



CEU
*Universidad
San Pablo*

Time Series Analysis

Session III: Probability models for time series

Carlos Óscar Sánchez Sorzano, Ph.D.
Madrid, July 19th 2006

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

1. Goal

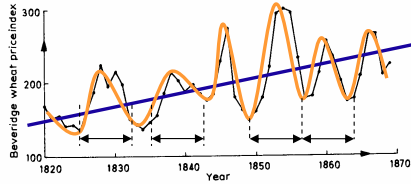
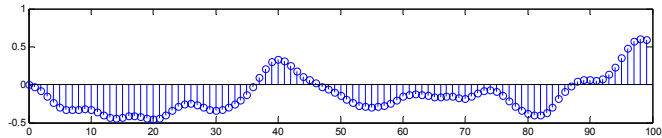
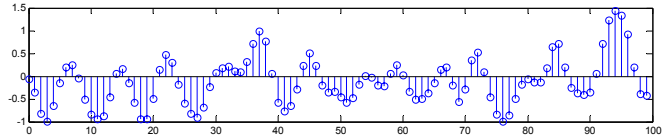
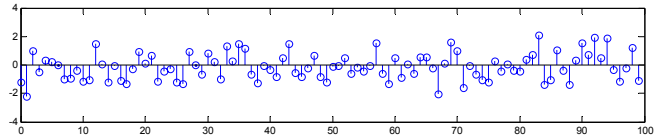


Figure 1.1 Part of the Beverage wheat price index series.

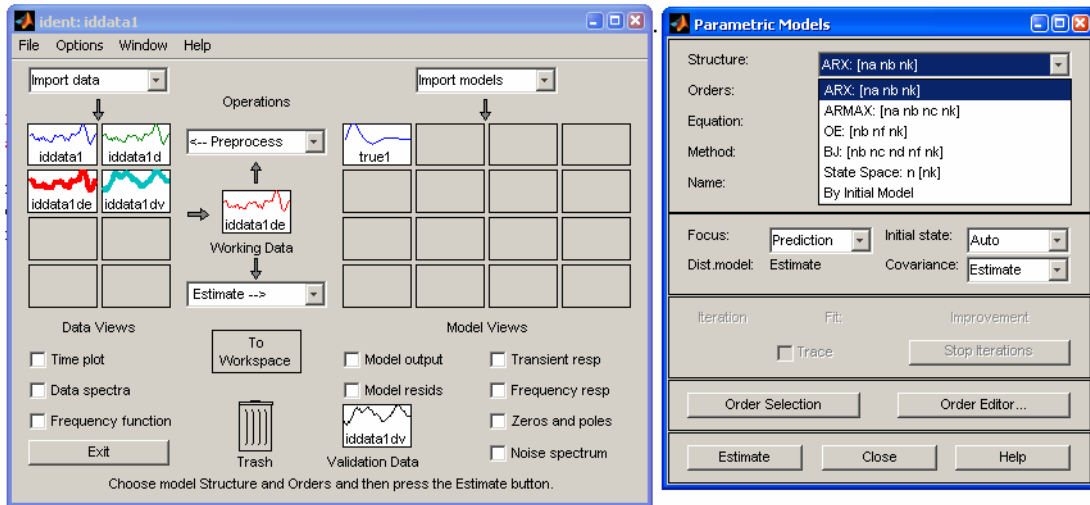


$$x[n] = \text{trend}[n] + \text{periodic}[n] + \text{random}[n]$$

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

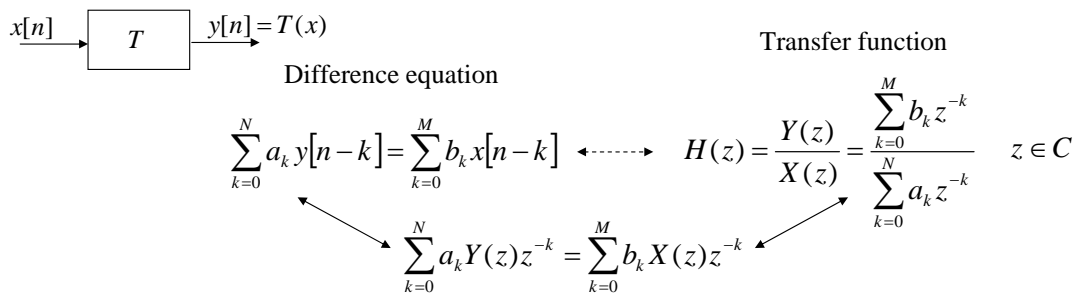
The goal of the probability models for time series is to characterize the kind of “randomness” of the random component. The figure on the right represent three different stochastic processes, all of them are completely random. However, the kind of randomness is different from one to the other. The difference comes from their corresponding correlation structure, i.e., the autocorrelation function for each time series is different. The objective of parametric models is to define a controlled time series whose correlation structure is similar to the one of the input process.

