Time Series Analysis

Lasse Engbo Christiansen

Department of Applied Mathematics and Computer Science
Technical University of Denmark

September 28, 2017

Outline of the lecture

- Stochastic processes, 1st part:
  - Stochastic processes in general: Sec 5.1, 5.2, 5.3 [except 5.3.2], 5.4.
  - MA, AR, and ARMA-processes, Sec. 5.5
  - Non-stationary models, Sec. 5.6
  - Optimal Prediction, Sec. 5.7

Stochastic Processes – in general

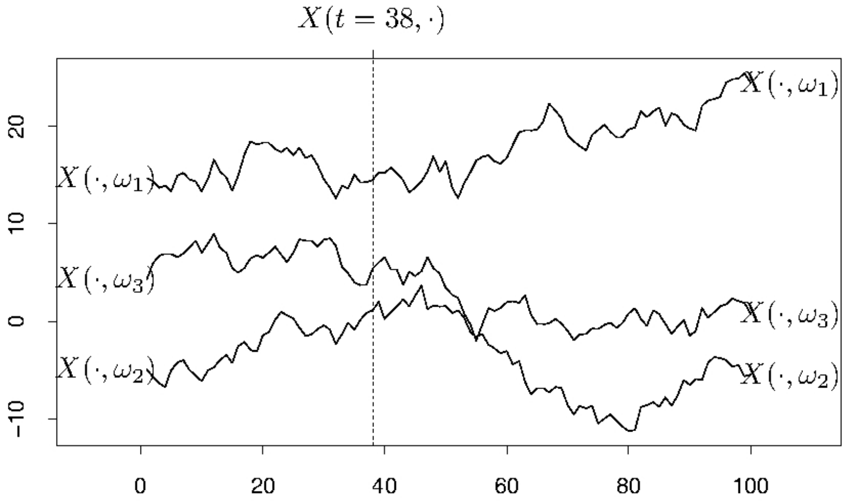
- Function: $X(t, \omega)$
- Time: $t \in T$
- Realization: $\omega \in \Omega$

- Index set: T
- Sample Space: Ω

- $X(t = t_0, \cdot)$ is a random variable
- $X(\cdot, \omega)$ is a time series (i.e. *one* realization). This is what we often denote $\{x_t\}$.

- In this course we consider the case where time is discrete, and the realizations take values on the real numbers (continuous range).

Stochastic Processes – illustration



Complete Characterization

n-dimensional probability density:

$$f_{X(t_1),\ldots,X(t_n)}(x_1,\ldots,x_n)$$

Family of probability density functions, i.e.:

- For all $n = 1, 2, 3, \ldots$
- and all t

is called the *family of finite-dimensional probability density functions (pdf's) for the process*. This family completely characterize the stochastic process.

2nd order moment representation

Mean function:

$$\mu(t) = E[X(t)] = \int_{-\infty}^{\infty} x\, f_{X(t)}(x)\, dx,$$

Autocovariance function:

$$\begin{aligned}\gamma_{XX}(t_1, t_2) &= \gamma(t_1, t_2) = \mathrm{Cov}\left[X(t_1), X(t_2)\right] \\ &= E\left[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))\right]\end{aligned}$$

The variance function is obtained from $\gamma(t_1, t_2)$ when $t_1 = t_2 = t$:

$$\sigma^2(t) = V[X(t)] = E\left[(X(t) - \mu(t))^2\right]$$

Stationarity

- A process $\{X(t)\}$ is said to be *strongly stationary* if all finite-dimensional distributions are invariant for changes in time, i.e. for every n, and for any set $(t_1, t_2, \ldots, t_n)$ and for any h it holds

$$f_{X(t_1),\cdots,X(t_n)}(x_1, \cdots, x_n) = f_{X(t_1+h),\cdots,X(t_n+h)}(x_1, \cdots, x_n)$$

