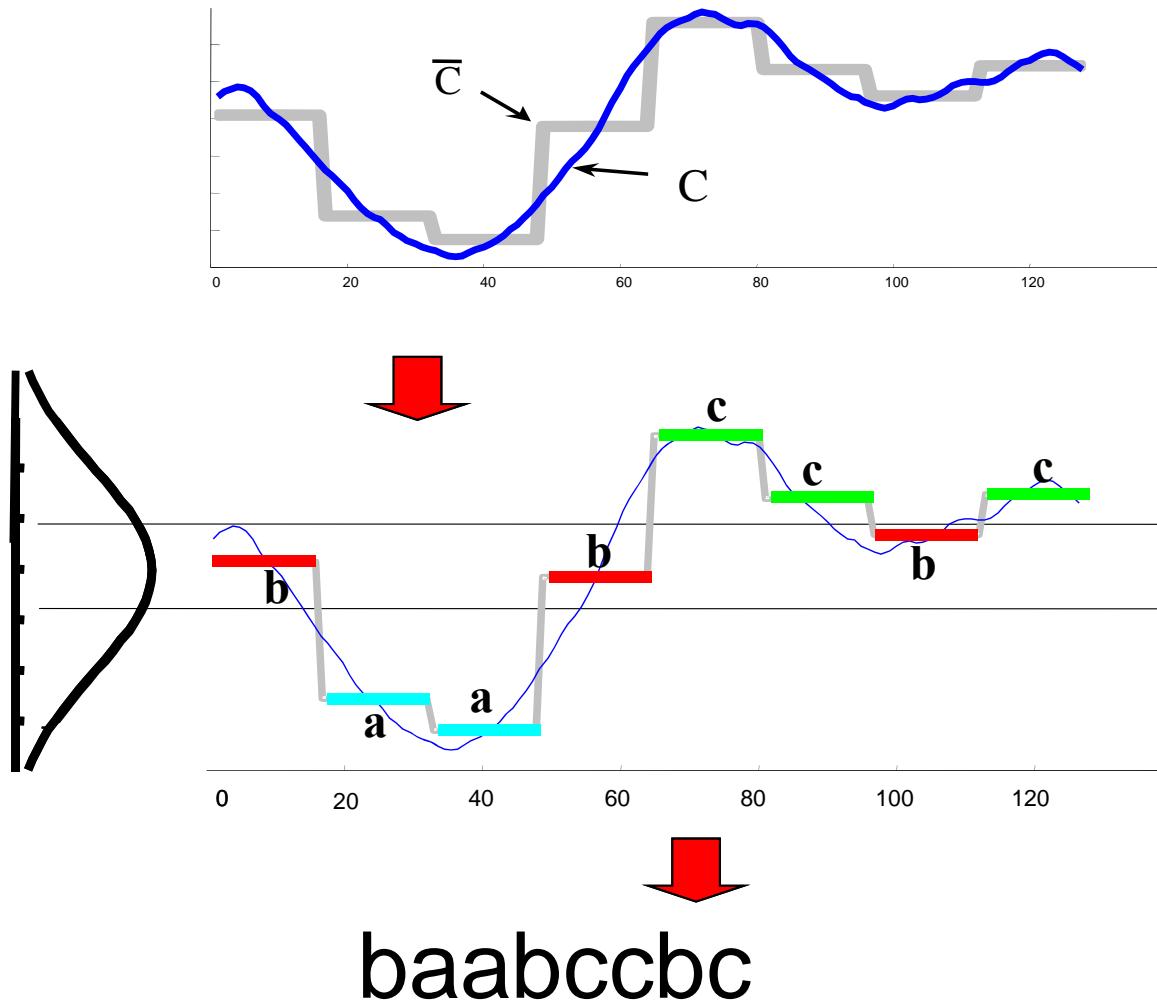
Time series representation

Time Series Representations



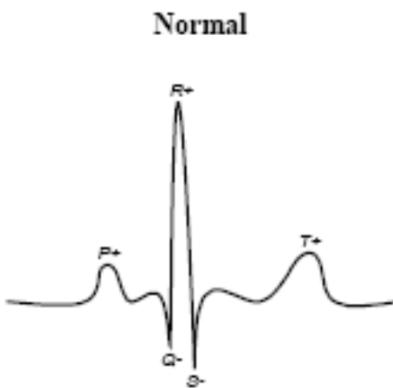
5. Time series data mining

Time series representation



5. Time series data mining

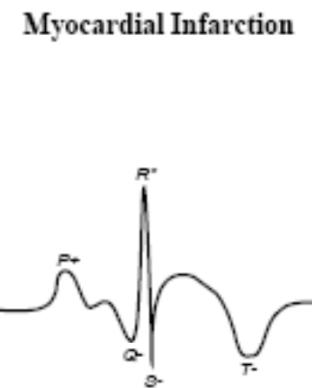
Time series representation



Parse String: P+ Q- R+ S- T+

Grammar:

```
<NORMAL ECG> → <P><QRS><T>
<P> → <P+>
<QRS> → <Q><R><S>
<Q> → <Q->
<R> → <R+>
<S> → <S->
<T> → <T+>
<P+> : positive wave, expected shape
<Q-> : negative wave, expected shape
<R+> : positive wave, expected shape
<R*> : positive wave, invalid amplitude
<S-> : negative wave, expected shape
<T+> : positive wave, expected shape
<T-> : negative wave, expected shape
```



