Time Series Analysis

Week 6 - ACF and PACF with a focus on model order selection

Peder Bacher
Department of Applied Mathematics and Computer Science
Technical University of Denmark

March 8, 2024

Week 6: Outline of the lecture

- Estimation of auto-covariance and -correlation, Sec. 6.2.1 (and the intro. to 6.2)
- Using the SACF and SPACF for model order selection Sec. 6.5
- Model validation, Sec. 6.6

Autocorrelation and Partial Autocorrelation

Autocorrelation

Partial autocorrelation

Autocorrelation and Partial Autocorrelation

Autocorrelation

$$\rho(k) = \mathsf{Cor}[\,Y_t\,,\,Y_{t+k}]$$

Partial autocorrelation

Autocorrelation and Partial Autocorrelation

Autocorrelation

$$\rho(k) = \mathsf{Cor}[\, Y_t ,\, Y_{t+k}]$$

Partial autocorrelation

$$\phi_{kk} = \mathsf{Cor}[\, Y_t ,\, Y_{t+k} \,|\, Y_{t+1}, \ldots, Y_{t+k-1}]$$

Autocorrelation and Partial Autocorrelation

Autocorrelation

$$\rho(k) = \text{Cor}[Y_t, Y_{t+k}]$$

- Sample autocorrelation function (SACF): $\widehat{\rho}(k) = r_k = C(k)/C(0)$
- For white noise and $k \neq 0$ it holds that $E[\widehat{\rho}(k)] \simeq 0$ and $V[\widehat{\rho}(k)] \simeq 1/N$, this gives the bounds $\pm 2/\sqrt{N}$ for deciding when it is not possible to distinguish a value from zero.
- R: `acf(x)`

Partial autocorrelation

$$\phi_{kk} = \text{Cor}[Y_t, Y_{t+k} | Y_{t+1}, \ldots, Y_{t+k-1}]$$

Autocorrelation and Partial Autocorrelation

Autocorrelation

$$\rho(k) = \text{Cor}[Y_t, Y_{t+k}]$$

- Sample autocorrelation function (SACF): $\widehat{\rho}(k) = r_k = C(k)/C(0)$
- For white noise and $k \neq 0$ it holds that $E[\widehat{\rho}(k)] \simeq 0$ and $V[\widehat{\rho}(k)] \simeq 1/N$, this gives the bounds $\pm 2/\sqrt{N}$ for deciding when it is not possible to distinguish a value from zero.
- R: `acf(x)`

Partial autocorrelation

$$\phi_{kk} = \text{Cor}[Y_t, Y_{t+k} | Y_{t+1}, \ldots, Y_{t+k-1}]$$

- Sample partial autocorrelation function (SPACF): Use the Yule-Walker equations on $\widehat{\rho}(k)$ (exactly as for the theoretical relations Eq.(5.81)) or as in next slide
- It turns out that $\pm 2/\sqrt{N}$ is also appropriate for deciding when the SPACF is zero
- R: `acf(x, type="partial")` or `pacf(x)`

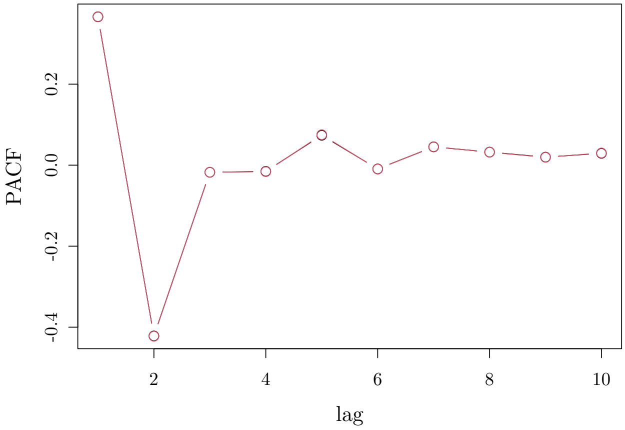
Autocorrelation and Partial Autocorrelation

```r
# Example to show how the PACF is calculated
set.seed(972)
n <- 1000
x <- arima.sim(list(ar=c(0.5,-0.4)), n=n)
#acf(x)
#pacf(x)
library(onlineforecast)
D <- lagdf(c(x), 0:50)

# A way to calculate the PACF
lag.max <- 10
pacf1 <- numeric(lag.max)
# First, calculate it with the function
val <- pacf(x, lag.max, plot=FALSE)
# Then calc on our own
for(k in 1:lag.max){
  (frml <- pst("k0 ~ 1 + ",pst("k",1:k, collapse=" + ")))
  fit <- lm(frml, D)
  pacf1[k] <- fit$coef[pst("k",k)]
}
```

Autocorrelation and Partial Autocorrelation

```r
# It's very close!
pacf1 - val$acf
plot(val$acf, type="b", xlab="lag", ylab="PACF")
lines(pacf1, type="b", col=2)
```