- A process $\{X(t)\}$ is said to be *weakly stationary of order* k if all the first k moments are invariant to changes in time
- A weakly stationary process of order 2 is simply called *weakly stationary* or just *stationary*:

$$\mu(t) = \mu \quad \sigma^2(t) = \sigma^2 \quad \gamma(t_1, t_2) = \gamma(t_1 - t_2)$$

Ergodicity

- In time series analysis we normally assume that we have access to one realization only
- We therefore need to be able to determine characteristics of the random variable X_t from one realization x_t
- It is often enough to require the process to be mean-ergodic:

$$E[X(t)] = \int_{\Omega} x(t,\omega) f(\omega)\, d\omega = \lim_{T\to\infty} \frac{1}{2T} \int_{-T}^{T} x(t,\omega)\, dt$$

i.e. if the *mean of the ensemble* equals the *mean over time*

Some intuitive examples, not directly related to time series:
`http://news.softpedia.com/news/What-is-ergodicity-15686.shtml`

Special processes

- *Normal processes* (also called *Gaussian processes*): All finite-dimensional distribution functions are (multivariate) normal distributions
- *Markov processes*: The conditional distribution depends only on the latest state of the process:

$$\mathrm{P}\{X(t_n) \leq x | X(t_{n-1}), \cdots, X(t_1)\} = \mathrm{P}\{X(t_n) \leq x | X(t_{n-1})\}$$

- *Deterministic processes*: Can be predicted without uncertainty from past observations
- *Pure stochastic processes*: Can be written as a (infinite) linear combination of uncorrelated random variables
- Decomposition: $X_t = S_t + D_t$

Autocovariance and autocorrelation

- For stationary processes: Only dependent on the time difference $\tau = t_2 - t_1$
- Autocovariance:

$$\gamma(\tau) = \gamma_{XX}(\tau) = \mathrm{Cov}[X(t), X(t+\tau)] = E[X(t)X(t+\tau)]$$

- Autocorrelation:

$$\rho(\tau) = \rho_{XX}(\tau) = \gamma_{XX}(\tau)/\gamma_{XX}(0) = \gamma_{XX}(\tau)/\sigma_X^2$$

- Some properties of the autocovariance function:
  - $\gamma(\tau) = \gamma(-\tau)$
  - $|\gamma(\tau)| \leq \gamma(0)$

Linear processes

- A linear process $\{Y_t\}$ is a process that can be written on the form

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

  where μ is the mean value of the process and
- $\{\varepsilon_t\}$ is white noise, i.e. a sequence of uncorrelated, identically distributed random variables.
- $\{\varepsilon_t\}$ can be scaled so that $\psi_0 = 1$
- Without loss of generality we assume $\mu = 0$

ψ- and π-weights

- Transfer function and linear process:

$$\psi(B) = 1 + \sum_{i=1}^{\infty} \psi_i B^i \qquad Y_t = \psi(B)\varepsilon_t$$

- Inverse operator (if it exists) and the linear process:

$$\pi(B) = 1 + \sum_{i=1}^{\infty} \pi_i B^i \qquad \pi(B)\, Y_t = \varepsilon_t,$$

- Autocovariance using ψ-weights:

$$\gamma(k) = \mathrm{Cov}\left[\sum_{i=0}^{\infty} \psi_i \varepsilon_{t-i}, \sum_{i=0}^{\infty} \psi_i \varepsilon_{t+k-i}\right] = \sigma_\varepsilon^2 \sum_{i=0}^{\infty} \psi_i \psi_{i+k}$$

Stationarity and invertibility

- The linear process $Y_t = \psi(B)\varepsilon_t$ is *stationary* if

$$\psi(z) = \sum_{i=0}^{\infty} \psi_i z^{-i}$$

  converges for $|z| \geq 1$ (i.e. old values of ε_t are down-weighted)
- The linear process $\pi(B)\,Y_t = \varepsilon_t$ is said to be *invertible* if

$$\pi(z) = \sum_{i=0}^{\infty} \pi_i z^{-i}$$

  converges for $|z| \geq 1$ (i.e. ε_t can be calculated from recent values of Y_t)