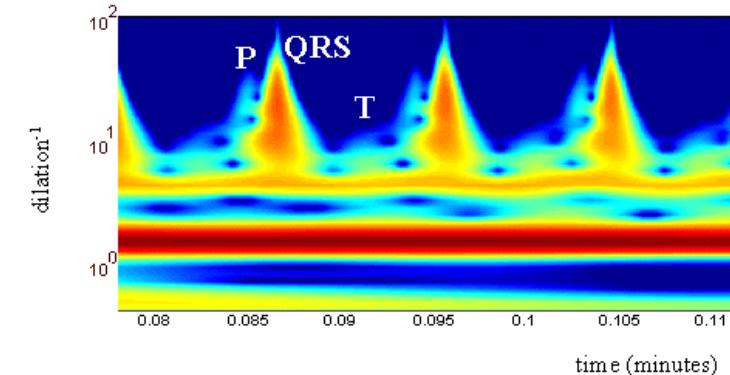
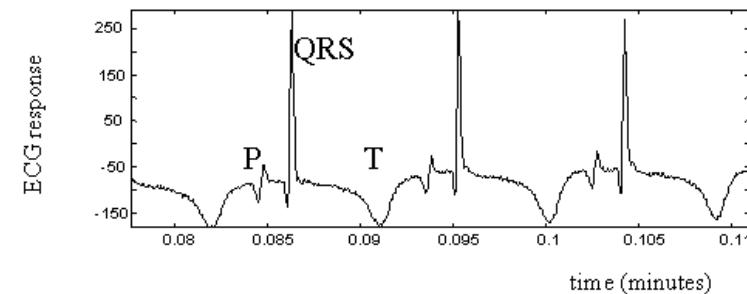
Parse String: P+ Q- R* S- T-

Grammar:

```
<MYO INF ECG> → <P><QRS><T>
<P> → <P+>
<QRS> → <Q><R*><S>
<Q> → <Q->
<R*> → <R*>
<S> → <S->
<T> → <T->
<P+> : positive wave, expected shape
<Q-> : negative wave, expected shape
<R+> : positive wave, expected shape
<R*> : positive wave, invalid amplitude
<S-> : negative wave, expected shape
<T+> : positive wave, expected shape
<T-> : negative wave, expected shape
```

Source: <http://www.cs.cmu.edu/~bobski/pubs/tr01108-onesided.pdf>

Figure 2.5 An example of structural pattern recognition applied to time-series data for electrocardiogram diagnosis. Each waveform traces the electrical activity recorded during one cardiac cycle (i.e., heartbeat) by a single electrode: the leftmost represents data recorded during a normal heartbeat, and the rightmost represents data recorded during a heartbeat exhibiting behavior indicative of a cardiac condition called myocardial infarction. The primitives extracted from each data set are labeled on the waveforms. Concatenating the primitives by time constitutes a parse string, and a context-free grammar can be constructed to parse, and hence classify, each string of primitives.



