Time Series Analysis

Lasse Engbo Christiansen

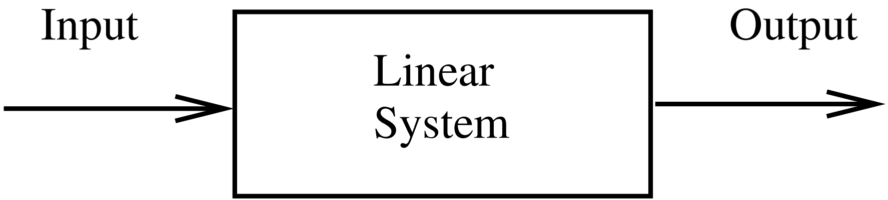
Department of Applied Mathematics and Computer Science
Technical University of Denmark

October 28, 2016

Outline of the lecture

- Input-Output systems, sec. 4 introduction and 4.1
- Linear system notation
- The z-transform, section 4.4
- Cross Correlation Functions – from Sec. 6.2.2
- Transfer function models; identification, estimation, validation, prediction, Chap. 8

Linear Dynamic Systems



- We are going to study the case where we measure the input and the output to/from a system
- Here we will discuss some theory and descriptions for such systems

Dynamic response

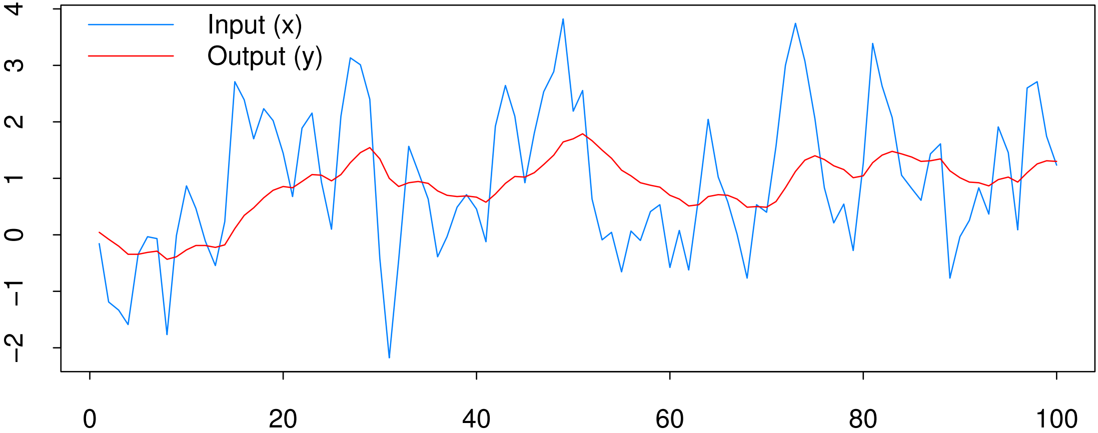
What would happen to the temperature inside a hollow, insulated, concrete block, if you

- place it in a controlled temperature environment at $20^o C$,
- wait until everything is settled (all temperatures are equal), and then
- suddenly raise the temperature by $100^o C$ outside the block?

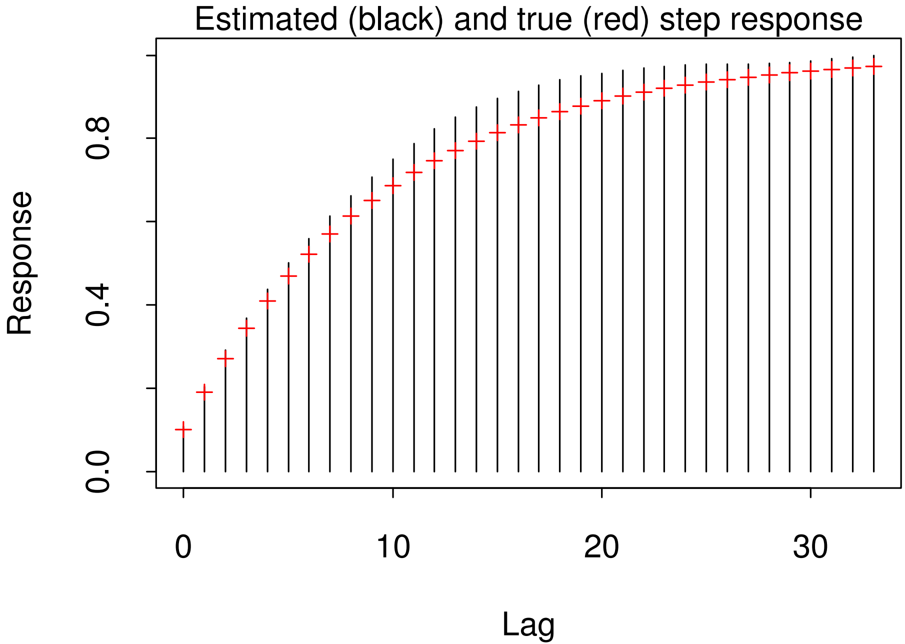
Sketch the temporal development of the temperature outside and inside the block