Linear process as a statistical model?

$$Y_t = \varepsilon_t + \psi_1 \varepsilon_{t-1} + \psi_2 \varepsilon_{t-2} + + \psi_3 \varepsilon_{t-3} + \dots$$

- Observations: Y_1, Y_2, $Y_3, \dots,$ Y_N
- Task: Find an infinite number of parameters from N observations!
- Solution: Restrict the sequence $1, \psi_1, \psi_2, \psi_3, \dots$

$MA(q)$, $AR(p)$, and $ARMA(p, q)$ processes

$$\begin{aligned} Y_t &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \\ Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} &= \varepsilon_t \\ Y_t + \phi_1 Y_{t-1} + \cdots + \phi_p Y_{t-p} &= \varepsilon_t + \theta_1 \varepsilon_{t-1} + \cdots + \theta_q \varepsilon_{t-q} \end{aligned}$$

$\{\varepsilon_t\}$ is white noise

$$\begin{aligned} Y_t &= \theta(B)\varepsilon_t \\ \phi(B)\, Y_t &= \varepsilon_t \\ \phi(B)\, Y_t &= \theta(B)\varepsilon_t \end{aligned}$$

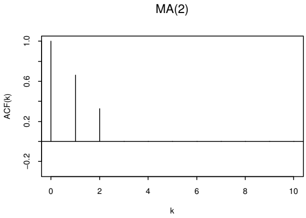
$\phi(B)$ and $\theta(B)$ are polynomials in the backward shift operator B, ($BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$)

Stationarity and invertibility

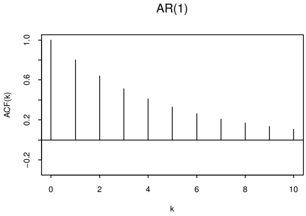
- $MA(q)$
  - Always stationary
  - Invertible if the roots in $\theta(z^{-1}) = 0$ with respect to z all are within the unit circle
- $AR(p)$
  - Always invertible
  - Stationary if the roots of $\phi(z^{-1}) = 0$ with respect to z all lie within the unit circle
- $ARMA(p, q)$
  - Stationary if the roots of $\phi(z^{-1}) = 0$ with respect to z all lie within the unit circle
  - Invertible if the roots in $\theta(z^{-1}) = 0$ with respect to z all are within the unit circle

Autocorrelations

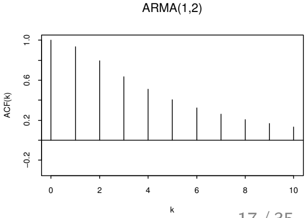
MA(2):
$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$
zero after lag 2



AR(1):
$(1 - 0.8B)\,Y_t = \varepsilon_t$
exponential decay (damped sine in case of complex roots)

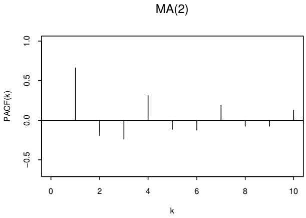


ARMA(1,2):
$(1 - 0.8B)\,Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$
exponential decay from lag $q+1-p = 2+1-1 = 2$
(damped sine in case of complex roots)

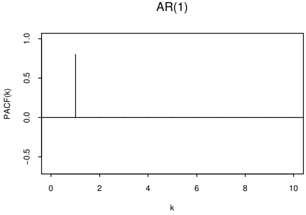


Partial autocorrelations (Appendix B)

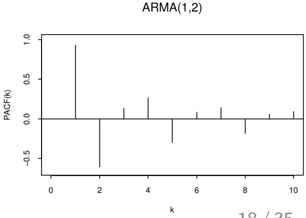
MA(2):
$Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$