5. Time series data mining

Time series distance measure

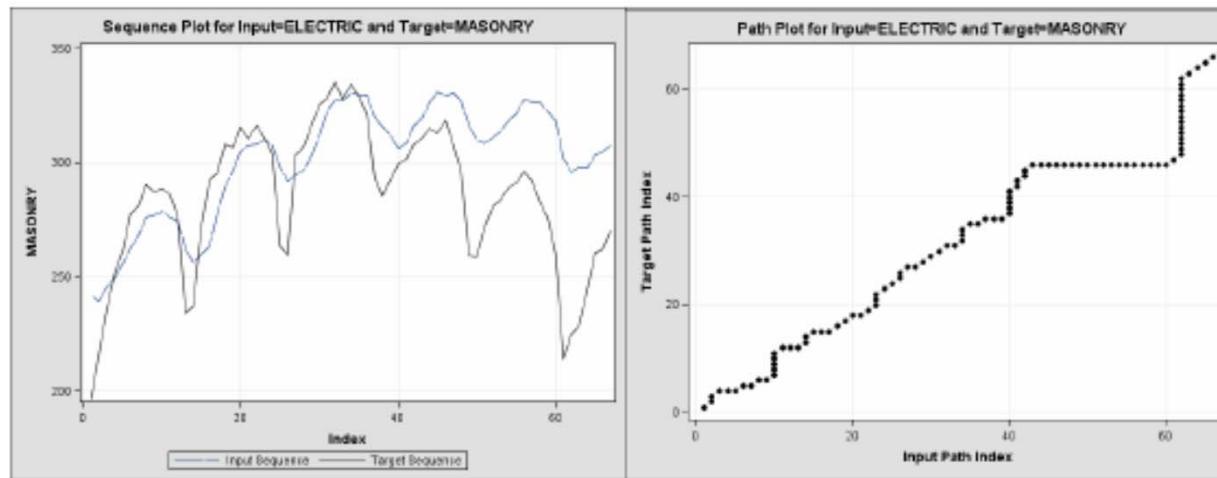


Figure 26: Input and Target Plot

Figure 27: Warping Path

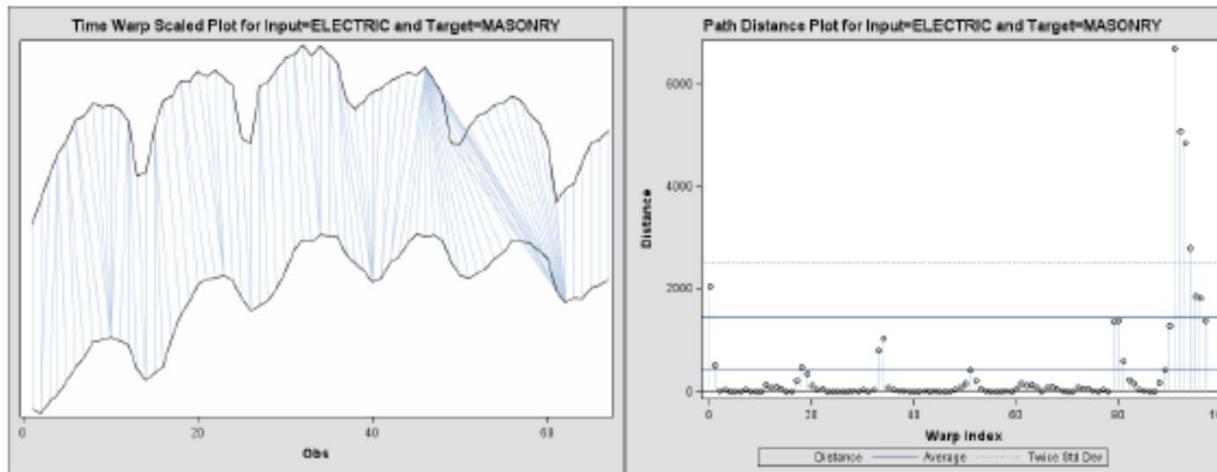
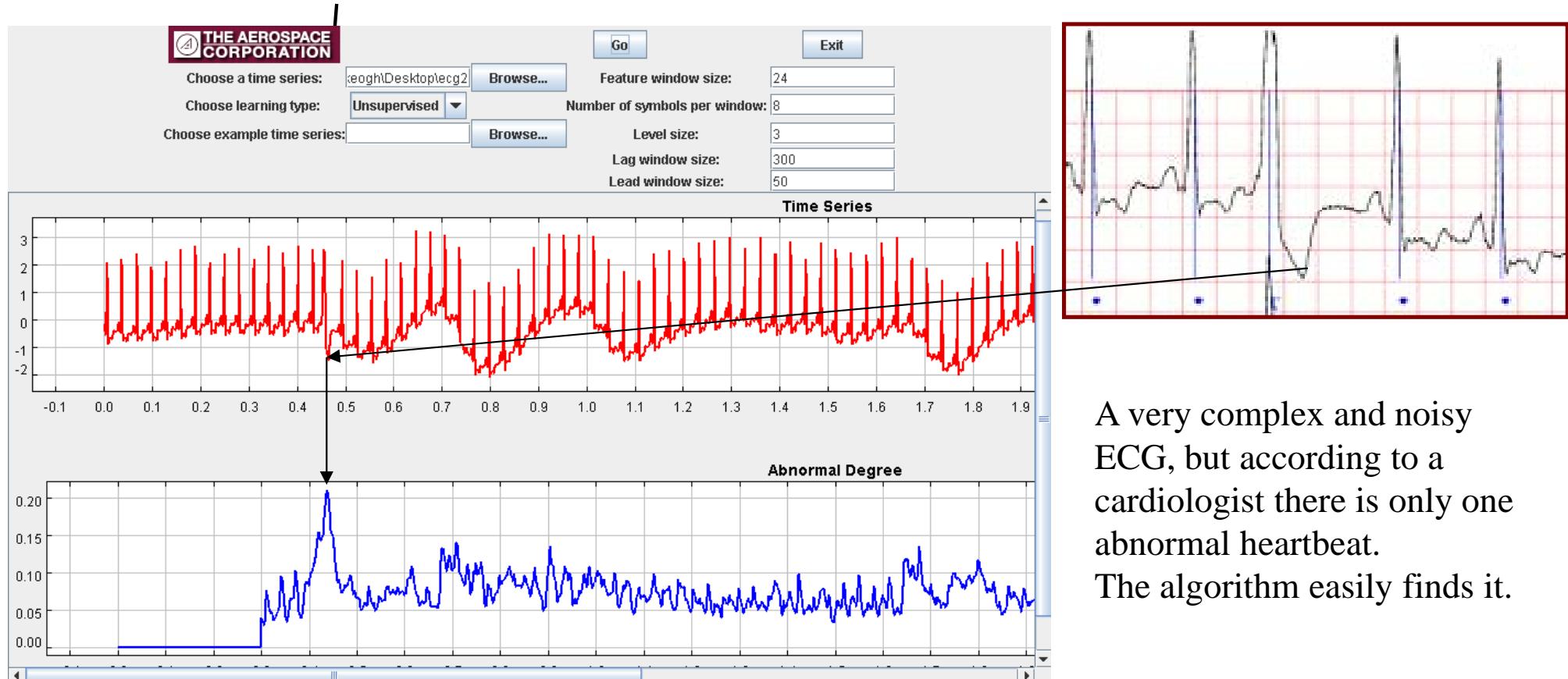