1. Goal



Although, probably none of us will have to program any of this time series models, it is necessary in order to use the programs to know exactly to what they are referring when they are going to fit an ARX(n_a, n_b, n_k) model. For being able of understanding what the program is exactly going to do is necessary to review all the mathematics that follow.

2. A short introduction to system analysis



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

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

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

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

All linear systems can be expressed by a difference equation involving the input series and the output series. If we Z-transform the difference equation, and reorganize the equation we can compute what is called the transfer function of the system. This function completely defines the behaviour of a LTI system. In particular, it relates the spectral content of the input signal to the spectral content of the output. Expliciting the definition of the Z transform falls outside of the scope of this introductory course. The interested student may read Oppenheim2000b.

2. A short introduction to system analysis

Poles/Zeros

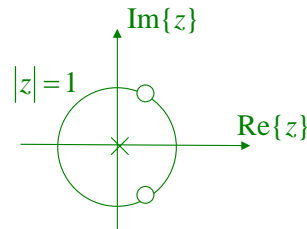
z_0 is a pole of $H(z)$ iff $|H(z_0)| = \infty$

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

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

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

Poles: $z = 0, \infty$ 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

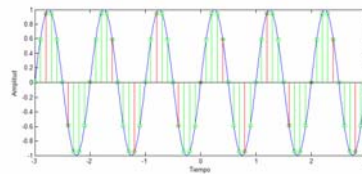
The transfer function of the inverse system of a LTI system whose transfer function is $H(z)$ is $\frac{1}{H(z)}$. Therefore, the zeros of one system are the poles of its inverse, and viceversa.

Poles and zeros play a key role in the analysis of LTI systems since they define (up to a multiplying constant) the transfer function of a given system. Furthermore, they define whether a causal LTI system is stable or not. An interesting property of inverse LTI systems is that they are the reciprocal of the original system. Thus, zeros of the original system become poles of the inverse systems and viceversa. If we wonder about the stability of the inverse system, we could say that the inverse of a given system is stable iff all the zeros of that system are inside the unit circle.

2. A short introduction to system analysis

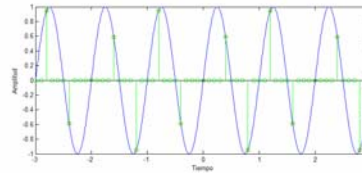
Downsampling

$$x[n] \rightarrow \boxed{\downarrow M} \rightarrow x_d[n] = x[nM]$$



Upsampling

$$x[n] \rightarrow \boxed{\uparrow L} \rightarrow x_u[n] = \begin{cases} x[n/L] & n = 0, \pm L, \pm 2L, \dots \\ 0 & \text{resto} \end{cases} = \sum_{k=-\infty}^{\infty} x[k] \delta[n - kL]$$

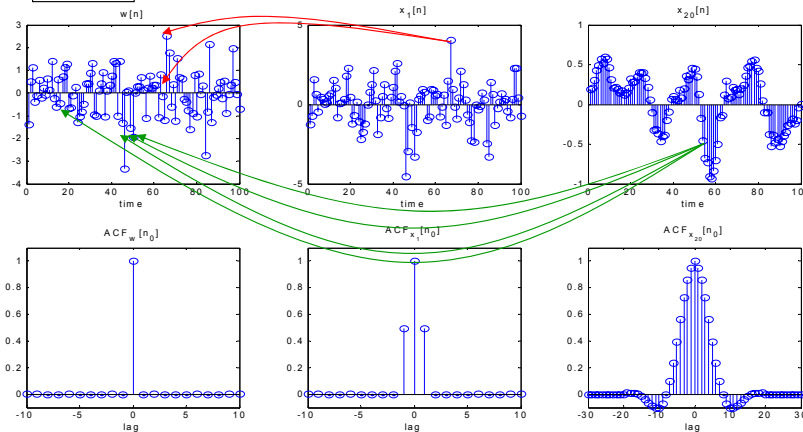


3. Moving average processes: MA(q)

Definition

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

$$w[n] \xrightarrow{MA(q)} x[n] = b_0 w[n] + b_1 w[n-1] + \dots + b_q w[n-q]$$



A MA(q) stochastic process is one that is generated using a difference equation like the one shown in the slide. Note that it only uses previous samples of the input signal. The main features of the associated generating system are that it is LTI, causal and stable. The MA system is FIR and, therefore, an all-zero system.