AR(1):
$(1 - 0.8B)\, Y_t = \varepsilon_t$
zero after lag 1



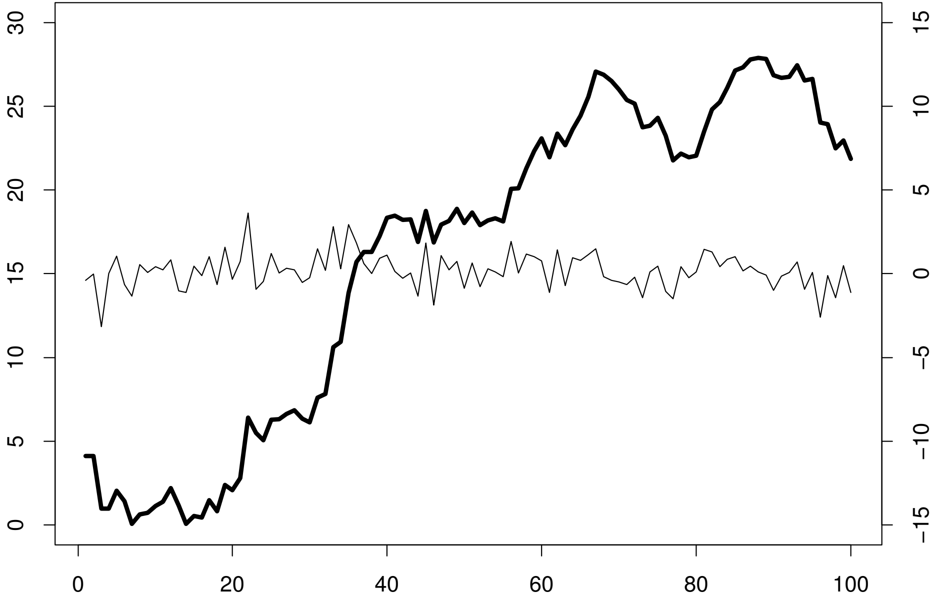
ARMA(1,2):
$(1 - 0.8B)\, Y_t = (1 + 0.9B + 0.8B^2)\varepsilon_t$



Inverse autocorrelation

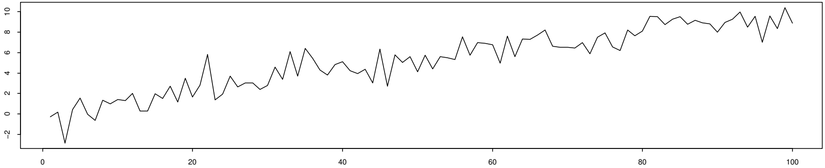
- The process: $\phi(B)Y_t = \theta(B)\varepsilon_t$
- The dual process: $\theta(B)Z_t = \phi(B)\varepsilon_t$
- The inverse autocorrelation is the autocorrelation for the dual process
- Thus, the IACF can be used in a similar way as the PACF.

Non-stationary series

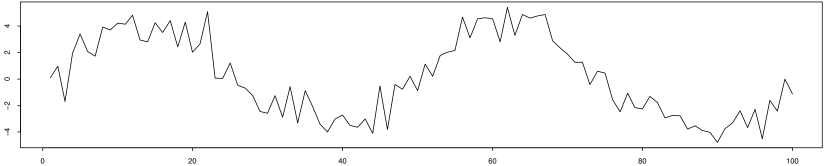


Some types of non-stationarity

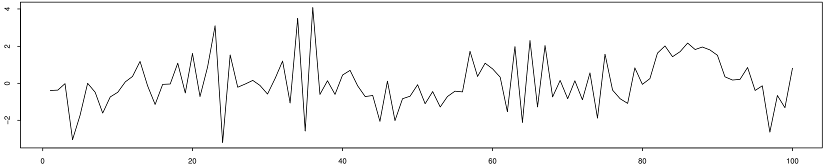
Long term trends



Periodic trends



General time varying behavior



The $ARIMA(p, d, q)$-process

- An $ARMA(p, q)$ model for:

$$W_t = \nabla^d Y_t = (1 - B)^d Y_t$$

where $\{Y_t\}$ is the series

- That is:

$$\phi(B)\nabla^d Y_t = \theta(B)\varepsilon_t$$

- If we consider stationarity:

$$\phi(z^{-1})(1 - z^{-1})^d = 0$$

i.e. d roots in $z = 1 + 0i$, and the rest inside the unit circle

The $(p, d, q) \times (P, D, Q)_s$ seasonal process

- A multiplicative (stationary) $ARMA(p, q)$ model for:

$$W_t = \nabla^d \nabla_s^D Y_t = (1 - B)^d (1 - B^s)^D Y_t$$

  where $\{Y_t\}$ is the series

- That is:

$$\phi(B)\Phi(B^s)\nabla^d \nabla_s^D Y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

- If we consider stationarity:

$$\phi(z^{-1})\Phi(z^{-s})(1 - z^{-1})^d (1 - z^{-s})^D = 0$$

  i.e. d roots in $z = 1 + 0i$, $D \times s$ roots on the unit circle, and the rest inside the unit circle

The case $d = D = 0$; stationary seasonal process

- General:

$$\phi(B)\Phi(B^s)\,Y_t = \theta(B)\Theta(B^s)\varepsilon_t$$

- Example:

$$(1 - \Phi B^{12})\,Y_t = \varepsilon_t$$

- Which can also be written:

$$Y_t = \Phi\,Y_{t-12} + \varepsilon_t$$

  i.e. Y_t depend on $Y_{t-12},\ Y_{t-24},\ \ldots$ (thereof the name)
- How would you think that the auto correlation function looks?
- Take a look at Example 5.10 also.

Prediction

- At time t we have observations Y_t, Y_{t-1}, Y_{t-2}, $Y_{t-3}, \ldots$
- We want a prediction of Y_{t+k}, where $k \geq 1$
- If we want to minimize the expected squared error the optimal prediction is the conditional expectation:

$$\widehat{Y}_{t+k|t} = E[\,Y_{t+k}\,|\,Y_t,\, Y_{t-1},\, Y_{t-2}, \cdots]$$

Example – prediction in the $AR(1)$ model

- We write the model like $Y_{t+1} = \phi Y_t + \varepsilon_{t+1}$ (note the sign on ϕ)
- 1-step prediction:

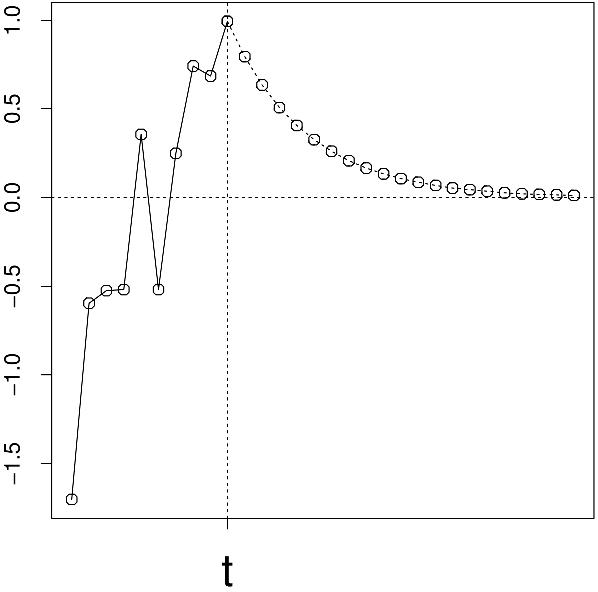
$$\begin{aligned}\widehat{Y}_{t+1|t} &= E[Y_{t+1} | Y_t, Y_{t-1}, \cdots] = E[\phi Y_t + \varepsilon_{t+1} | Y_t, Y_{t-1}, \cdots] \\ &= \phi Y_t + 0 = \phi Y_t\end{aligned}$$

- 2-step prediction:

$$\begin{aligned}\widehat{Y}_{t+2|t} &= E[Y_{t+2} | Y_t, Y_{t-1}, \cdots] = E[\phi Y_{t+1} + \varepsilon_{t+2} | Y_t, Y_{t-1}, \cdots] \\ &= \phi \widehat{Y}_{t+1|t} + 0 = \phi^2 Y_t\end{aligned}$$