Figure 28: Scaled Warp Plot

Figure 29: Path Distances

5. Time series data mining

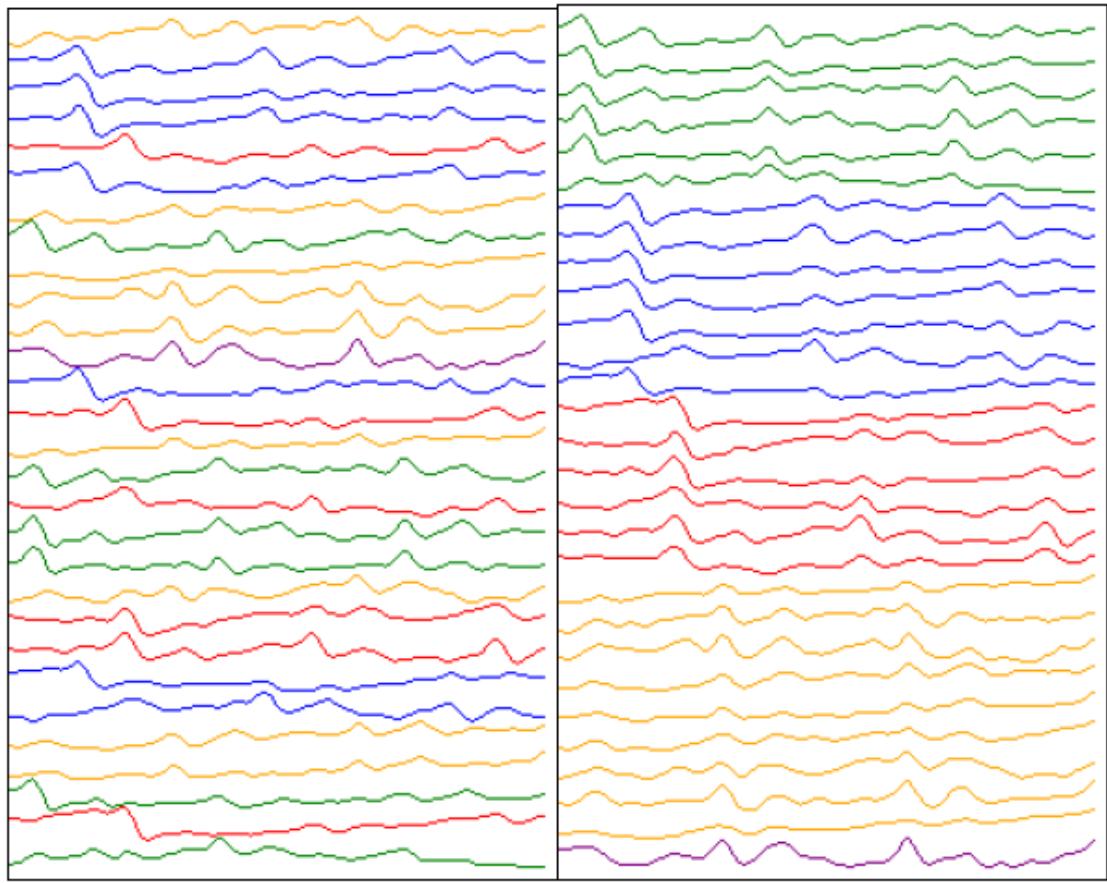
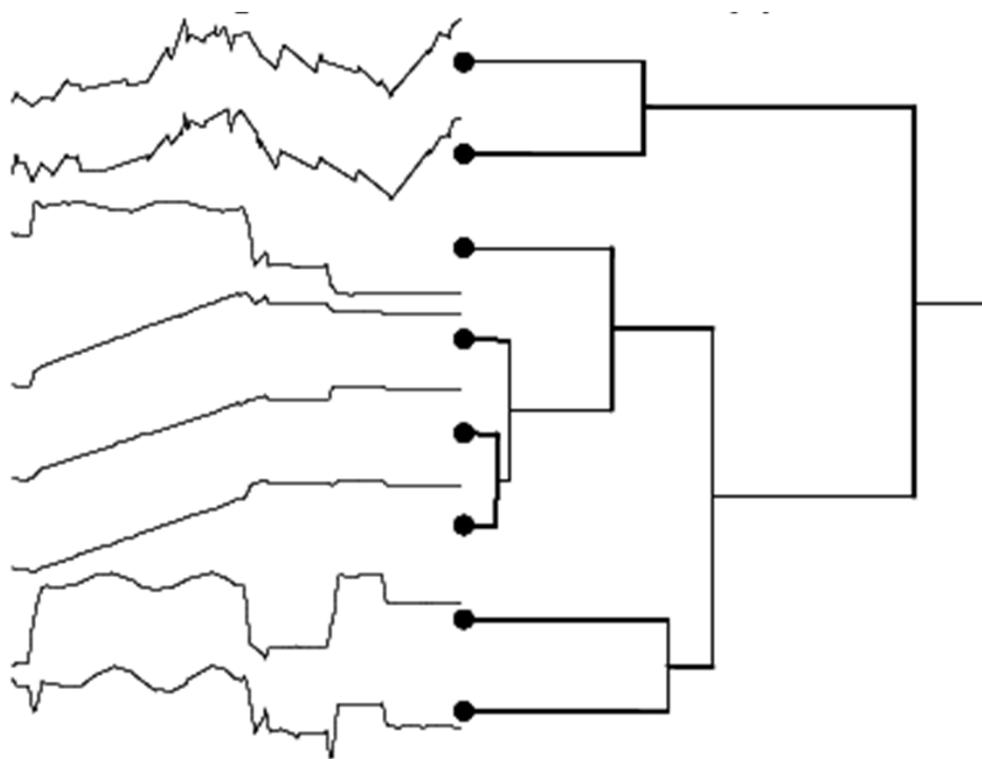
Anomaly/Novelty detection



Source: <http://www.cs.ucr.edu/~wli/SSDBM05/>

5. Time series data mining

Classification/Clustering



5. Time series data mining

Indexing

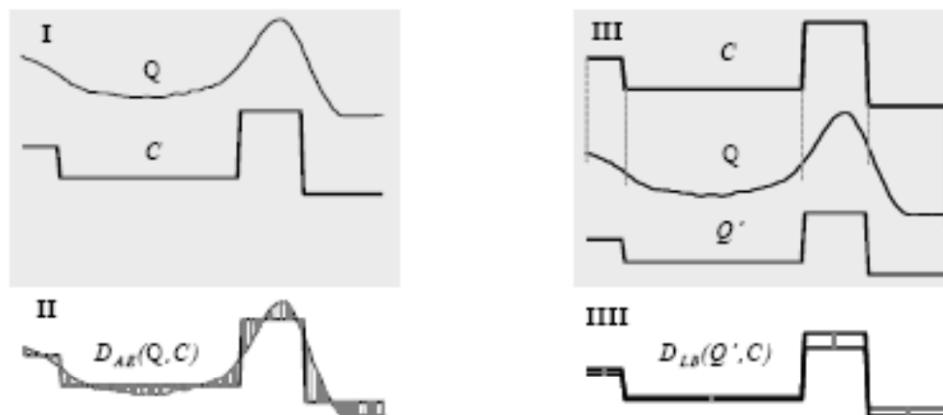


Figure 6: A visualization of the two distance measures define on the APCA representation. I) A query time series Q and a APCA object C . II) The D_{AE} measure can be visualized as the Euclidean distance between Q and the reconstruction of C . III) Q' is obtained by projecting the endpoints of C onto Q and calculating the mean values of the sections falling within the projected lines. IIII) The D_{LB} measure can be visualized as the square root of the sum of the product of squared length of the gray lines with the length of the segments they join.