Selecting the right b coefficients, many different correlation structures might be created. The more complex (and wider) is the correlation we want to reproduce, the higher must be the q. In the examples shown the following b's have been used

```
b1: [1 0.9]
b20: [-0.0438 -0.0151 -0.0079 0.0074 0.0304 0.0591 0.0899 0.1186 0.1408 0.1529
      0.1529 0.1408 0.1186 0.0899 0.0591 0.0304 0.0074 -0.0079 -0.0151 -0.0438]
```

The MATLAB code used for generating these plots is:

```
function MA_process
N=200;

w=randn(1,N);
x1=generate_MA1_process(w);
x20=generate_MA20_process(w);
subplot(231); stem(1:N/2,w(N/2+1:N)); title('w[n]'); xlabel('time');
subplot(232); stem(1:N/2,x1(N/2+1:N)); title('x_1[n]'); xlabel('time');
subplot(233); stem(1:N/2,x20(N/2+1:N)); title('x_{20}[n]'); xlabel('time');

K=1000;
corr_w =xcorr(w,'coeff');
corr_x1 =xcorr(x1,'coeff');
corr_x20=xcorr(x20,'coeff');
for k=1:K
    w=randn(1,N);
    x1=generate_MA1_process(w);
    x20=generate_MA20_process(w);
    corr_w =corr_w +xcorr(w,'coeff');
    corr_x1 =corr_x1 +xcorr(x1,'coeff');
    corr_x20=corr_x20+xcorr(x20,'coeff');
end
corr_w =corr_w/(K+1);
corr_x1 =corr_x1/(K+1);
corr_x20=corr_x20/(K+1);
subplot(234); stem(-10:10,corr_w(N-10:N+10)); title('ACF_w[n_0]');
xlabel('lag'); axis([-10 10 -0.1 1.1]);
subplot(235); stem(-10:10,corr_x1(N-10:N+10)); title('ACF_{x_1}[n_0]');
xlabel('lag'); axis([-10 10 -0.1 1.1]);
```


3. Moving average processes: MA(q)

Statistical properties

$$w[n] \longrightarrow \boxed{MA(q)} \longrightarrow x[n] = \sum_{k=0}^q b_k w[n-k]$$

$$N(0, \sigma_w^2) \longrightarrow N(0, \sigma_w^2 \sum_{k=0}^q b_k^2)$$

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

It has limited support!!

Proof

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

One of the most interesting things of stochastic processes is to characterize them statistically. Normally, we are only interested on their characterization up to a second order (mean, variance, and autocorrelation). If the input time series is normally distributed with zero mean and a given variance, it is easy to show that the output time series is also normally distributed with zero mean and whose variance is a function of the input variance and the system coefficients as shown in the slide.

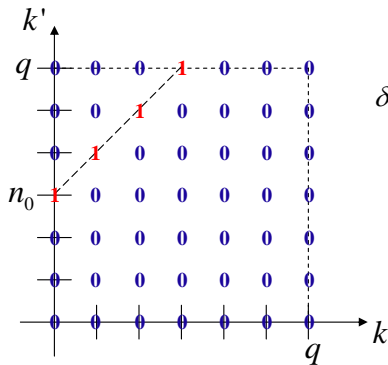
The computation of the autocorrelation function is aq little bit more involved. First, we show the final result and then we prove it. This proof will only be performed for MA processes since for the rest of processes computations are similarly carried out.

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)

Invertibility

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

$$H(z) = \sum_{k=0}^q b_k z^{-k} \longrightarrow H_{inv}(z) = \frac{1}{H(z)} = \frac{1}{\sum_{k=0}^q b_k z^{-k}}$$

$$\sum_{k=0}^q b_k w[n-k] = x[n] \longrightarrow w[n] = \frac{1}{b_0} \left(x[n] - \sum_{k=1}^q b_k w[n-k] \right)$$

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

Invertibility of a stochastic MA(q) process is quite well understood in terms of system analysis. The process is invertible iff all zeros of the MA(q) transfer function are within the unit circle. If there are zeros on or outside the unit circle, then the system is not invertible.

Invertibility is an important issue in stochastic processes because if a given MA(q) is invertible, then there exist a bijective relationship between the process and its corresponding autocorrelation function. That is, there exist a unique MA(q) process that has a given ACF.

The reader interested in knowing more about MA(q) processes may read Chatfield1996b.