Dynamic response characteristics from data

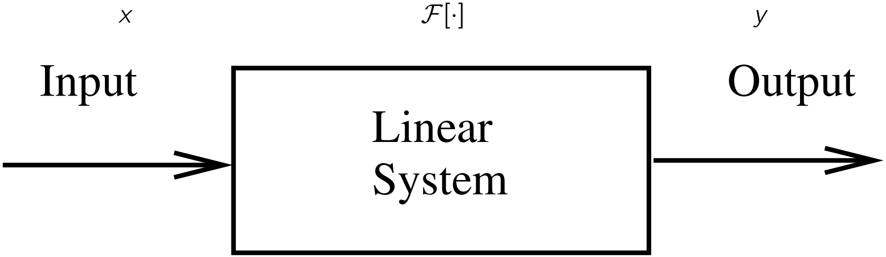
- An important aspect of what we aim at later on is to identify the characteristics of the dynamic response based on measurements of input and output signals



Dyn. response characteristics from data II



Linear Dynamic Systems – notation



$x(t)$	Differential eq., $h(u)$	$y(t)$
x_t	Difference eq., h_k, $h(B)$	y_t
$X(\omega)$	$\mathcal{H}(\omega)$	$Y(\omega)$
$X(z)$	$H(z)$	$Y(z)$
$(X(s)$	$H(s)$	$Y(s))$

Dynamic Systems – Some characteristics

Def. Linear system:

$$\mathcal{F}\left[\lambda_1 x_1(t) + \lambda_2 x_2(t)\right] = \lambda_1 \mathcal{F}\left[x_1(t)\right] + \lambda_2 \mathcal{F}\left[x_2(t)\right]$$

Def. Time invariant system:

$$y(t) = \mathcal{F}\left[x(t)\right] \Rightarrow y(t-\tau) = \mathcal{F}\left[x(t-\tau)\right]$$

Def. Stable system: A system is said to be *stable* if any constrained input implies a constrained output.

Def. Causal system: A system is said to be *physically feasible* or *causal*, if the output at time t does not depend on future values of the input.

Description in the time domain (Convolution)

For *linear time invariant systems*:

- Continuous time:
$$y(t) = \int_{-\infty}^{\infty} h(u)x(t-u)\,du \tag{1}$$
- Discrete time:
$$y_t = \sum_{k=-\infty}^{\infty} h_k x_{t-k} \tag{2}$$

- $h(u)$ or h_k is called the *impulse response*
- $S_k = \sum_{j=-\infty}^{k} h_j$ is called the *step response* (similar def. in continuous time)
- The impulse response can be determined by "sending a 1 through the system" - $x_t = 1$ for $t = 0$ and zero otherwise

Example

- System: $y_t - ay_{t-1} = bx_t$
- Can be written: $y_t = bx_t + ay_{t-1} = bx_t + a(bx_{t-1} + ay_{t-2})$ or

$$y_t = b(x_t + ax_{t-1} + a^2x_{t-2} + a^3x_{t-3} + \ldots) = b\sum_{k=0}^{\infty} a^k x_{t-k}$$

- Is the system *linear* and *time invariant*? – Yes and yes.
- The *impulse response* is $h_k = ba^k, \;\; k \geq 0$ (0 otherwise)
- Is the system causal? – Yes

$$\sum_{k=-\infty}^{\infty} |h_k| = \sum_{k=0}^{\infty} |b||a|^k = \begin{cases} |b|/(1-|a|) & ; \quad |a| < 1 \\ \infty & ; \quad |a| \geq 1 \end{cases}$$

- Is the system *stable*? – Yes, for $|a| < 1$ (stability does not depend on b)

Stability based on the impulse response function

If the impulse response function is absolutely convergent, the system is stable (Theorem 4.3).

- Continuous time:

$$\int_{-\infty}^{\infty} |h(u)| \mathrm{d}u < \infty$$

- Discrete time:

$$\sum_{k=-\infty}^{\infty} |h_k| < \infty$$

Example: Calculating the impulse response function.

The impulse response can be determined by 'sending a 1 through the system'. Consider the linear, time invariant system

$$y_t - 0.8y_{t-1} = 2x_t - x_{t-1}$$

By putting $x = \delta_k = \delta_{0k}$ (Kronecker delta) we see that $y_k = h_k = 0$ for $k < 0$. For $k = 0$ we get

$$\begin{aligned} y_0 &= 0.8y_{-1} + 2\delta_0 - \delta_{-1} \\ &= 0.8 \times 0 + 2 \times 1 - 0 = 2 \end{aligned}$$

i.e. $h_0 = 2$.