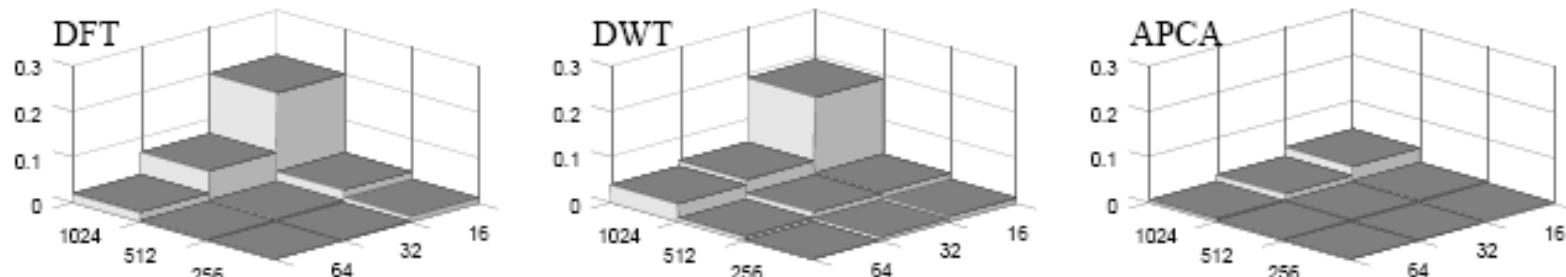
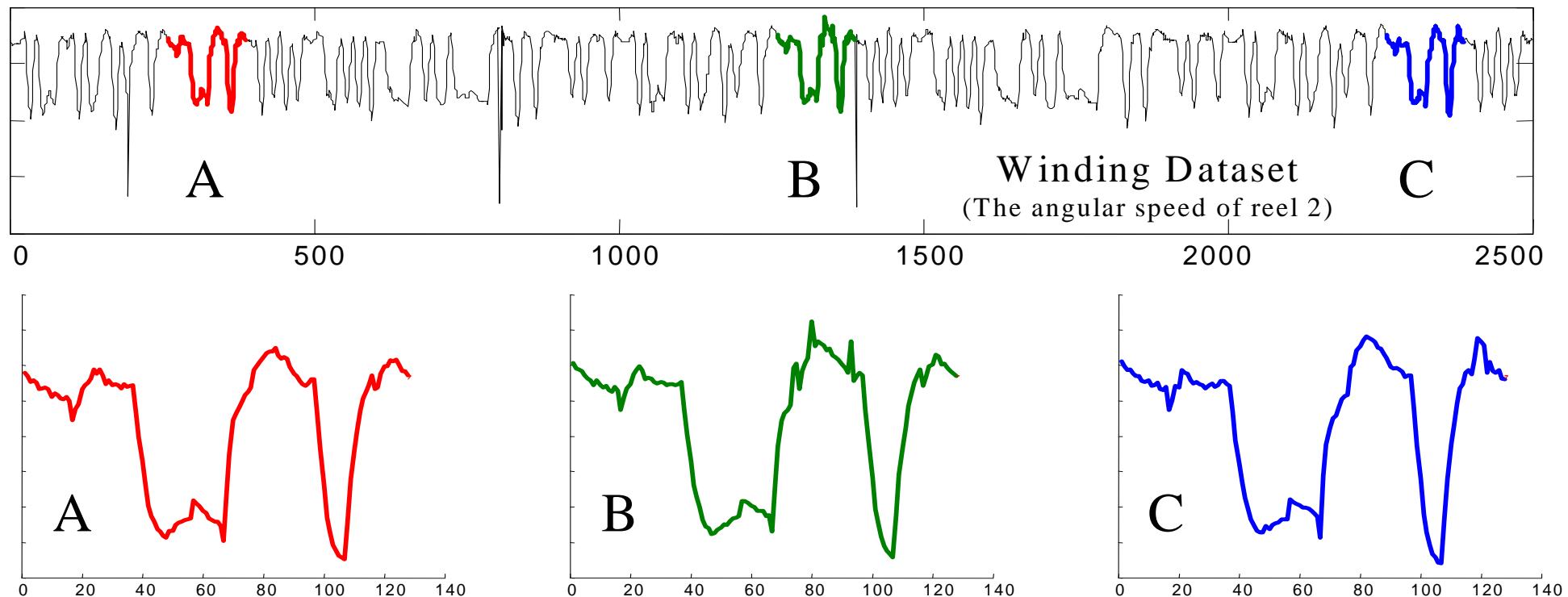


Figure 11: The fraction P , of the ECG database that must be examined by the three dimensionality reduction techniques being compared over a range of query lengths (256-1024) and dimensionalities (16-64).