- k-step prediction: $\boxed{\widehat{Y}_{t+k|t} = \phi^k Y_t}$

Example – prediction in $Y_t = 0.8\,Y_{t-1} + \varepsilon_t$



Variance of prediction error for the $AR(1)$-process

Prediction error:

$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = Y_{t+k} - \phi^k Y_t$$

$$\begin{aligned}
Y_{t+k} &= \phi Y_{t+k-1} + \varepsilon_{t+k} \\
&= \phi(\phi Y_{t+k-2} + \varepsilon_{t+k-1}) + \varepsilon_{t+k} \\
&= \phi^2 Y_{t+k-2} + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k} \\
&= \phi^2(\phi Y_{t+k-3} + \varepsilon_{t+k-2}) + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k} \\
&= \phi^3 Y_{t+k-3} + \phi^2\varepsilon_{t+k-2} + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k} \\
&\vdots \\
&= \phi^k Y_t + \phi^{k-1}\varepsilon_{t+1} + \phi^{k-2}\varepsilon_{t+2} + \ldots + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k}
\end{aligned}$$

Variance of prediction error for the $AR(1)$-process

Variance of prediction error:

$$\begin{aligned} V[e_{t+k|t}] &= V[\phi^{k-1}\varepsilon_{t+1} + \phi^{k-2}\varepsilon_{t+2} + \ldots + \phi\varepsilon_{t+k-1} + \varepsilon_{t+k}] \\ &= (\phi^{2(k-1)} + \phi^{2(k-2)} + \ldots + \phi^2 + 1)\sigma_\varepsilon^2 \end{aligned}$$

$(1-\alpha) \times 100\%$ prediction interval:

$$\widehat{Y}_{t+k|t} \pm u_{\alpha/2}\sqrt{V[e_{t+k|t}]}$$

$u_{\alpha/2}$ is the $\alpha/2$-quantile in the standard normal distribution

k-step prediction in $ARMA(p, q)$–models

We assume that $k > max(p, q)$. The process:

$$Y_{t+k} + \phi_1 Y_{t+k-1} + \cdots + \phi_p Y_{t+k-p} = \\ \varepsilon_{t+k} + \theta_1 \varepsilon_{t+k-1} + \cdots + \theta_q \varepsilon_{t+k-q}$$

Using conditional expectation on both sides we get:

$$\begin{aligned} \widehat{Y}_{t+k|t} &= -\phi_1 \widehat{Y}_{t+k-1|t} - \cdots - \phi_p \widehat{Y}_{t+k-p|t} \\ &\quad + \widehat{\varepsilon}_{t+k|t} + \theta_1 \widehat{\varepsilon}_{t+k-1|t} + \cdots + \theta_q \widehat{\varepsilon}_{t+k-q|t} \end{aligned}$$

This results in a recursive method for calculating the predictions – how would you find $\widehat{\varepsilon}_{t+k-q|t}$?

Inverse form

For an invertible process, the π-weights

$$\varepsilon_t = Y_t + \pi_1 Y_{t-1} + \pi_2 Y_{t-2} + \cdots$$

goes to zero sufficiently fast and only recent values of the process is needed.

Variance of prediction error

Process written with ψ-weights:

$$Y_{t+k} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \cdots + \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \cdots$$

k-step prediction:

$$\widehat{Y}_{t+k|t} = \psi_k \varepsilon_t + \psi_{k+1} \varepsilon_{t-1} + \cdots$$

k-step prediction error:

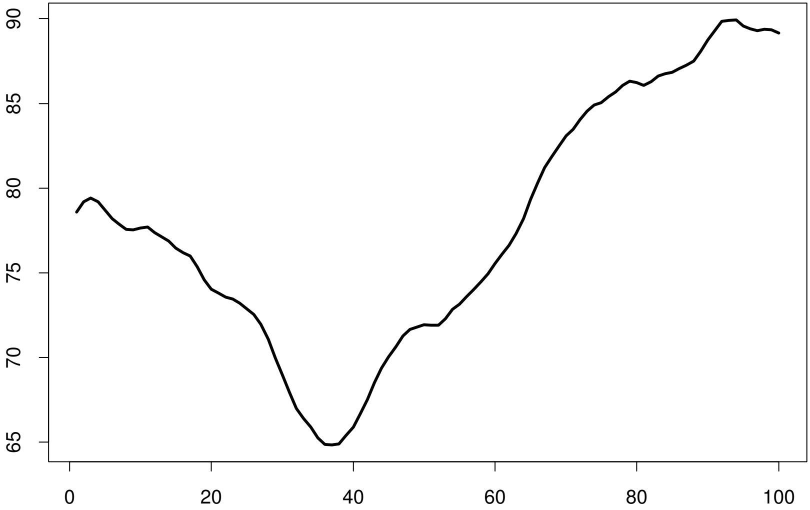
$$e_{t+k|t} = Y_{t+k} - \widehat{Y}_{t+k|t} = \varepsilon_{t+k} + \psi_1 \varepsilon_{t+k-1} + \cdots + \psi_{k-1} \varepsilon_{t+1}$$

Variance of k-step prediction error:

$$\sigma_k^2 = (1 + \psi_1^2 + \cdots + \psi_{k-1}^2) \sigma_\varepsilon^2$$

Prediction of bond prices

$$(1 - 1.274B + 0.3867B^2)\nabla Y_t = \varepsilon_t; \quad \sigma_\varepsilon^2 = 0.201^2; \quad \mu_Y = 84.3$$



Prediction of bond prices

- Price last 6 days: $\ldots, 90.79, 89.90, 88.88, 87.98, 87.41, 87.16$
- Prediction of price in two days:

$$
\begin{aligned}
\varphi(B) &= \phi(B)\nabla = (1+\phi_1 B+\phi_2 B^2)(1-B) \\
&= 1+(\phi_1-1)B+(\phi_2-\phi_1)B^2-\phi_2 B^3 \\
\widehat{Y}_{t+1|t} &= (1-\phi_1)\,Y_t+(\phi_1-\phi_2)\,Y_{t-1}+\phi_2\,Y_{t-2} = 87.06 \\
\widehat{Y}_{t+2|t} &= (1-\phi_1)\,\widehat{Y}_{t+1|t}+(\phi_1-\phi_2)\,Y_t+\phi_2\,Y_{t-1} = \underline{\underline{87.03}}
\end{aligned}
$$

- Variance of prediction error:

$$V[\varepsilon_{t+2}+(1-\phi_1)\varepsilon_{t+1}] = (1+(1-\phi_1)^2)\sigma_\varepsilon^2 = 0.499^2$$

- 95% prediction interval: $87.03 \pm 1.96 \cdot 0.50 = [86.05;\ 88.01]$

Highlights

- Stochastic process $X(t, \omega)$
  - $X(t = t_0, \cdot)$ is a random variable
  - $X(\cdot, \omega)$ is a time series (i.e. *one* realization).
- Stationarity
- Autocovariance:

$$\begin{aligned}\gamma_{XX}(t_1, t_2) &= \gamma(t_1, t_2) = \mathrm{Cov}\left[X(t_1), X(t_2)\right] \\ &= E\left[(X(t_1) - \mu(t_1))(X(t_2) - \mu(t_2))\right]\end{aligned}$$

- $MA(q)$, $AR(p)$, and $ARMA(p, q)$ processes.
- $ARIMA(p, d, q) \times (P, D, Q)_s$ for seasonal and non stationary processes.
- The optimal prediction is the conditional expectation:

$$\widehat{Y}_{t+k|t} = E[\,Y_{t+k} \,|\, Y_t,\, Y_{t-1},\, Y_{t-2}, \cdots]$$