3. Moving average processes: generalizations

Model not restricted to be causal

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

$$+ b_{-1} w[n + 1] + \dots + b_{q_f} w[n + q_f] \quad \text{Anticausal component}$$

Model not restricted to be linear

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

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

Volterra Kernels

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



Depending on the application, the MA model needs not be causal and may use future samples of the time series. This is possible for instance in recorded signals or the a posteriori analysis of time series.

A further generalization makes use of nonlinear filtering. This can be performed thanks to the following result:

If $x[n]$ is strictly stationary with finite moments, then it can be represented as the multi-order convolution of a set of causal, stable, non-linear, time-invariant filters. Those interested in this result may read I. W. Sandberg, "Expansions for Discrete-Time Nonlinear Systems," *Circuits, Systems, and Signal Processing* 3, 180-192, (1983).

The student interested in nonlinear time series may read Dwyer2003.

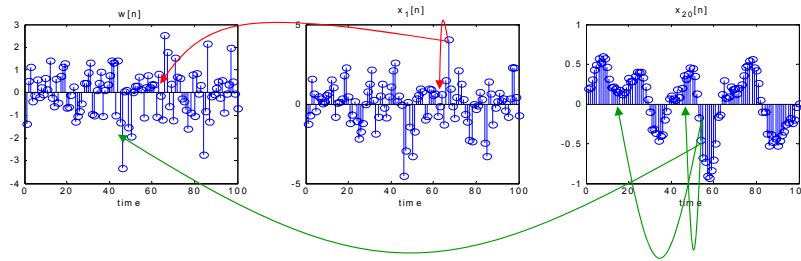
Another nonlinear generalization of the moving average is provided by an expression similar to the linear MA except that, each term $w[n-k]$ is affected by nonlinear function $f(x)$. This model is called GMA (generalized moving average).

4. Autoregressive processes: AR(p)

Definition

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

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



An AR(p) stochastic process is one that is generated using a difference equation like the one shown in the slide. This is a quite general situation in which it is reasonable to think that a given sample of a time series, depends linearly on previous samples plus some random error.

4. Autoregressive processes: AR(p)

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

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

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



The impulse response of the associated system is IIR and its transfer function is of the kind all-pole. Note that, this time, the autocorrelation is not limited and it tends to 0 when the lag tends to infinity, only if the module of all its poles is strictly smaller than 1. That means that if this condition is met, then the AR(p) process is ergodic.

Any AR(p) process can be modelled by a MA process of infinite order. The relationship between this two kind of processes is given by the power series expansion of the transfer function.

The reader interested in knowing more about AR(p) processes may read Chatfield1996b.

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:

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

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

$$z_i \in R \Rightarrow a_1^2 + 4a_2 > 0$$

$$r_X[n_0] = A_1' z_1^{|n_0|} + A_2' z_2^{|n_0|} \longrightarrow \begin{cases} r_X[0] = A_1' + A_2' = 1 \\ r_X[1] = A_1' z_1 + A_2' z_2 = a_1 r_X[0] + a_2 r_X[-1] \end{cases}$$



In general, it is difficult to solve for the A_i coefficients in the autocorrelation function since that would imply a priori knowledge about the input noise power (variance of W). However, it is possible to normalize the autocorrelation function to obtain the correlation coefficients. These coefficients do not depend on the input power and are normalized to 1. So, the first thing to do is to translate the conditions and solution form of the autocorrelation function into conditions and solution form for the correlation coefficients.

Once this is done, it is possible to find two equations to determine the A_i' coefficients. One of the equations is always given by the normalization fact ($r_X[0]=1$), the rest of conditions are given by the $p-1$ first Yule-Walker equations. One important thing to have a sensible solution is that the poles of the system have a module smaller than 1. If the two poles are not real, then the correlation coefficients follow a damped sinusoidal.

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

Definition

$$H(z) = H_{MA(q)}(z)H_{AR(p)}(z) = \frac{b_0 - b_1z^{-1} - b_2z^{-2} - \dots - b_qz^{-q}}{1 - a_1z^{-1} - a_2z^{-2} - \dots - a_pz^{-p}} = \frac{B(z)}{A(z)}$$

$$w[n] \rightarrow \boxed{AR(p)} \rightarrow \boxed{MA(q)} \rightarrow x[n] = \sum_{k=0}^q b_k w[n-k] + \sum_{k=1}^p a_k x[n-k]$$

Statistical properties

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

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

The ARMA(p,q) process corresponds to a mixture of AR(p) and MA(q) by concatenating both systems (the order of the concatenation is not important since both systems are LTI and can be interchanged). The process is ergodic (the autocorrelation function goes to zero as the lag goes to infinity) iff the module of all system poles is smaller than 1. Note, that the poles of the ARMA system are defined exclusively by the AR(p) process, therefore, the ergodic condition is exclusively related to the AR part of the ARMA model.