Example continued

$$y_t - 0.8y_{t-1} = 2x_t - x_{t-1}$$

Going on we get

$$\begin{aligned} y_1 &= 0.8y_0 + 2\delta_1 - \delta_0 = 0.8 \times 2 + 2 \times 0 - 1 = 0.6 \\ y_2 &= 0.8y_1 = 0.48 \\ y_k &= 0.8^{k-1}0.6 \ (k > 0) \end{aligned}$$

Hence, the impulse response function is

$$h_k = \begin{cases} 0 & \text{for } k < 0 \\ 2 & \text{for } k = 0 \\ 0.8^{k-1}0.6 & \text{for } k > 0 \end{cases}$$

which clearly represents a causal system. Furthermore, the system is stable since $\sum_0^\infty |h_k| = 2 + 0.6(1 + 0.8 + 0.8^2 + \cdots) = 5 < \infty$

The z-transform

- A way to describe dynamical systems in discrete time

$$Z(\{x_t\}) = X(z) = \sum_{t=-\infty}^{\infty} x_t z^{-t} \qquad (z \in \mathbb{C})$$

- The z-transform of a time delay: $Z(\{x_{t-\tau}\}) = z^{-\tau} X(z)$
- The *transfer function* of the system is called $H(z) = \sum_{t=-\infty}^{\infty} h_t z^{-t}$

$$y_t = \sum_{k=-\infty}^{\infty} h_k x_{t-k} \quad \Leftrightarrow \quad Y(z) = H(z)X(z)$$

Linear Difference Equation

$$y_t + a_1 y_{t-1} + \cdots + a_p y_{t-p} = b_0 x_{t-\tau} + b_1 x_{t-\tau-1} + \cdots + b_q x_{t-\tau-q}$$
$$(1 + a_1 z^{-1} + \cdots + a_p z^{-p}) Y(z) = z^{-\tau}(b_0 + b_1 z^{-1} + \cdots + b_q z^{-q}) X(z)$$

Transfer function:

$$\begin{aligned} H(z) &= \frac{z^{-\tau}(b_0 + b_1 z^{-1} + \cdots + b_q z^{-q})}{(1 + a_1 z^{-1} + \cdots + a_p z^{-p})} \\ &= \frac{z^{-\tau}(1 - n_1 z^{-1})(1 - n_2 z^{-1}) \cdots (1 - n_q z^{-1}) b_0}{(1 - \lambda_1 z^{-1})(1 - \lambda_2 z^{-1}) \cdots (1 - \lambda_p z^{-1})} \end{aligned}$$

Where the roots $n_1, n_2, \ldots, n_q$ are called the *zeros of the system* and $\lambda_1, \lambda_2, \ldots, \lambda_p$ are called the *poles of the system*

The system is stable if all poles lie within the unit circle

Relation to the backshift operator

$$
\begin{aligned}
y_t + a_1 y_{t-1} + \cdots + a_p y_{t-p} &= b_0 x_{t-\tau} + b_1 x_{t-\tau-1} + \cdots + b_q x_{t-\tau-q} \\
(1 + a_1 z^{-1} + \cdots + a_p z^{-p}) Y(z) &= z^{-\tau}(b_0 + b_1 z^{-1} + \cdots + b_q z^{-q}) X(z) \\
(1 + a_1 B^1 + \cdots + a_p B^p) y_t &= B^\tau (b_0 + b_1 B^1 + \cdots + b_q B^q) x_t \\
\varphi(B) y_t &= \omega(B) B^\tau x_t
\end{aligned}
$$

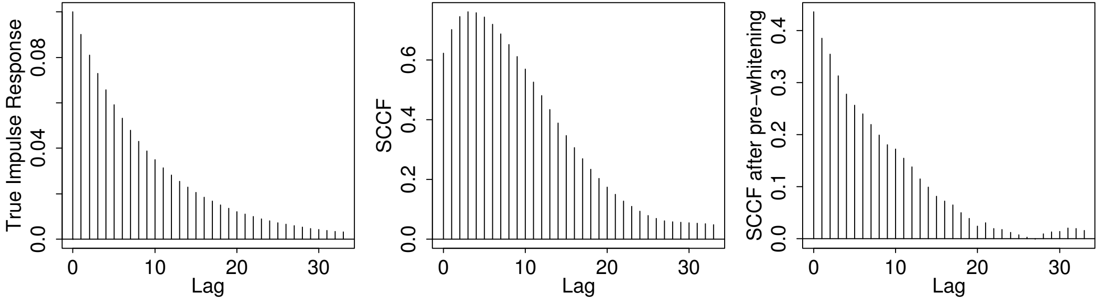
The output can be written:

$$
y_t = \varphi^{-1}(B)\omega(B)B^\tau x_t = h(B)x_t = \left[\sum_{i=0}^{\infty} h_i B^i\right] x_t = \sum_{i=0}^{\infty} h_i x_{t-i}
$$