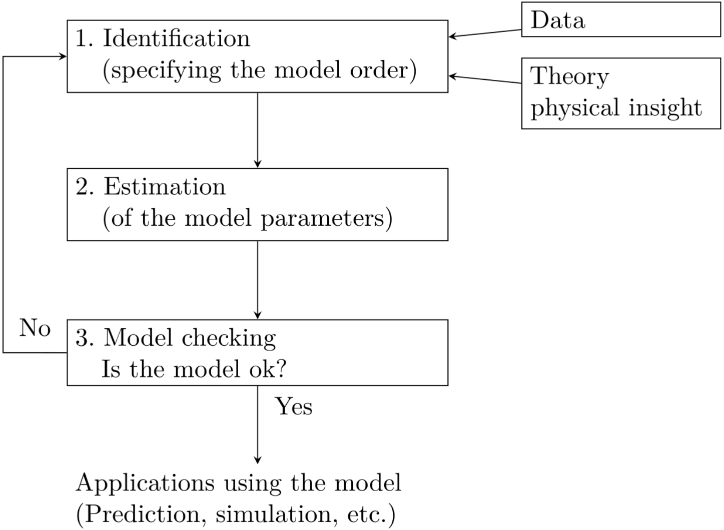
ARIMA models

- Today we will see how to identify ARIMA model orders

ARIMA models

- Today we will see how to identify ARIMA model orders
- Basically ARIMA(p, d, q) has
  - **A**uto-**R**egression order p
  - **M**oving-**A**verage order q
  - **I**ntegration order d

Model building in general

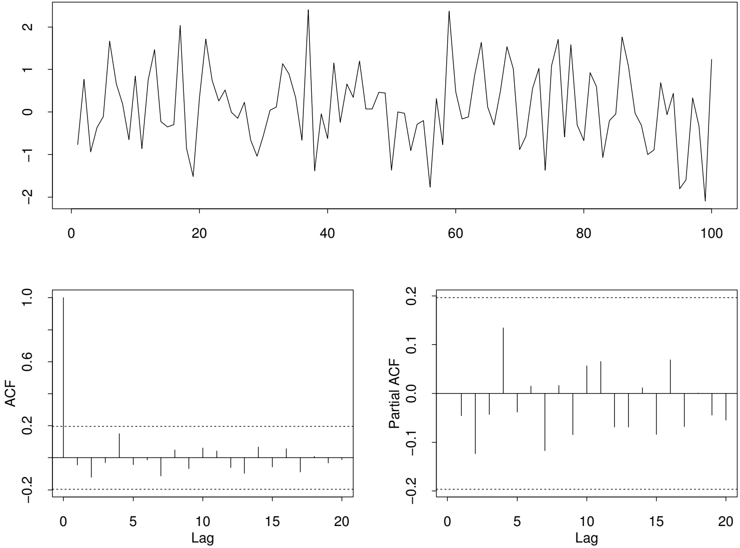


The golden table for ARMA identification

(Table 6.1)

	ACF $\rho(k)$	PACF ϕ_{kk}
AR(p)	Damped exponential and/or sine functions	$\phi_{kk} = 0$ for $k > p$
MA(q)	$\rho(k) = 0$ for $k > q$	Dominated by damped exponential and or/sine functions
ARMA(p, q)	Damped exponential and/or sine functions after lag max$(0, q - p)$	Dominated by damped exponential and/or sine functions after lag max$(0, p - q)$

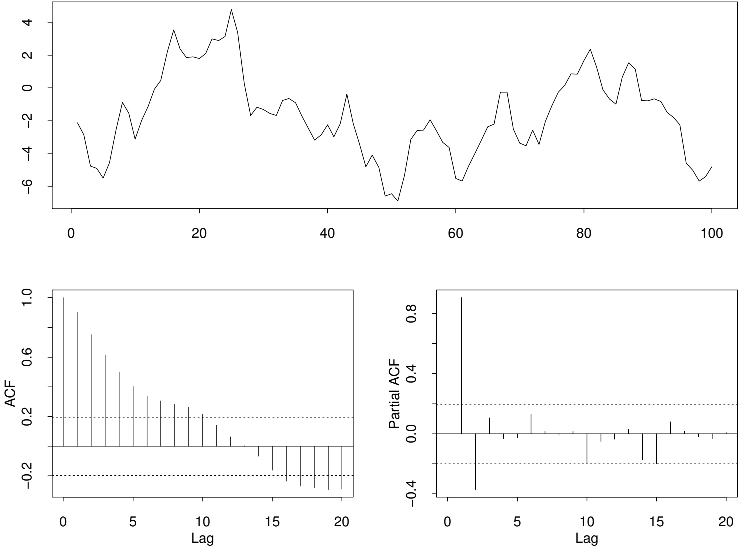
What would be an appropriate structure?



1: White noise 2: AR(1) 3: AR(2)

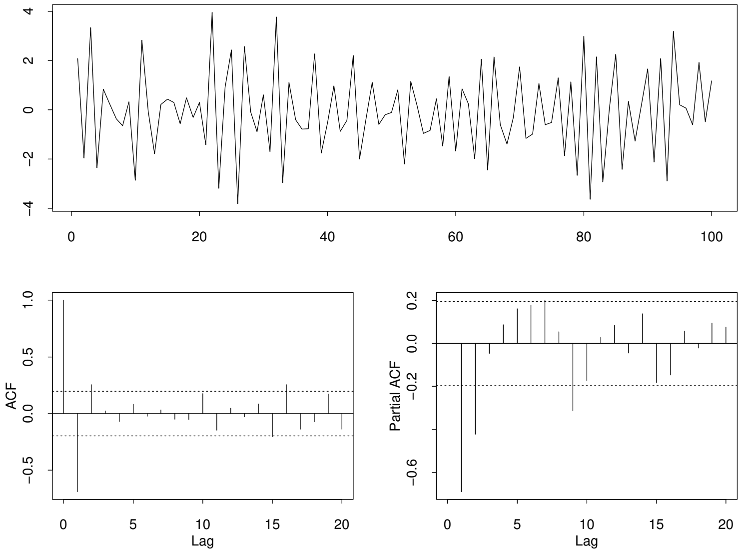
4: MA(1) 5: MA(2) 6: ARMA(1,1)

What would be an appropriate structure?