The importance of the ARMA model is that it usually describes random processes with fewer parameters than the AR and MA processes alone.

The reader interested in knowing more about ARMA(p,q) processes may read Chatfield1996b.

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

Definition

$$x_d[n] = \nabla_t^d x[n] = \overbrace{(\delta[n] - \delta[n-1]) * \dots * (\delta[n] - \delta[n-1])}^d * x[n]$$

$$X_d(z) = (1 - z^{-1})^d X(z) \Rightarrow H_{Int}(z) = \frac{X(z)}{X_d(z)} = \frac{1}{(1 - z^{-1})^d} \rightarrow \text{Poles: } z = 1 \text{ (Multiplicity=d)}$$

↓ Unit root

$$H(z) = \frac{X_d(z)}{W(z)} = H_{ARMA(p,q)}(z)$$

$d \in \mathbb{Q} \Rightarrow \text{FARIMA or ARFIMA}$

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

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

One of the conditions for modelling a time series with AR, MA or ARMA processes is that the time series is stationary. Remember that one of the consequences of stationarity is that the mean and variance at each time instant n is the same for all n . If the time series has a polynomial trend, then it is possible to detrend it by differentiation at an appropriate order. The resulting differentiated time series is now stationary. ARIMA models state that the differentiated time series follow an ARMA model. After integrating, the output will be a reasonable model of the time series being studied.

Usually it is not necessary to differentiate more than once ($d=1$), that would be removing a linear trend. However, it is always difficult to decide how many times we have to integrate. To elucidate these questions there is a number of tests we can perform. All these tests are called “unit root tests”. Among the most famous are the Dickey-Fuller test and the Perron test. The student interested in these tests may read Kwiatkowski1992.

If d is a fraction, then the model is also called FARIMA or ARFIMA (Fractional ARIMA). Fractional differentiation is not something that is intuitive in the time domain. However, fractional differentiation is trivial in Fourier space. This will be further studied in the next session.

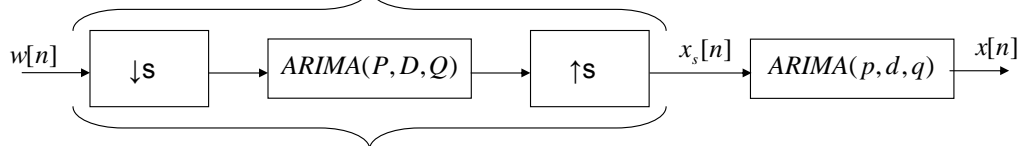
The reader interested in knowing more about ARIMA(p,d,q) processes may read Chatfield1996b.

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

Definition

$$d = 1 \quad (x[n] - x[n-1]) = \sum_{k=0}^q b_k x_s[n-k] + \sum_{k=1}^p a_k (x[n-k] - x[n-k-1])$$

$$D = 1 \quad (x_s[n] - x_s[n-s]) = \sum_{k=0}^Q B_k w[n-ks] + \sum_{k=1}^P A_k (x_s[n-ks] - x_s[n-(k-1)s])$$



$$\frac{X_s(z)}{W(z)} = H_{ARIMA(P,D,Q)}(z^s)$$

$$\frac{X_s(z)}{X(z)} = H_{ARIMA(p,d,q)}(z)$$

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

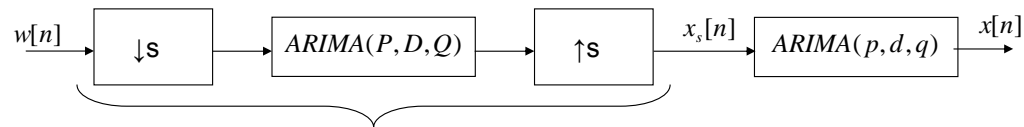
The Seasonal ARIMA is also called the Box-Jenkins model. This model takes into account the possible seasonal component which might be random, too. The seasonal period is assumed to be known, for instance yearly, quarterly, weekly, etc. It will be referred to as s .

The way the time series is generated is by taking white noise, creating a seasonal ARIMA model that would account for the relationship among years (if the seasonality is yearly) and then creating an ARIMA model for what happens inside each year.

The student interested in knowing more about the ARIMA model may read Chatfield1996c.