$h(B)$ is *also* called the *transfer function*. Using $h(B)$ the system is assumed to be causal; compare with $H(z) = \sum_{t=-\infty}^{\infty} h_t z^{-t}$

Estimating the impulse response

- The poles and zeros characterize the impulse response (Appendix A and Chapter 8)
- If we can estimate the impulse response from recordings of input and output we can get information that allows us to *suggest a structure for the transfer function*



Cross covariance and cross correlation functions

Estimate of the cross covariance function:

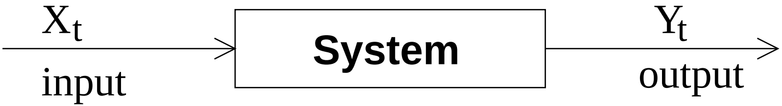
$$\begin{aligned} C_{XY}(k) &= \frac{1}{N}\sum_{t=1}^{N-k}(X_t - \overline{X})(Y_{t+k} - \overline{Y}) \\ C_{XY}(-k) &= \frac{1}{N}\sum_{t=1}^{N-k}(X_{t+k} - \overline{X})(Y_t - \overline{Y}) \end{aligned}$$

Estimate of the cross correlation function:

$$\widehat{\rho}_{XY}(k) = C_{XY}(k)/\sqrt{C_{XX}(0)C_{YY}(0)}$$

If at least one of the processes is white noise and if the processes are uncorrelated then $\widehat{\rho}_{XY}(k)$ is approximately normally distributed with mean 0 and variance $1/N$

Systems without measurement noise

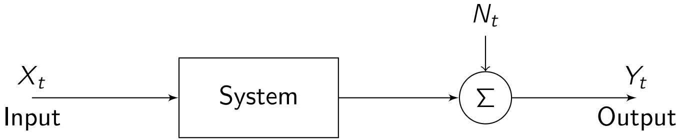


$$Y_t = \sum_{i=-\infty}^{\infty} h_i X_{t-i}$$

Given γ_{XX} and the system description we obtain

$$\begin{aligned}
\gamma_{YY}(k) &= \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j \gamma_{XX}(k-j+i) \\
\gamma_{XY}(k) &= \sum_{i=-\infty}^{\infty} h_i \gamma_{XX}(k-i).
\end{aligned}$$

Systems with measurement noise



$$Y_t = \sum_{i=-\infty}^{\infty} h_i X_{t-i} + N_t.$$

Given γ_{XX} and γ_{NN} we obtain

$$\gamma_{YY}(k) = \sum_{i=-\infty}^{\infty} \sum_{j=-\infty}^{\infty} h_i h_j \gamma_{XX}(k-j+i) + \gamma_{NN}(k)$$

$$\gamma_{XY}(k) = \sum_{i=-\infty}^{\infty} h_i \gamma_{XX}(k-i).$$

IMPORTANT ASSUMPTION: No feedback in the system.

Estimating the impulse response

- On a previous slide we saw that we got a good picture of the true impulse response when *pre-whitening* the data
- The reason is

$$\gamma_{XY}(k) = \sum_{i=-\infty}^{\infty} h_i \gamma_{XX}(k-i)$$

- and only if $\{X_t\}$ is white noise then we get $\boxed{\gamma_{XY}(k) = h_k \sigma_X^2}$
- Therefore if $\{X_t\}$ is white noise the SCCF $\hat{\rho}_{XY}(k)$ is proportional to $\hat{h}_k$
- Normally $\{X_t\}$ is not white noise – we fix this using pre-whitening

Pre-whitening

a) A suitable ARMA-model is applied to the input series:

$$\eta(B)X_t = \nu(B)\alpha_t.$$

b) We perform a *prewhitening* of the input series

$$\alpha_t = \nu(B)^{-1}\eta(B)X_t$$

c) The output–series $\{Y_t\}$ is filtered with the same model, i.e.