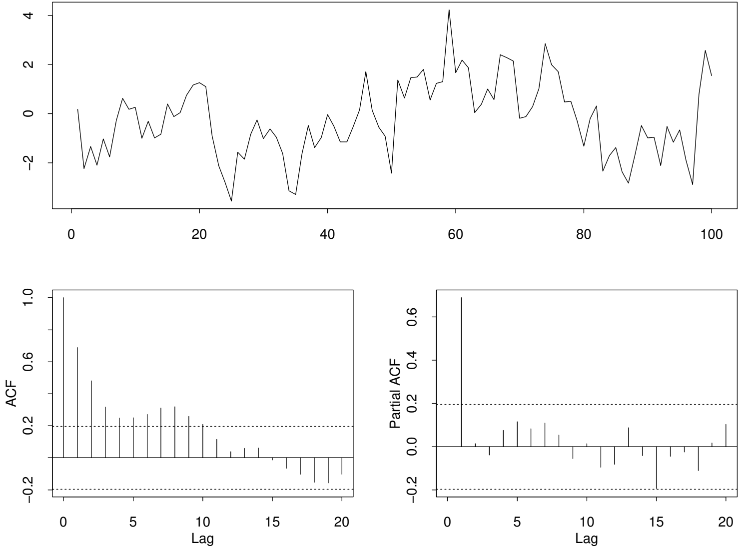
1: White noise 2: AR(1) 3: AR(2)
4: MA(1) 5: MA(2) 6: ARMA(1,1)

What would be an appropriate structure?



1: White noise	2: AR(1)	3: AR(2)
4: MA(1)	5: MA(2)	6: ARMA(1,1)

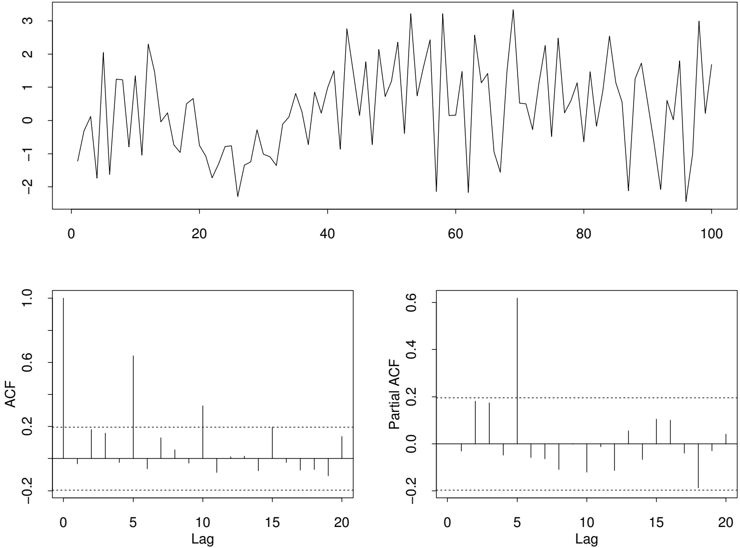
What would be an appropriate structure?



1: White noise 2: AR(1) 3: AR(2)

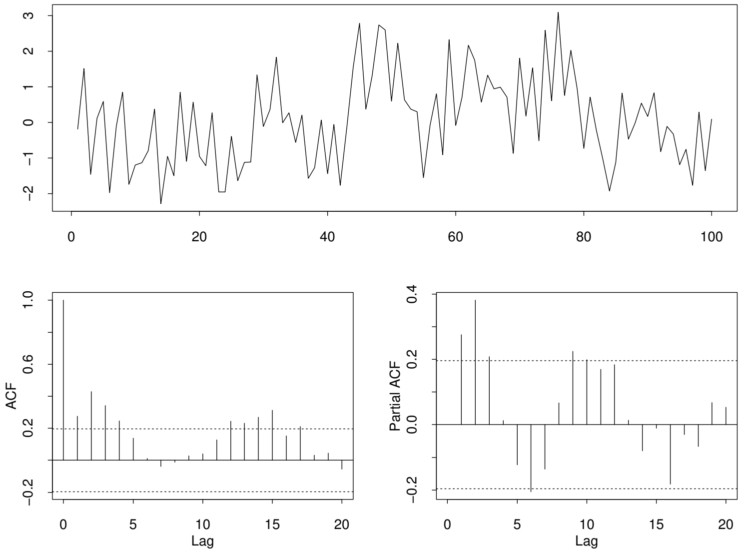
4: MA(1) 5: AR(8) 6: ARMA(1,1)

What would be an appropriate structure?