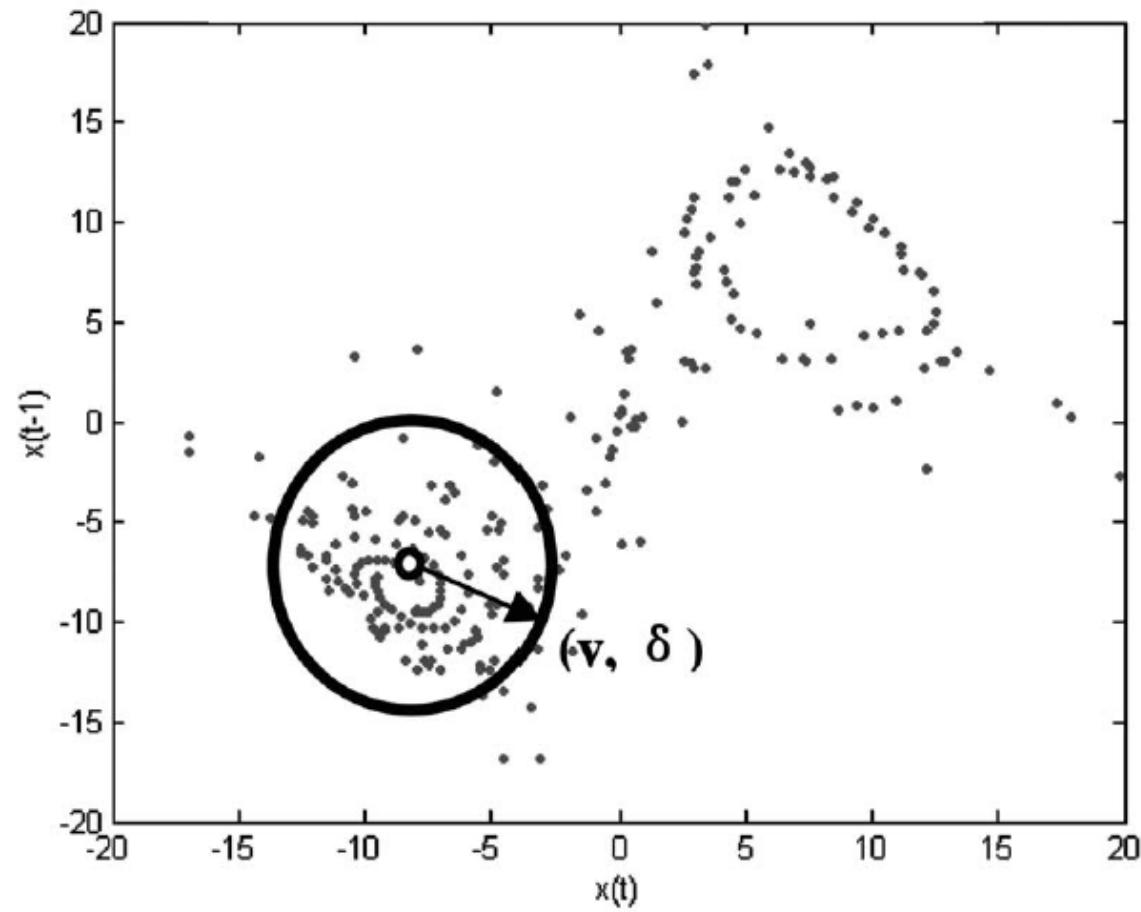
5. Time series data mining

Motif discovery



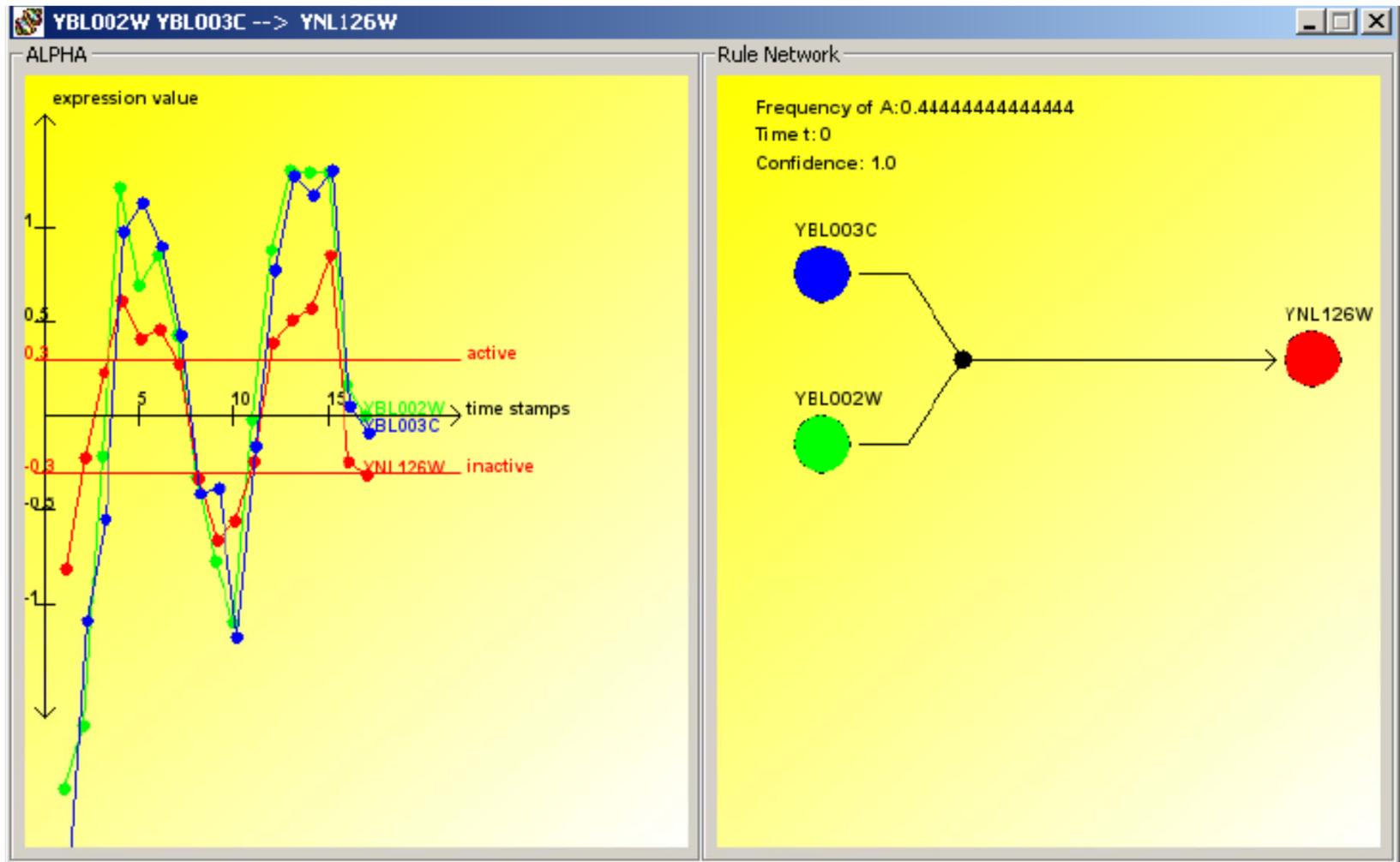
5. Time series data mining

Motif discovery



5. Time series data mining

Rule extraction

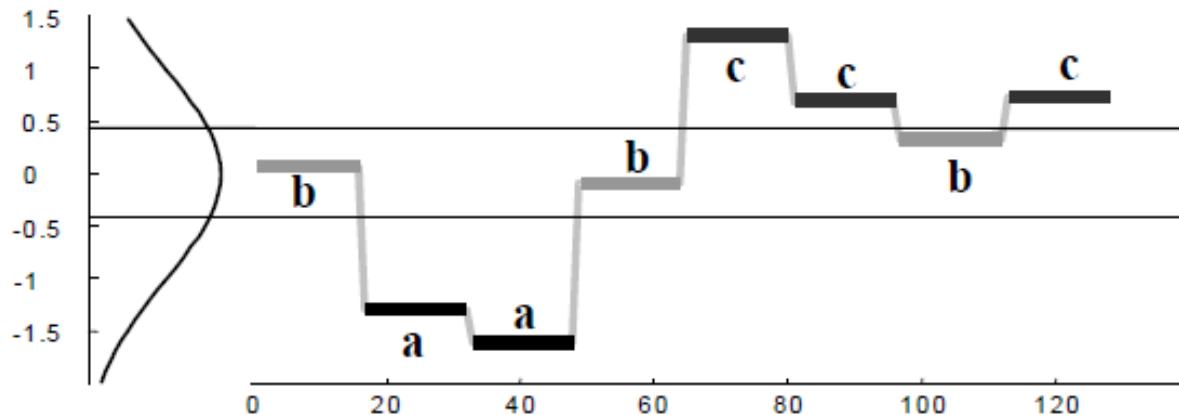


5. Time series data mining

Rule extraction

Step 1: Convert time series into a symbol sequence