$$\beta_t = \nu(B)^{-1}\eta(B)Y_t.$$

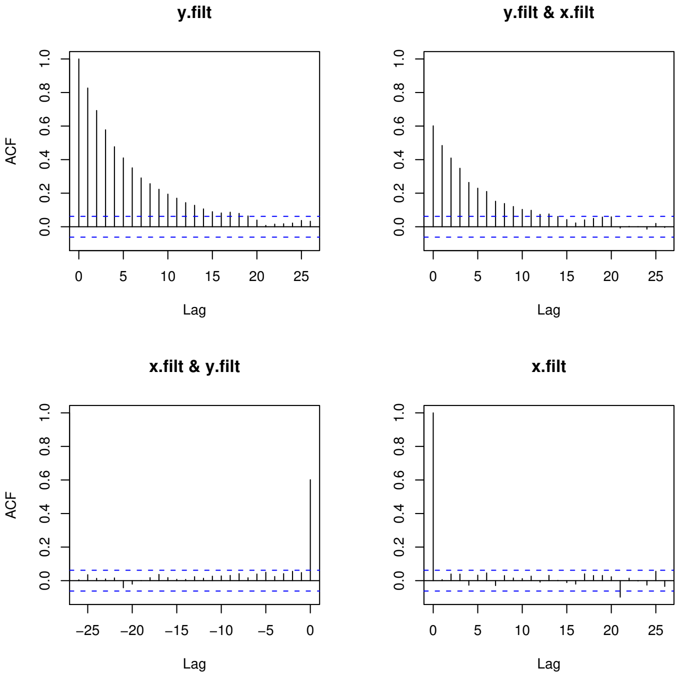
d) Now the *impulse response function is estimated* by

$$\widehat{h}_k = C_{\alpha\beta}(k)/C_{\alpha\alpha}(0) = C_{\alpha\beta}(k)/S_\alpha^2.$$

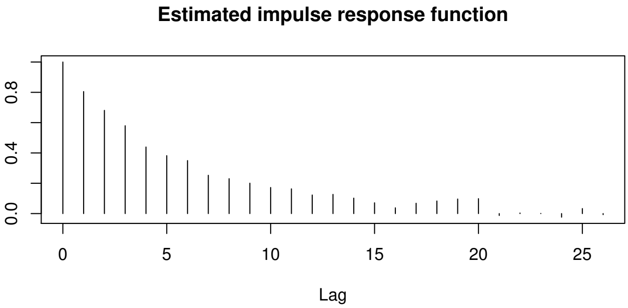
Example using R

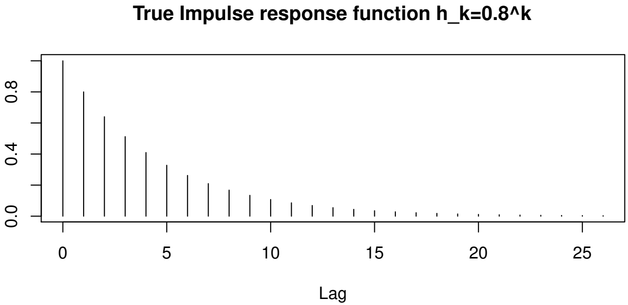
```
## ARMA structure for x; AR(1)
x.struct<-c(1,0,0)
## Estimate the model (check for convergence):
x.fit <- arima(x, order=x.struct, include.mean=F)
## Filter x and y:
x.filt <- x - x.fit$coef[1] * c(0,x[1:(length(x)-1)])
y.filt <- y - x.fit$coef[1] * c(0,y[1:(length(y)-1)])
##Estimate SCCF for the filtered series:
acf(cbind(y.filt, x.filt), type="correlation")
```

Graphical output

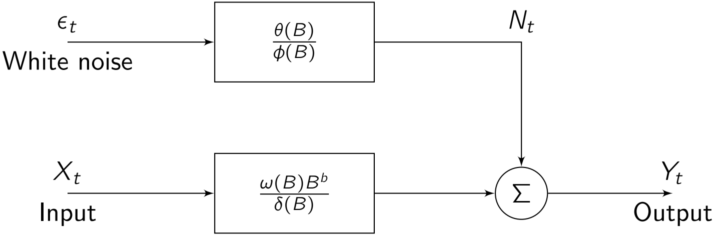


Impulse response functions





Transfer function models



$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + \frac{\theta(B)}{\varphi(B)} \varepsilon_t$$

- Also called Box-Jenkins models
- Can be extended to include more inputs – see the book.

Some names

The following are all sub-models of transfer function models.