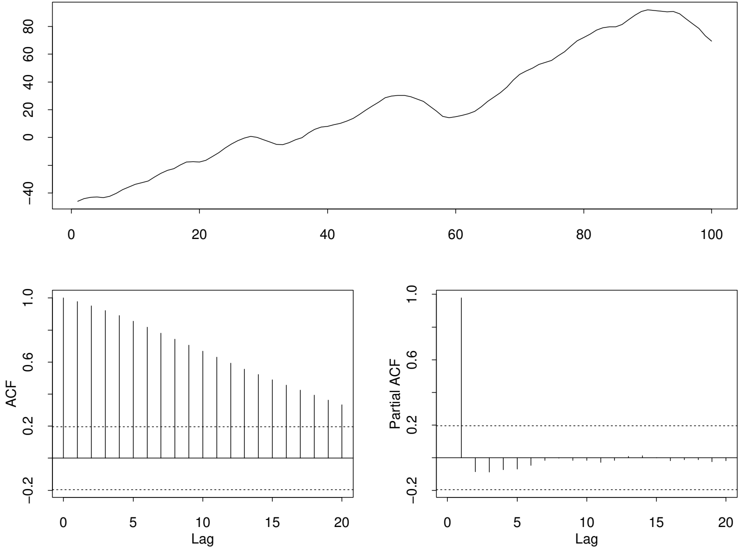
1: AR(1)	2: AR(5)	3: AR(0)x(2)$_5$
4: AR(0)x(1)$_4$	5: AR(0)x(1)$_5$	6: AR(10)

What would be an appropriate structure?



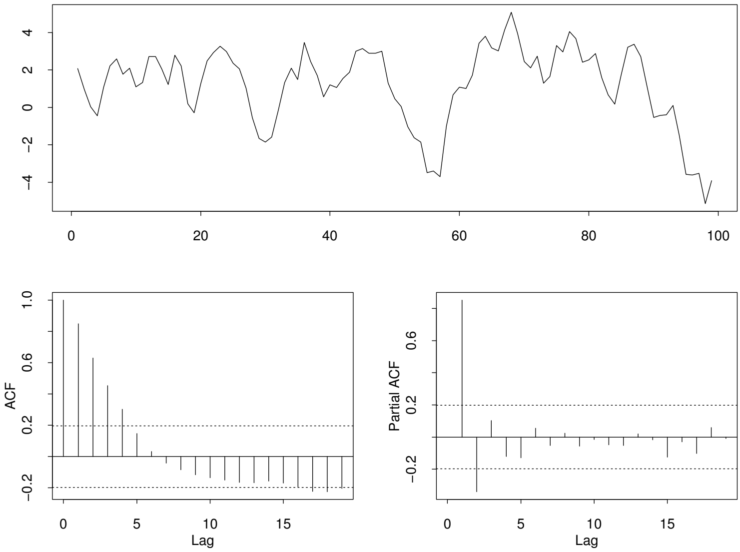
1: AR(4) 2: AR(5) 3: MA(4)
4: ARMA(1,1) 5: AR(2) 6: ARMA(2,2)

What would be an appropriate structure?



1: AR(3)	2: AR(2)	3: ARIMA(0,1,0)
4: ARMA(1,1)	5: ARIMA(2,1,0)	6: ARIMA(0,2,2)

Same series; analysing $\nabla Y_t = (1 - B) Y_t = Y_t - Y_{t-1}$



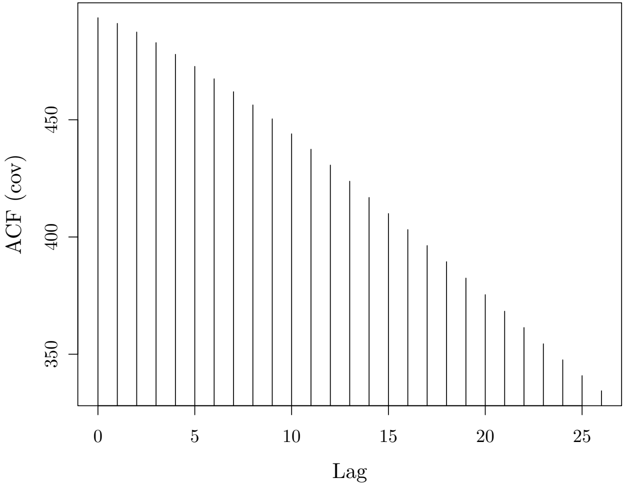
1: AR(3)	2: AR(2)	3: ARIMA(0,1,0)
4: ARMA(1,1)	5: ARIMA(2,1,0)	6: ARIMA(0,2,2)

How does $C(k)$ behave for non-stationary series?

$$C(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} (Y_t - \overline{Y})(Y_{t+|k|} - \overline{Y})$$

How does $C(k)$ behave for non-stationary series?

$$C(k) = \frac{1}{N} \sum_{t=1}^{N-|k|} (Y_t - \overline{Y})(Y_{t+|k|} - \overline{Y})$$



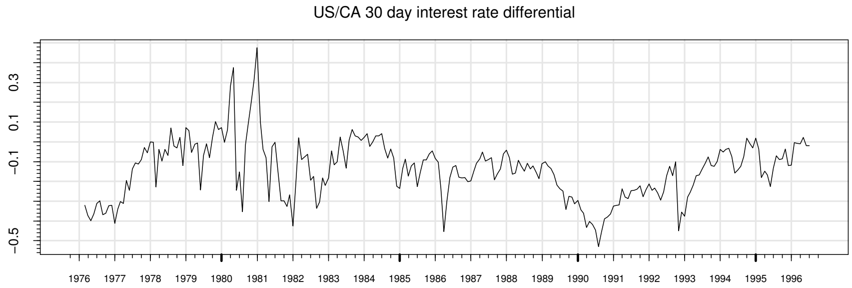
Identification of the order of differencing