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

Example: SARIMA(1,0,0)x(0,1,1)₁₂



$$(P, D, Q) = (0, 1, 1) \quad x_s[n] - x_s[n-12] = B_0 w[n] + B_1 w[n-12]$$

$$(p, d, q) = (1, 0, 0)$$

$$x[n] = x_s[n] + a_1 x[n-1] \longrightarrow x_s[n] = x[n] - a_1 x[n-1]$$

$$(x[n] - a_1 x[n-1]) - (x[n-12] - a_1 x[n-13]) = B_0 w[n] + B_1 w[n-12]$$

$$x[n] = x[n-12] + B_0 w[n] + B_1 w[n-12] + a_1 (x[n-1] - x[n-13])$$

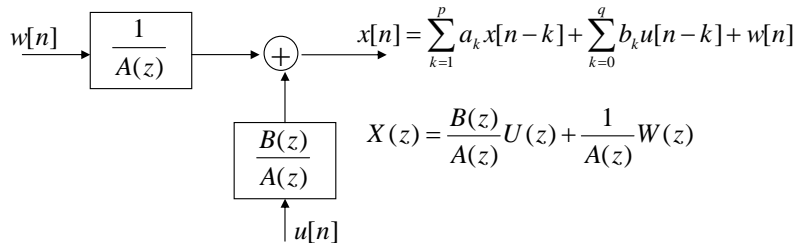


An example, let's study the model (1,0,0)x(0,1,1) with s=12. The seasonal ARIMA model yields $x_s[n]$ whose difference equation corresponds to the (0,1,1) s=12 model as shown in the slide. $x_s[n]$ is the input to the within year ARIMA model (1,0,0) whose difference equation is also shown in the slide. Eliminating $x_s[n]$ in both equations give us the final model.

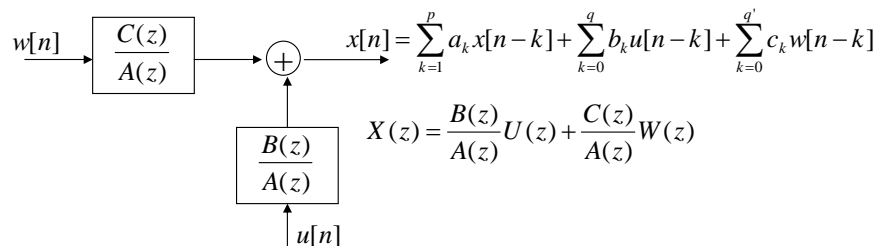
Note that this final model is something appealing when working with seasonal data, it states that the current sample is what happened 12 months ago, plus some random input ($B_0 w[n] + B_1 w[n-12]$) plus something that accounts for the differences between the behaviour one month ago and 13 months ago.

8. Known external inputs: System identification

ARX



ARMAX



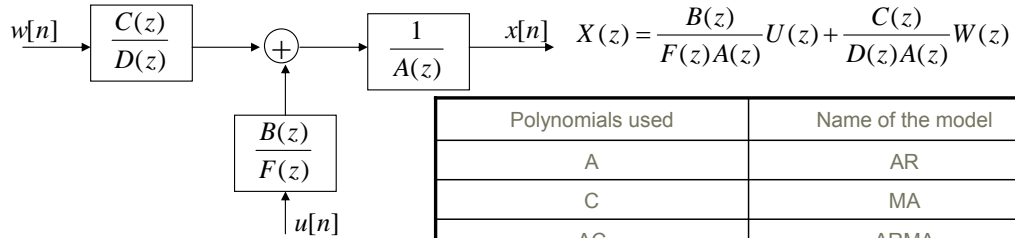
There are situations in which the time series directly depends on a known input plus some other random (unknown) effect. For instance, the price of natural gas in international markets is computed as a factor of the current price of oil plus some extra costs due to transporting, manufacturing, etc. A possible model for the price of natural gas might be a linear function of the oil price (known), plus a linear function of unknown costs. This problem is known as system identification.

Depending on the specific way in which these factors are combined, the model receives a name or another.

The student interested in knowing more about these models may read Ljung1987.

9. A family of models

General model

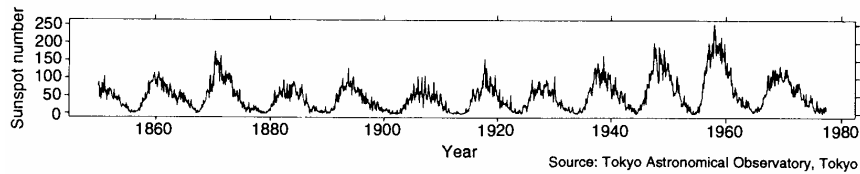


Polynomials used	Name of the model
A	AR
C	MA
AC	ARMA
ACD	ARIMA
AB	ARX
ABC	ARMAX
ABD	ARARX
ABCD	ARARMAX
BFCD	Box-Jenkins

Most of the models studied so far are part of a big family of linear models. The scheme shown in the slide represents the most general situation one can think of when using linear models. Depending on which polynomials we use, we have one or another model.