- FIR: Finite Impulse Response (impulse response function(s) of finite length)
- ARX: Auto Regressive with eXogenous input
- ARMAX/CARMA: Auto Regressive Moving Average with eXogenous input / Controlled ARMA (Common poles in transfer functions)
- OE: Output Error (No model for the observation noise)

Regression models with ARMA noise (the `xreg` option to `arima` in R)

Identification of transfer function models

$$h(B) = \frac{\omega(B)B^b}{\delta(B)} = h_0 + h_1 B + h_2 B^2 + h_3 B^3 + h_4 B^4 + \dots$$

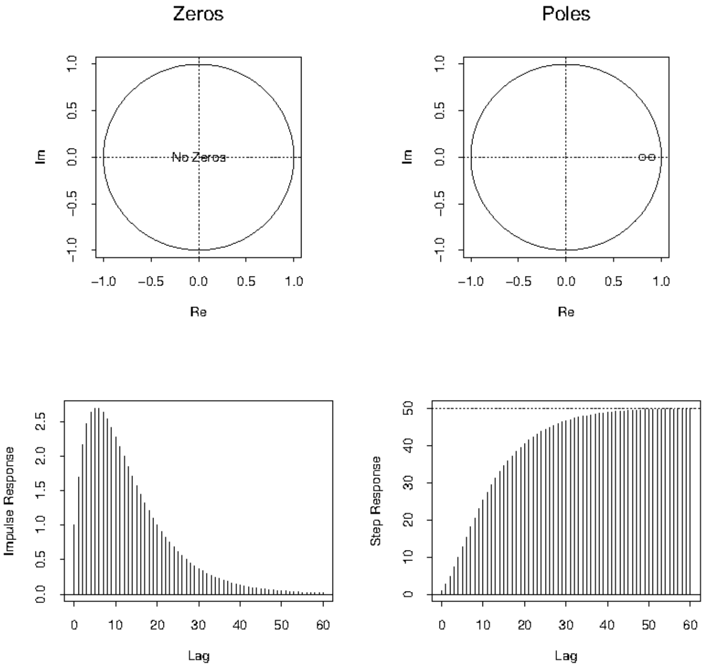
- Using pre-whitening we estimate the impulse response and "guess" an appropriate structure of $h(B)$ based on this.
- It is a good idea to experiment with some structures. With Matlab's "ident" toolbox (use q^{-1} instead of B):

```
A = 1; B = 1; C = 1; D = 1; F = [1 -2.55 2.41 -0.85];
mod = idpoly(A, B, C, D, F, 1, 1)
impulse(mod)
```

- PEZdemo (complex poles/zeros should be in pairs): `http://users.ece.gatech.edu/mcclella/matlabGUIs/`