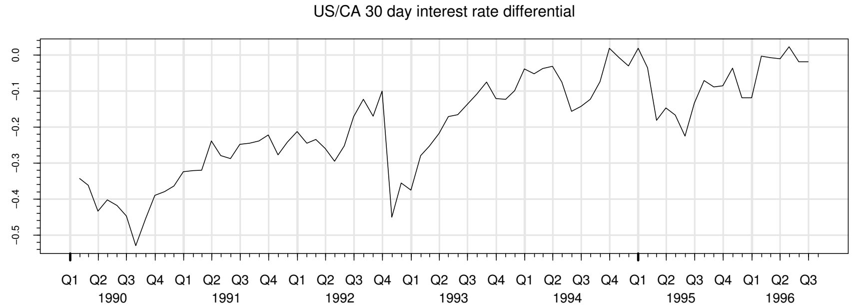
- Select the order of differencing d as the first order for which the autocorrelation decreases sufficiently fast towards 0
- In practice d is 0, 1, or maybe 2
- Sometimes a periodic difference is required, e.g. $Y_t - Y_{t-12}$

Identification of the order of differencing

- Select the order of differencing d as the first order for which the autocorrelation decreases sufficiently fast towards 0
- In practice d is 0, 1, or maybe 2
- Sometimes a periodic difference is required, e.g. $Y_t - Y_{t-12}$
- Remember to consider the practical application. E.g. it may be that the system is stationary, but you measured over a too short period

Stationarity vs. length of measuring period





Selection of the Model Order

- The model order of an ARMA process model:
  The number of parameters for the AR and MA part; (p, q).
- The autocorrelation functions can be used - as we just did

Selection of the Model Order

- The model order of an ARMA process model:
  The number of parameters for the AR and MA part; (p, q).
- The autocorrelation functions can be used - as we just did
- If that method fails to identify (p, q) because the process:
  - Is not a standard AR-proces.
  - Is not a standard MA-proces.
  - Is not a directly identifiable ARMA proces

Selection of the Model Order

- The model order of an ARMA process model:
  The number of parameters for the AR and MA part; (p, q).
- The autocorrelation functions can be used - as we just did
- If that method fails to identify (p, q) because the process:
  - Is not a standard AR-proces.
  - Is not a standard MA-proces.
  - Is not a directly identifiable ARMA proces
- then try a small model and analyse the residuals

Selection of the Model Order

- The model order of an ARMA process model:
  The number of parameters for the AR and MA part; (p, q).
- The autocorrelation functions can be used - as we just did
- If that method fails to identify (p, q) because the process:
  - Is not a standard AR-proces.
  - Is not a standard MA-proces.
  - Is not a directly identifiable ARMA proces
- then try a small model and analyse the residuals
- and/or Consider transformations
  Typically sqrt, log, square or inverse.

Iterative model building

Iterative model building

1. (Identification step): Construct a model for your data:

$$\phi(B)\,Y_t = \theta(B)\varepsilon_t$$

2. (Estimation step): Estimate the coefficients $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and calculate the model residuals $\hat{\varepsilon}_{t|t-1}$

Iterative model building

1. (Identification step): Construct a model for your data:

$$\phi(B)\,Y_t = \theta(B)\varepsilon_t$$

2. (Estimation step): Estimate the coefficients $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and calculate the model residuals $\hat{\varepsilon}_{t|t-1}$
3. (Model checking step):
   - Are the estimated coefficients significant?

Iterative model building

1. (Identification step): Construct a model for your data:

$$\phi(B)\,Y_t = \theta(B)\varepsilon_t$$

2. (Estimation step): Estimate the coefficients $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and calculate the model residuals $\hat{\varepsilon}_{t|t-1}$
3. (Model checking step):
   - Are the estimated coefficients significant?
   - Does $\hat{\varepsilon}_{t|t-1}$ resemble white noise?

Iterative model building

1. (Identification step): Construct a model for your data:

$$\phi(B)\,Y_t = \theta(B)\varepsilon_t$$

2. (Estimation step): Estimate the coefficients $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and calculate the model residuals $\hat{\varepsilon}_{t|t-1}$
3. (Model checking step):
    - Are the estimated coefficients significant?
    - Does $\hat{\varepsilon}_{t|t-1}$ resemble white noise?
    - If so, the model can be described by the ϕ and θ polynomials.

Iterative model building

1. (Identification step): Construct a model for your data:

$$\phi(B)\,Y_t = \theta(B)\varepsilon_t$$

2. (Estimation step): Estimate the coefficients $(\phi_1, \ldots, \phi_p, \theta_1, \ldots, \theta_q)$ and calculate the model residuals $\hat{\varepsilon}_{t|t-1}$
3. (Model checking step):
    - Are the estimated coefficients significant?
    - Does $\hat{\varepsilon}_{t|t-1}$ resemble white noise?
    - If so, the model can be described by the ϕ and θ polynomials.

- If the model residuals do not resemble white noise, then what do they look like?

Iterative model building II

- $\hat{\varepsilon}_{t+k|t}$ will often have a simpler behavior than Y, if the original model $\phi(B)\,Y_t = \theta(B)\varepsilon_t$ captures the essential terms of Y's behavior.

Iterative model building II

- $\hat{\varepsilon}_{t+k|t}$ will often have a simpler behavior than Y, if the original model $\phi(B)\,Y_t = \theta(B)\varepsilon_t$ captures the essential terms of Y's behavior.