The student interested in knowing more about these models may read Ljung1987.

10. Nonlinear models



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

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

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

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

The plot in the slide shows the monthly average number of sunspots between 1850 and 1977. It can be seen that there is a cyclic component of 11 years, and that the rising slope is higher than the dropping slope. This kind of asymmetry cannot be modelled with a single sinusoidal seasonal component (if we try to do so, we should use a non linear model). Alternatively, this kind of asymmetries can be studied in Fourier space (harmonic analysis), but we will do this in the next chapter.

There are a number of nonlinear models. Here, we only show a few of them:

- NLAR: Non linear AR, where the function that relates the current sample to past samples is nonlinear
- Time varying parameter models: the combination coefficients change with time
- Random coefficient models: the combination coefficient is known up to a given degree of certainty
- Bilinear models: where the time series can also be explained in terms of the products of previous samples with the input white process.

10. Nonlinear models

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

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

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

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

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

(Neural networks)

(Chaos)



More nonlinear models:

-Threshold AR: The time series can be explained by two independent models depending on whether a previous sample is bigger than a given threshold or not

-Smooth Threshold AR: The time series can be continuously explained by two models. $S(x)$ acts as a smooth threshold. Usually $S(x) = 1 / (1 + \exp(-(r-x)/t))$.

There are some other models to handle, for instance, situations in which the variance is changing with time (heteroscedasticity). After removing all trends and seasonal components, the residual of the time series may have a time varying variance. If this is the case, the time series may be explained by an ARCH model (Autoregressive Conditional Heterocedastic)

The student interested in knowing more about nonlinear models may read Chatfield1996a.

11. Parameter estimation

Maximum Likelihood Estimates (MLE)

$$\begin{aligned}
 X_1 | \theta &\rightarrow N\left(0, \frac{\sigma_w^2}{1-a_1^2}\right) \\
 X_2 | X_1, \theta &\rightarrow N(a_1 x_1, \sigma_w^2) \\
 X_3 | X_2, X_1, \theta &\rightarrow N(a_1 x_2, \sigma_w^2) \\
 &\dots \\
 X_N | X_{N-1}, \dots, X_1, \theta &\rightarrow N(a_1 x_{N-1}, \sigma_w^2)
 \end{aligned}
 \left. \vphantom{\begin{aligned} X_1 | \theta \\ X_2 | X_1, \theta \\ X_3 | X_2, X_1, \theta \\ \dots \\ X_N | X_{N-1}, \dots, X_1, \theta \end{aligned}} \right\} \rightarrow f_{X_1, X_2, \dots, X_N | \theta}(x_1, x_2, \dots, x_N) = \\
 = f_{X_1 | \theta}(x_1) f_{X_2 | X_1, \theta}(x_2) f_{X_3 | X_2, \theta}(x_3) \dots f_{X_N | X_{N-1}, \theta}(x_N)$$

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

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

Confidence intervals



It is clearly seen, that the probability of observing X3 given X2 is independent from X1. Therefore, the joint probability of all the observed variables is the product of observing X1, times observing X2 given X1, times observing X3 given X2, etc. The logarithm of this joint probability is called the likelihood function.

Maximum likelihood looks for the model parameters that maximize the observation of the provided samples. This in turn boils down to solving a pair of nonlinear equations that have to be solved numerically using some iterative algorithm.

An important issue of the maximum likelihood approach is that we can compute confidence intervals for the model parameters. In system identification, these confidence intervals are usually translated into confidence regions for the system poles and zeros.

The reader interested in the estimation of the model parameters using the Maximum Likelihood approach may read Hamilton1994a.

11. Parameter estimation

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

Least Squares Estimates (LSE)

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

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

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

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

The least squares estimate minimizes the power of the input random signal. The rationale behind it is that the input variance will not be minimum if I could have explained $x[n]$ using a better a_1 . Minimizing the variance of the input signal with respect to the model coefficients yield a set of equations (called the Yule-Walker equations already seen in the explanation of the AR(p) model). Once all the model coefficients are estimated, we can estimate the input variance. The reader interested in the LSE method to estimate the model may read Proakis1988a.