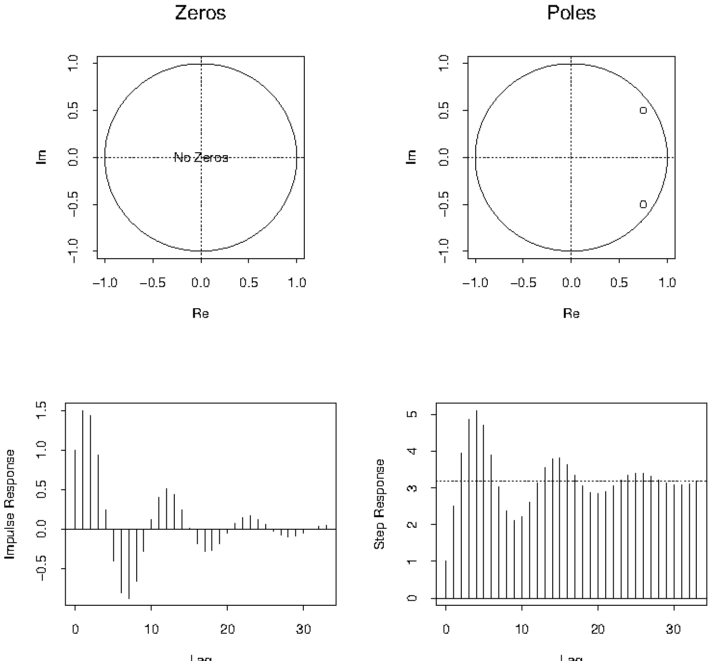
2 real poles

$$h(B) = \frac{1}{1 - 1.7B + 0.72B^2}$$



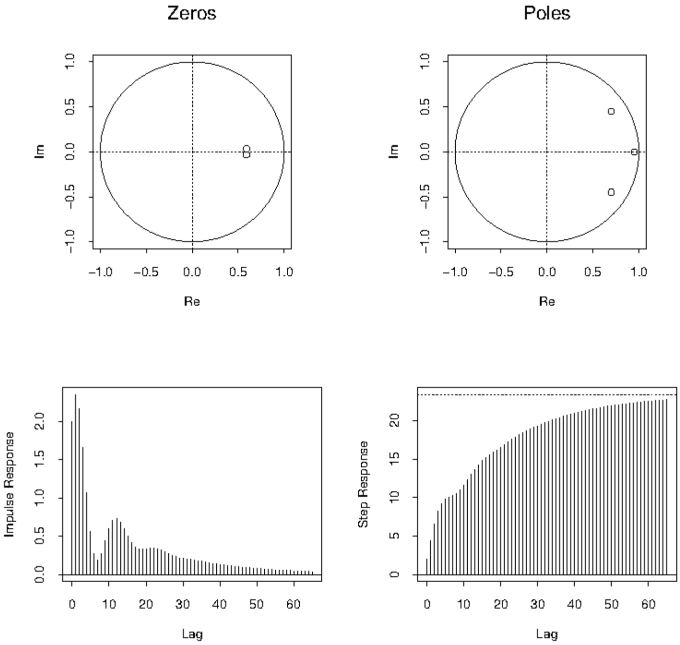
2 complex

$$h(B) = \frac{1}{1 - 1.5B + 0.81B^2}$$



1 real, 2 comp

$$h(B) = \frac{2 - 2.35B + 0.69B^2}{1 - 2.35B + 2.02B^2 - 0.66B^3}$$



Identification of the transfer function for the noise

- After selection of the structure of the transfer function of the input we estimate the parameters of the model
$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + N_t$$
- We extract the residuals $\{N_t\}$ and identifies a structure for an ARMA model of this series
$$N_t = \frac{\theta(B)}{\varphi(B)} \varepsilon_t \quad \Leftrightarrow \quad \varphi(B) N_t = \theta(B) \varepsilon_t$$
- We then have the full structure of the model and we estimate all parameters simultaneously

Estimation

- Form 1-step predictions, treating the input $\{X_t\}$ as known in the future (if $\{X_t\}$ is really stochastic we *condition* on the observed values)
- Select the parameters so that the sum of squares of these errors is as small as possible
- If $\{\varepsilon_t\}$ is normal then the ML estimates are obtained
- For FIR and ARX models we can write the model as $\boldsymbol{Y}_t = \boldsymbol{X}_t^T \boldsymbol{\theta} + \varepsilon_t$ and use LS-estimates
- Moment estimates: Based on the structure of the transfer function we find the theoretical impulse response and we make a match with the lowest lags in the estimated impulse response
- Output error estimates . . .

Model validation

As for ARMA models with the additions:

- Test for cross correlation between the residuals and the input

$$\hat{\rho}_{\varepsilon X}(k) \sim Norm(0, 1/N)$$

  which is (approximately) correct when $\{\varepsilon_t\}$ is white noise and when there is no correlation between the input and the residuals

- A *Portmanteau test*(*Ljung-Box*) can also be performed

Prediction $\widehat{Y}_{t+k|t}$

We must consider two situations

- The input is controllable, i.e. we can decide it and we can predict under different input-scenarios. In this case the prediction error variance is originating from the ARMA-part only (N_t).
- The input is only known until the present time point t and to predict the output we must predict the input. In this case the prediction error variance depend also on the autocovariance of the input process. In the book the case where the input can be modelled as an ARMA-process is considered.

Prediction II

$$\widehat{Y}_{t+k|t} = \sum_{i=0}^{k-1} h_i \widehat{X}_{t+k-i|t} + \sum_{i=k}^{\infty} h_i X_{t+k-i} + \widehat{N}_{t+k|t}.$$

$$Y_{t+k} - \widehat{Y}_{t+k|t} = \sum_{i=0}^{k-1} h_i (X_{t+k-i} - \widehat{X}_{t+k-i|t}) + N_{t+k} - \widehat{N}_{t+k|t}$$

- If the input is controllable then $\widehat{X}_{t+k-i|t} = X_{t+k-i}$
- The book also considers the case where output is known until time t and input until time $t + j$

Prediction III

- We have

$$N_t = \sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}$$

- And if we model the input as an ARMA-process we have

$$X_t = \sum_{i=0}^{\infty} \overline{\psi}_i \eta_{t-i}$$

- And thereby we get:

$$V[Y_{t+k} - \widehat{Y}_{t+k|t}] = \sigma_\eta^2 \sum_{\ell=0}^{k-1} \left(\sum_{i_1+i_2=\ell} h_{i_1} \overline{\psi}_{i_2} \right)^2 + \sigma_\varepsilon^2 \sum_{i=0}^{k-1} \psi_i^2$$

Example: Prediction of a transfer function model

- Consider the system

$$
\begin{aligned}
Y_t &= h(B)X_t + N_t \\
&= h(B)X_t + \Psi(B)\epsilon_t \\
&= \frac{0.4}{1-0.6B}X_t + \frac{1}{1-0.4B}\varepsilon_t, \;\; \sigma_\varepsilon^2 = 0.036
\end{aligned}
$$

- The following data is available:

t	1	2	3	4	5	6	7	8	9	10
Y	2.040	3.050	2.340	2.490	3.300	3.530	2.720	2.460		
X	1.661	4.199	1.991	2.371	3.521	3.269	0.741	2.238	2.544	3.201