1. Construct an ARMA description for $\hat{\varepsilon}_{t|t-1}$: $\phi^*(B)\varepsilon_t = \theta^*(B)\varepsilon_t^*$.

Iterative model building II

- $\hat{\varepsilon}_{t+k|t}$ will often have a simpler behavior than Y, if the original model $\phi(B)Y_t = \theta(B)\varepsilon_t$ captures the essential terms of Y's behavior.

1. Construct an ARMA description for $\hat{\varepsilon}_{t|t-1}$: $\phi^*(B)\varepsilon_t = \theta^*(B)\varepsilon_t^*$.
2. Insert $\varepsilon_t = \phi^{*-1}(B)\theta^*(B)\varepsilon_t^*$ into the original model to obtain the model

$$\phi^*(B)\phi(B)Y_t = \theta(B)\theta^*(B)\varepsilon_t^*$$

Iterative model building II

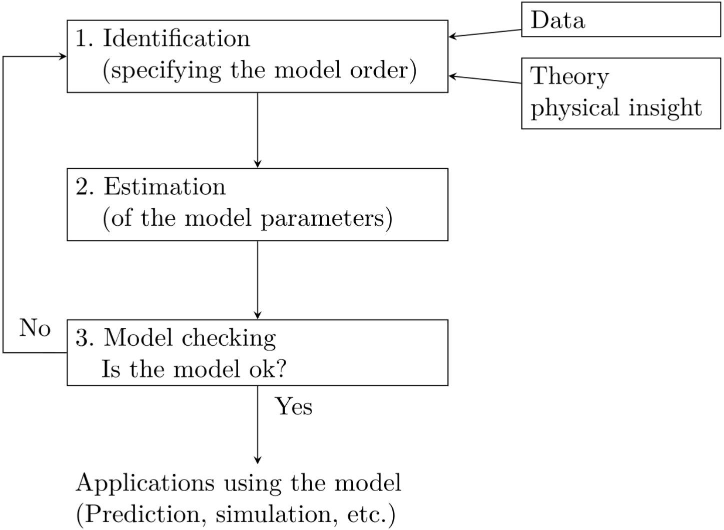
- $\hat{\varepsilon}_{t+k|t}$ will often have a simpler behavior than Y, if the original model $\phi(B)Y_t = \theta(B)\varepsilon_t$ captures the essential terms of Y's behavior.

1. Construct an ARMA description for $\hat{\varepsilon}_{t|t-1}$: $\phi^*(B)\varepsilon_t = \theta^*(B)\varepsilon_t^*$.
2. Insert $\varepsilon_t = \phi^{*-1}(B)\theta^*(B)\varepsilon_t^*$ into the original model to obtain the model

$$\phi^*(B)\phi(B)Y_t = \theta(B)\theta^*(B)\varepsilon_t^*$$

3. Estimate the parameters in the model above with coefficients in $\phi^* \cdot \phi$, $\theta \cdot \theta^*$ varying freely, and proceed to model check.

Model building in general



Residual Analysis

- The order of the model is the minimum order for which the model errors resemble white noise.

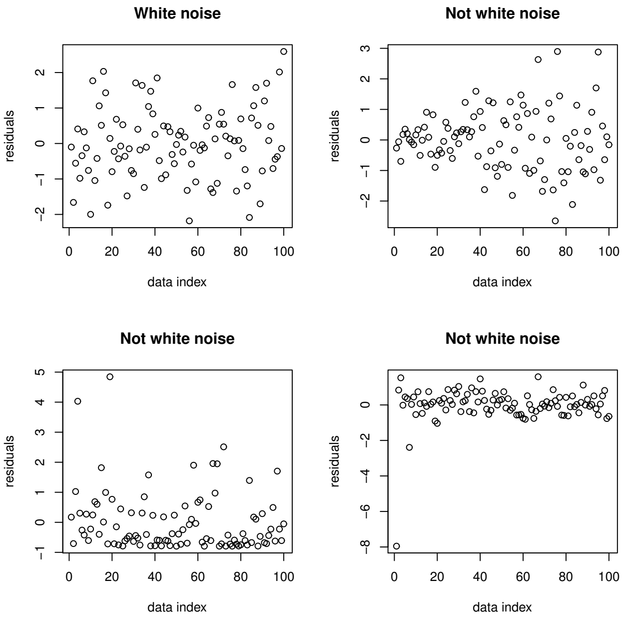
Residual Analysis

- The order of the model is the minimum order for which the model errors resemble white noise.
- How can we check that the model errors resemble white noise?