Note that the solution of this approach need not be the same as the one provided by MLE, although it much simpler to compute in practice. In engineering, the solution of this equation system is performed using the Levinson-Durbin algorithm. The LSE has the drawback that its accuracy relies on the accuracy of the determination of the time series autocorrelation.

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

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

The selection of the model order is a tricky topic. One is tempted to use the model that better fits the data. However, there is a point at which increasing the model order will marginally increase the fitting. Thus, it is also interesting to keep the model order as low as possible as long as it fits reasonably well the data. To choose such a model we must use some criterion. One that is quite intuitive is to analyze the ACF/PACF structure (this will be further explained in the next slides). Another possibility is to use any of the parsimony criteria available. In the slide we show three of them (AIC, BIC, FPE). They are designed to decrease as long as we are significantly fitting the data better, and to increase when the improvements are marginal.

The fit of the data is provided by the input variance which can also be seen as the variance of the misfit error. N is the number of available samples in the time series and p,q are the ARMA model orders.

12. Order selection

Partial correlation coefficients (PACF, First-order correlations)

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

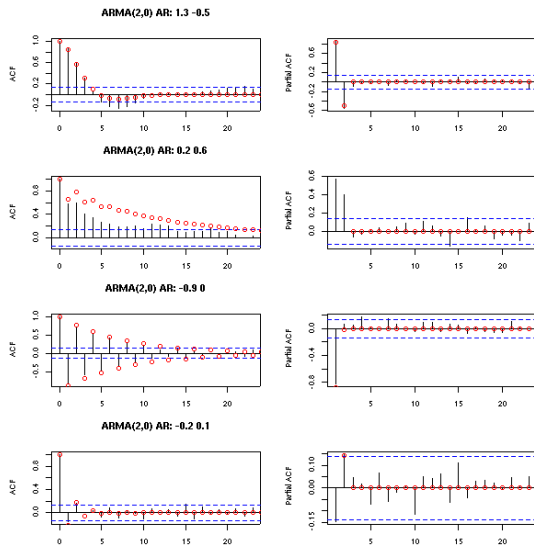
$$r[n_0] = \sum_{n=1}^{n_0} \phi_{n_0,n} r[n_0 - n]$$

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

The partial correlation coefficient is defined as the last coefficient of a partial autoregression model of order n_0 . It is important that $x[n]$ is zero mean.

The partial correlation between X_n and $X_{(n+n_0)}$ is the correlation between these two variables after removing all the linear relationships between the samples between X_n and $X_{(n+n_0)}$.

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(0,1): ACF: one peak; PACF: exponential decrease
 ARMA(0,2): ACF: two peaks; PACF: exponential decrease or waves
 ARMA(1,1): ACF&PACF: exponential decrease

For knowing more about the thumb rules to select the model order using ACF and PACF, please follow http://zoonek2.free.fr/UNIX/48_R/15.html

13. Model checking

Residual Analysis

Example: ARMA (1,1)

$$x[n] = ax[n-1] + b_0w[n] + b_1w[n-1] \Rightarrow \begin{cases} \hat{w}[n] = \frac{1}{b_0}(x[n] - ax[n-1] - b_1\hat{w}[n-1]) \\ \hat{w}[0] = 0 \end{cases}$$

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.

Model checking is as important as model formulation and estimation. After fitting our model we have to be sure that the model is valid and that none of the model hypothesis are violated. This is usually done through the residual analysis. This analysis aims at determining what is the one-sample-ahead prediction error which is also interpreted as the system input $w[n]$.

The main assumptions done in most models are: gaussianity, stationarity, residual independency. Be careful, that ARIMA models can deal with nonstationary signals because it explicitly considers multiple differentiation (d times) as a tool to produce a stationary time series.

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

Diagnostic checking is the general name used for all those different checkings that might point out that the model we have fitted is not adequate. If this is the case, you would have to revise the assumptions of your model, why they are not met and, eventually, change the model.

If stationarity is not met, it can be easily handled by removing a trend or seasonal component or by differencing enough times.

Many times residual gaussianity is not met, however, most of the times this is not a severe problem as long as its distribution has zero odd moments and the model still provide useful insight in the data structure.

A check that is particularly useful to test whether the model is overfitting or not is to keep the last M samples for testing prediction accuracy. Thus, if there are N samples in total in a time series, we use $N-M$ for training and the last M for testing. The residual analysis on these M samples should also pass all the diagnostic checkings.

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

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*. Academic Press, 1999.
- J. D. Hamilton. *Time series analysis*. Princeton Univ. Press, 1994.