- In order to perform a prediction, $Y_{9|8}$, we must filter X with $h(B)$ and forecast $N_{9|8} = \Psi(B)\epsilon_{9|8}$
- We can then evaluate

$$
\hat{Y}_{9|8} = h(B)X_9 + \hat{N}_{9|8}
$$

Example continued

Filtering

For $t \in \{1, \ldots, 10\}$ we evaluate

$$h(B)X_t = 0.4 \cdot \sum_{i=0}^{\infty} (0.6B)^i X_{t-i} \approx 0.4 \cdot \sum_{i=0}^{t-1} (0.6B)^i X_{t-i}$$

And for $t \in \{1, \ldots, 8\}$:

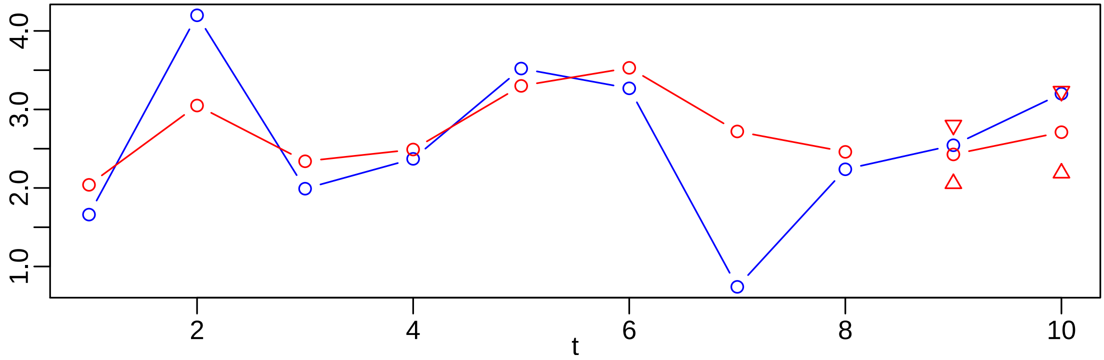
$$N_t = Y_t - h(B)X_t$$

Forecasting

For $k \in \{1, 2\}$:

$$\begin{aligned}
\widehat{N}_{8+k|8} &= N_8 \cdot 0.4^k \\
\widehat{Y}_{8+k|8} &= h(B)X_{8+k} + \widehat{N}_{8+k|8} \\
V(Y_{8+k} - \widehat{Y}_{8+k|8}) &= V(\widehat{Y}_{8+k|8}) = V(\sum_{i=0}^{k-1} \Psi^i \epsilon_{8+k-i}) \\
&= \begin{cases} V(\epsilon_9) = \sigma_\epsilon^2 \\ V(\epsilon_{10}) + 0.4^2 \cdot V(\epsilon_{10}) = (1 + 0.4^2)\sigma_\epsilon^2 \end{cases}
\end{aligned}$$

t	historic								future	
	1	2	3	4	5	6	7	8	9	10
data										
Y	2.04	3.05	2.34	2.49	3.30	3.53	2.72	2.46		
X	1.66	4.20	1.99	2.37	3.52	3.27	0.74	2.24	2.54	3.20
filtered										
$h(B)X_t$	0.66	2.08	2.04	2.17	2.71	2.94	2.06	2.13	2.30	2.66
N	1.38	0.97	0.30	0.32	0.59	0.59	0.66	0.33		
forecasted										
$N_{t\|8}$									0.13	0.05
$Y_{t\|8}$									2.43	2.71



blue: X, red: Y

Intervention models

$$I_t = \begin{cases} 1 & t = t_0 \\ 0 & t \neq t_0 \end{cases}$$

$$Y_t = \frac{\omega(B)}{\delta(B)} I_t + \frac{\theta(B)}{\phi(B)} \varepsilon_t$$

See a real life example in the book.

Highlights

- Def. linear system

$$\mathcal{F}\left[\lambda_1 x_1(t) + \lambda_2 x_2(t)\right] = \lambda_1 \mathcal{F}\left[x_1(t)\right] + \lambda_2 \mathcal{F}\left[x_2(t)\right]$$

- Estimating Cross Covariance function

$$\begin{aligned} C_{XY}(k) &= \frac{1}{N}\sum_{t=1}^{N-k}(X_t - \overline{X})(Y_{t+k} - \overline{Y}) \\ C_{XY}(-k) &= \frac{1}{N}\sum_{t=1}^{N-k}(X_{t+k} - \overline{X})(Y_t - \overline{Y}) \end{aligned}$$

- Transfer function model

$$Y_t = \frac{\omega(B)}{\delta(B)} B^b X_t + \frac{\theta(B)}{\varphi(B)}\varepsilon_t$$

- Use pre-whitening of input
- Fit ARMAX models using the `xreg` option to `arima`