$$y[n] = \frac{x[n] - x[n-1]}{x[n]}$$



5. Time series data mining

Rule extraction

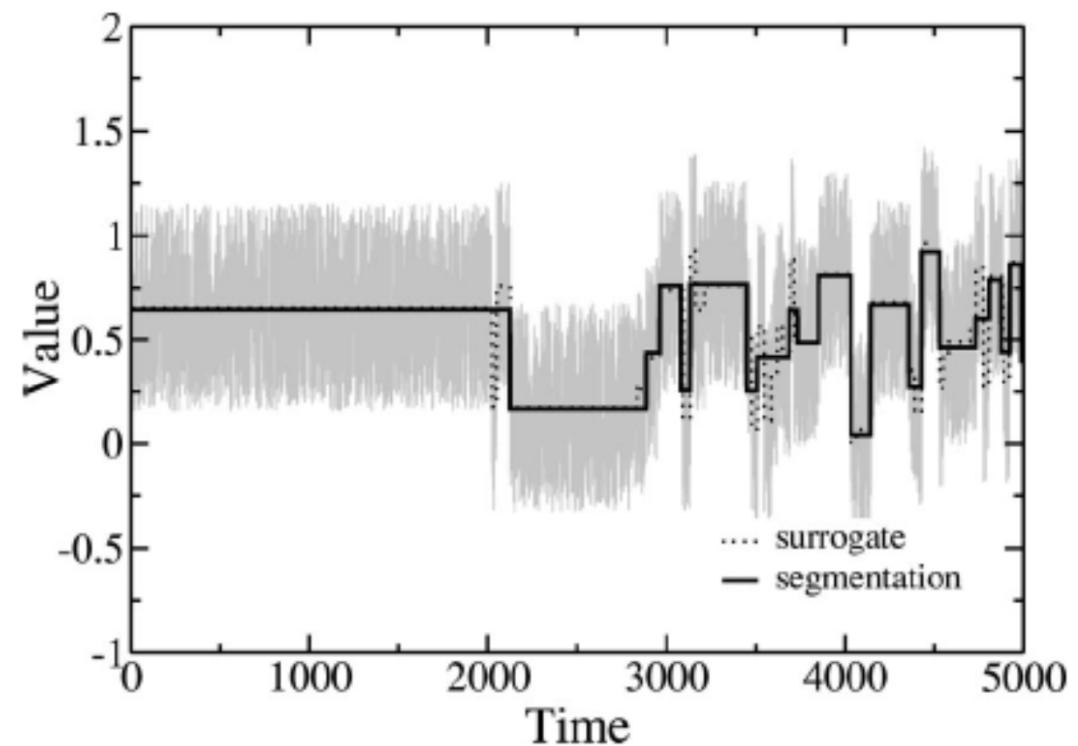
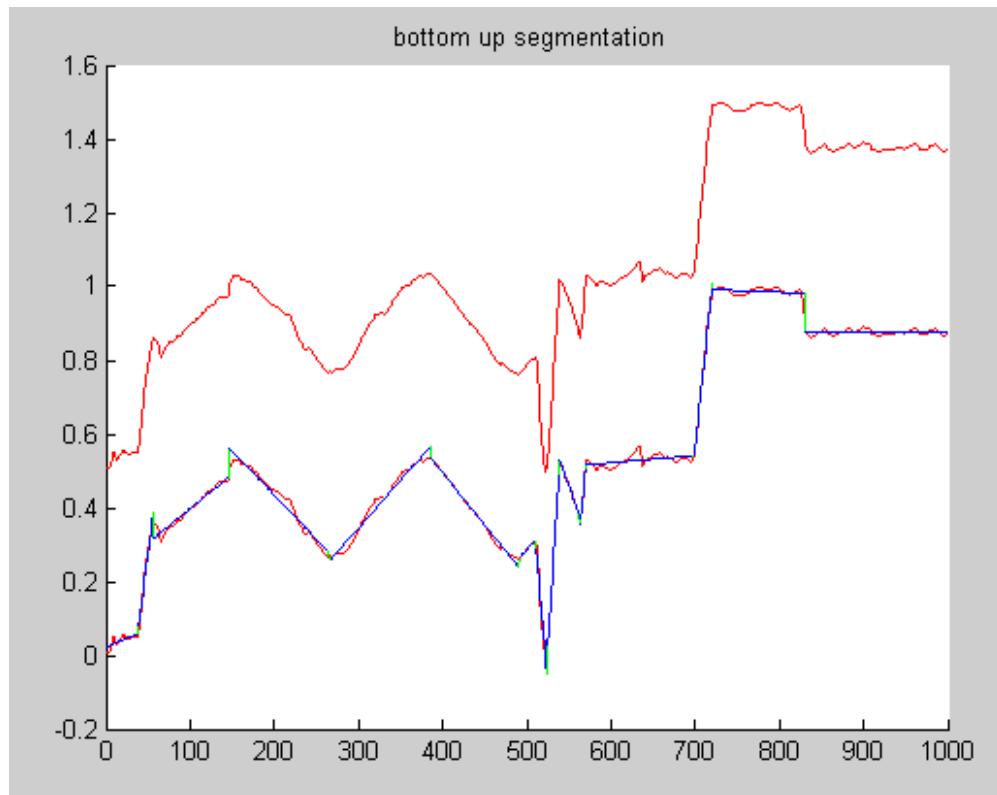
Step 2: Identify frequent itemsets

$x[n]=abbcaab\textcolor{red}{bcab}ababcaab\textcolor{red}{bcabb}cabc\textcolor{red}{bcab}baabc\textcolor{red}{bcab}bc\textcolor{black}{bc}$