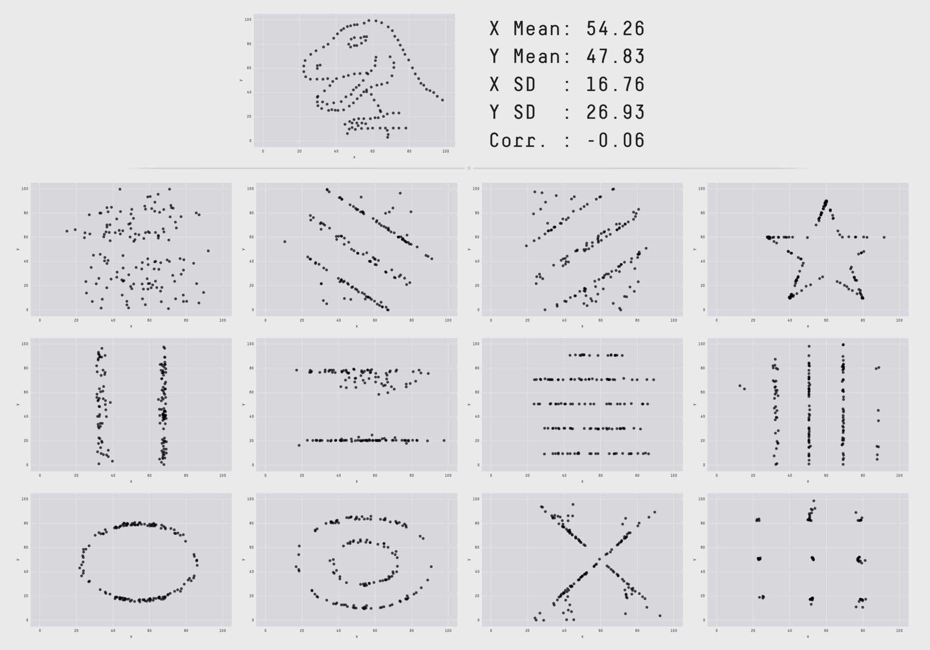
Residual Analysis

- The order of the model is the minimum order for which the model errors resemble white noise.
- How can we check that the model errors resemble white noise?
- First and most important - plot the data.

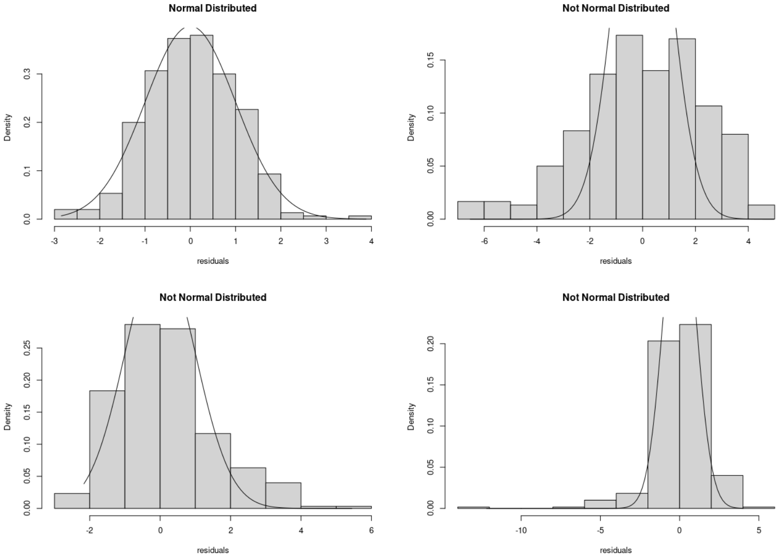
Residual analysis – Plot the data



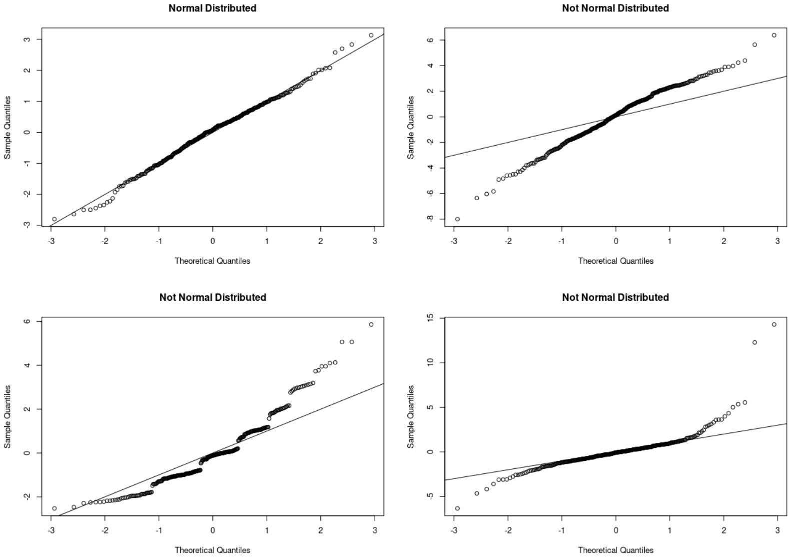
Residual analysis – Plot the data II



Residual analysis – Plot the data III

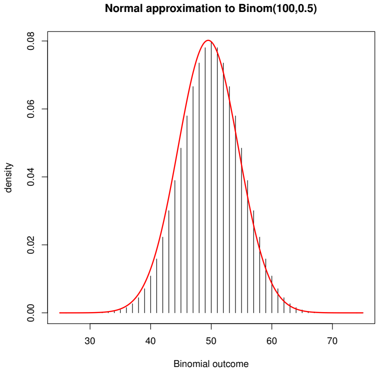


Residual analysis – Plot the data IV

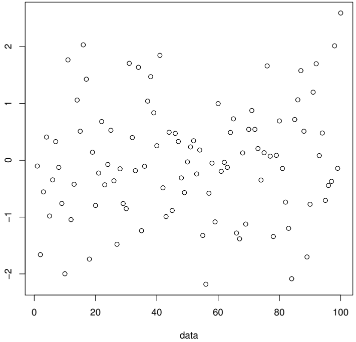


Residual analysis – sign test

- If (ε_t) is white noise, the probability that a new value has a different sign than the previous is $\frac{1}{2}$.
- Number of sign changes: $Binom(N-1, \frac{1}{2})$.
- Approx. normal distribution; $N((N-1)/2, (N-1)/4)$:



Residual analysis – sign test II



- 95% confidence interval for sign changes within 100 white noise residuals: [40; 59]. Actual sign changes from the 100 data: 47.