	<u>2-item set</u>	<u>3-item set</u>	<u>4-item set</u>	<u>5-item set</u>
	aa 3	aba 2	abba 1	bcaba 1
	ab 11	abb 5	abbb 0	bcabb 3
	ac 1	abc 4	abbc 4	bcabc 1
	ba 2	bba 1	bcaa 2	
	bb 5	bbb 0	bcab 5	
Min.Support=5	bc 11	bbc 4	bcac 0	
	ca 7	bca 7	caba 1	
	cb 3	bcb 3	cabb 3	
	cc 0	bcc 0	cabc 1	
		caa 2		
		cab 5		
		cac 0		

5. Time series data mining

Segmentation (Change Point Detection)



Summarization

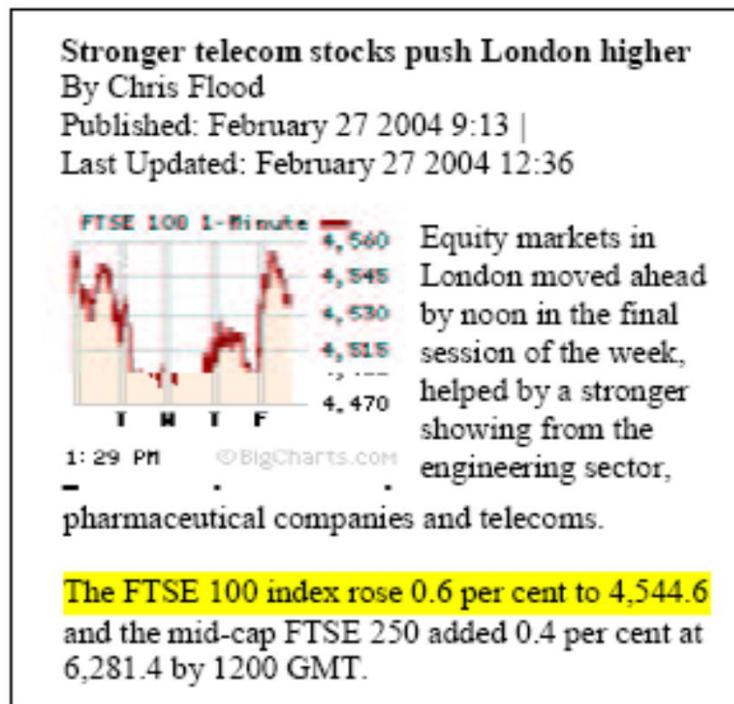


Figure 1: Excerpt from *Financial Times Online*

Table 2. Example Wind Speed and Direction Data

Day	Hour	Wind Direction	Wind Speed (Knots)
20-1-01	600	S	8
20-1-01	900	S	6
20-1-01	1200	S	7
20-1-01	1500	S	10
20-1-01	1800	S	12
20-1-01	2100	S	16
21-1-01	0000	S	20

S 08-12 (*just 1500 value*)

S 06-10 INCREASING 18-22 BY MIDNIGHT (0600, 0000 vals)

S 06-10 DECREASING 05-09 BY MIDDAY, INCREASING 10-14 BY EARLY EVENING AND 18-22 BY MIDNIGHT (0600, 1200, 1800, 0000 values)

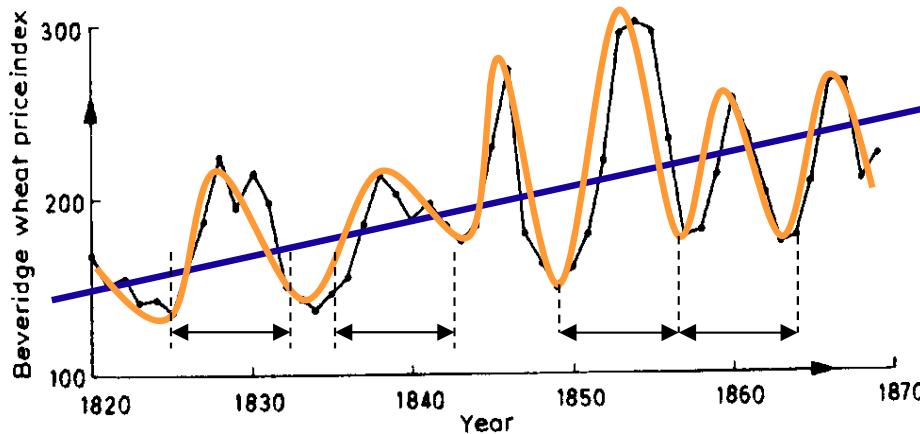
S 06-10 DECREASING 04-08 BY LATE MORNING, INCREASING 05-09 BY MIDDAY, 08-12 BY AFTERNOON, 10-14 BY EARLY EVENING, 14-18 BY MID EVENING, AND 18-22 BY MIDNIGHT (0600, 0900, 1200, 1500, 1800, 2100, 0000 values)

Figure 2. A few possible summaries of Table 2 Data

Session outline

1. Forecasting
2. Univariate forecasting
3. Intervention modelling
4. State-space modelling
5. Time series data mining
 1. Time series representation
 2. Distance measure
 3. Anomaly/Novelty detection
 4. Classification/Clustering
 5. Indexing
 6. Motif discovery
 7. Rule extraction
 8. Segmentation
 9. Summarization

Conclusions



Preprocessing
(heteroskedasticity,
gaussianity, outliers, ...)?



$$x[n] = \text{trend}[n] + \text{periodic}[n] + \text{random}[n]$$

(F)ARIMA SARIMA

Explained by statistical models (AR, MA, ARMA)

Harmonic analysis,
Filtering

Regression, Curve fitting

Bibliography

- C. Chatfield. *The analysis of time series: an introduction*. Chapman & Hall, CRC, 1996.
- C. Chatfield. *Time-series forecasting*. Chapman & Hall, CRC, 2000.
- D.S.G. Pollock. *A handbook of time-series analysis, signal processing and dynamics*. Academics Press, 1999.
- D. Peña, G. C. Tiao, R. S. Tsay. *A course in time series analysis*. John Wiley and Sons, Inc. 2001