Residual analysis – sign test III

Sign tests detects both asymmetry and correlation.

Residual analysis – sign test III

Sign tests detects both asymmetry and correlation.

- Too few may indicate positive one-step correlation;

Residual analysis – sign test III

Sign tests detects both asymmetry and correlation.

- Too few may indicate positive one-step correlation;
- Too many may indicate negative one-step correlation;

Residual analysis – sign test III

Sign tests detects both asymmetry and correlation.

- Too few may indicate positive one-step correlation;
- Too many may indicate negative one-step correlation;
- Too few or too many may indicate that P(being above the mean) $\neq \frac{1}{2}$ with no correlation.

Residual analysis - other tests

- There is a bunch of other tests out there.
- You are welcome to use them in assignments, as long as you are sure that you understand them.

Residual analysis – summary

- Plot $\hat{\varepsilon}_{t|t-1}$; do the residuals look stationary? Do they need a transformation?

Residual analysis – summary

- Plot $\hat{\varepsilon}_{t|t-1}$; do the residuals look stationary? Do they need a transformation?
- Plot estimated ACF and PACF, if there are significant lags, then can we use them to extend the model with an ARMA-structure?

Residual analysis – summary

- Plot $\hat{\varepsilon}_{t|t-1}$; do the residuals look stationary? Do they need a transformation?
- Plot estimated ACF and PACF, if there are significant lags, then can we use them to extend the model with an ARMA-structure?
- Plot histogram and/or qq-plot to see whether residuals are normal distributed, if not, then consider a transformation.

Residual analysis – summary

- Plot $\hat{\varepsilon}_{t|t-1}$; do the residuals look stationary? Do they need a transformation?
- Plot estimated ACF and PACF, if there are significant lags, then can we use them to extend the model with an ARMA-structure?
- Plot histogram and/or qq-plot to see whether residuals are normal distributed, if not, then consider a transformation.
- Perform a couple of statistical tests to get some quantitative measures of whether your residuals are alright.

Residual analysis – summary

- Plot $\hat{\varepsilon}_{t|t-1}$; do the residuals look stationary? Do they need a transformation?
- Plot estimated ACF and PACF, if there are significant lags, then can we use them to extend the model with an ARMA-structure?
- Plot histogram and/or qq-plot to see whether residuals are normal distributed, if not, then consider a transformation.
- Perform a couple of statistical tests to get some quantitative measures of whether your residuals are alright.
- Finally, see whether parameters are significant and if not, remove them (you do not need to redo residuals analysis after this).

Information criteria

When considering multiple non-nested candidate models, information criteria can be used:

- Select the model which minimizes some information criterion.
- Akaike's Information Criterion:
  $AIC = -2\log(L(Y_N; \hat{\theta}, \hat{\sigma}_\varepsilon^2)) + 2n_{\text{par}}$
- Bayesian Information Criterion (preferred):
  $BIC = -2\log(L(Y_N; \hat{\theta}, \hat{\sigma}_\varepsilon^2)) + \log(N)n_{\text{par}}$
- AIC is most commonly used, but BIC yields a consistent estimate of the model order.

Cross validation

Cross-validation is possible but slightly less efficient and cumbersome for time series analysis than for other kinds of data.

Cross validation

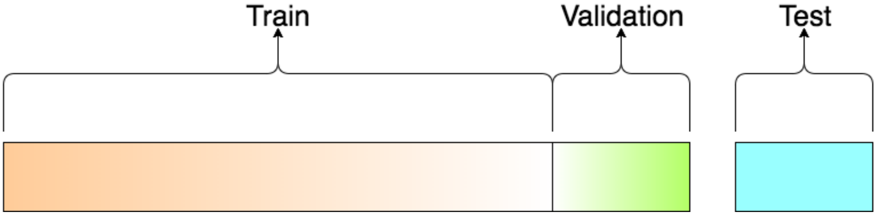
Cross-validation is possible but slightly less efficient and cumbersome for time series analysis than for other kinds of data.

- If we use future measurements we are cheating!

Cross validation

Cross-validation is possible but slightly less efficient and cumbersome for time series analysis than for other kinds of data.

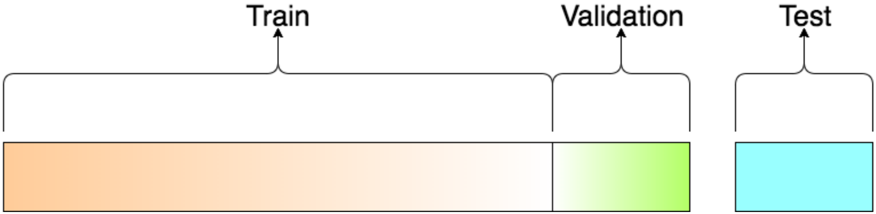
- If we use future measurements we are cheating!
- Thus, it is only possible to split data by having first part be for training, and last part testing.



Cross validation

Cross-validation is possible but slightly less efficient and cumbersome for time series analysis than for other kinds of data.

- If we use future measurements we are cheating!
- Thus, it is only possible to split data by having first part be for training, and last part testing.



- So we must gradually move the part used for training forward in time, it's called "rolling horizon" cross-validation
- Mainly used for forecasting applications
- Remember a burn-in period and then step forward from there