Time Series Analysis

Lasse Engbo Christiansen

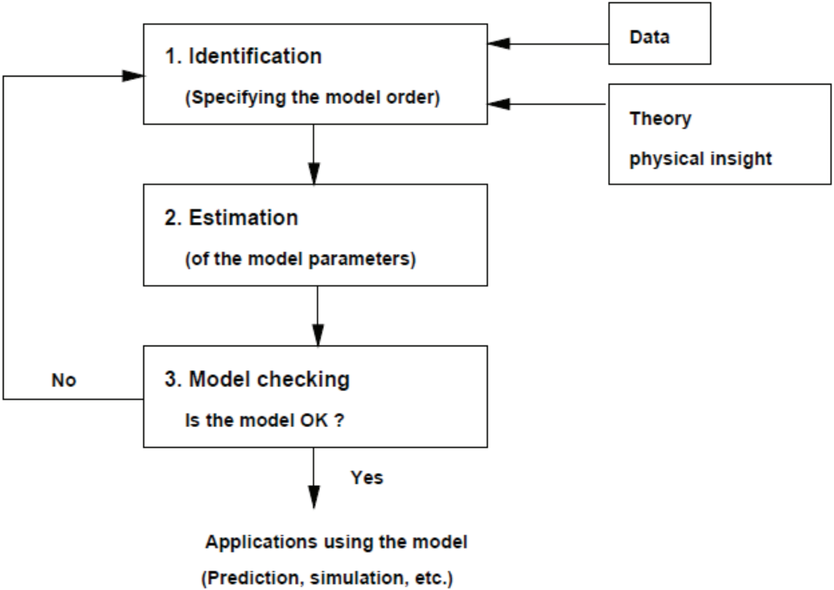
Department of Applied Mathematics and Computer Science
Technical University of Denmark

October 3, 2017

Outline of the lecture

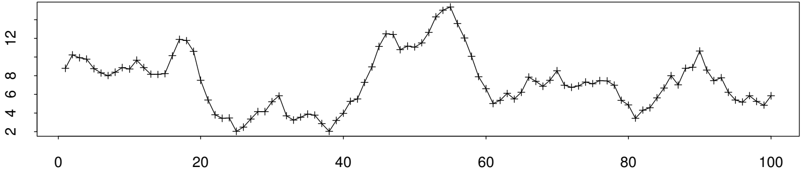
- Identification of univariate time series models, 1st part:
  - Introduction, Sec. 6.1
  - Estimation of auto-covariance and -correlation, Sec. 6.2.1 (and the intro. to 6.2)
  - Using the SACF and SPACF for model order selection
  - Model order selection, Sec. 6.5
  - Model validation, Sec. 6.6

Model building in general



Identification of univariate time series models

- What ARIMA structure would be appropriate for the data at hand? (If any)



- Given the structure we will then consider how to estimate the parameters (later)
- What do we know about ARIMA models which could help us?

Estimation of the autocovariance function

- Estimate of $\gamma(k)$

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

- It is enough to consider $k > 0$
- R: `acf(x, type = "covariance")`

Some properties of $C(k)$

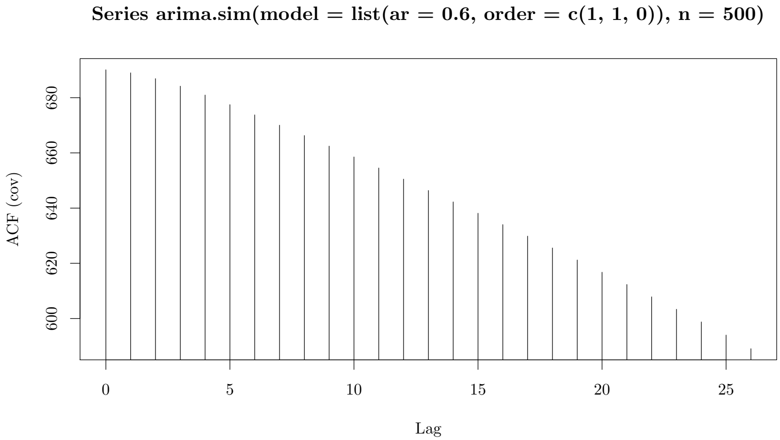
- The estimator is *non-central*:

$$E[C(k)] = \frac{1}{N} \sum_{t=1}^{N-|k|} \gamma(k) = \left(1 - \frac{|k|}{N}\right) \gamma(k)$$

- Asymptotically central (consistent) for fixed k:
  $E[C(k)] \to \gamma(k)$ for $N \to \infty$
- The estimates are correlated themselves (don't trust apparent correlation at random high lags too much)

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

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



Autocorrelation and Partial Autocorrelation

Autocorrelation

- 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

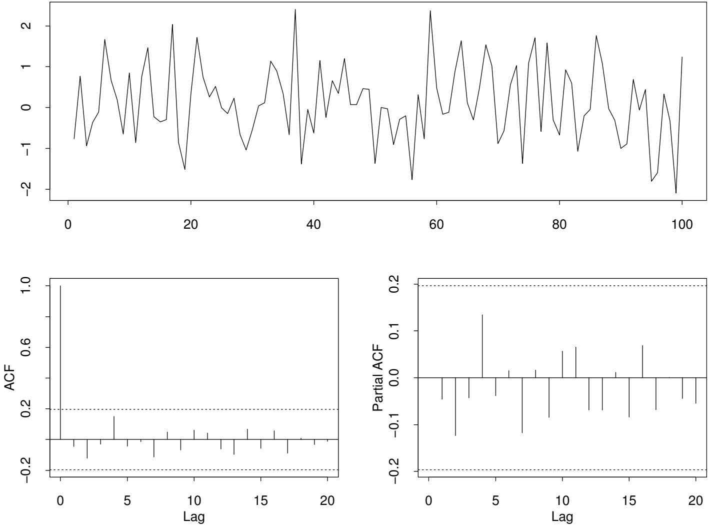
- Sample partial autocorrelation function (SPACF): Use the Yule-Walker equations on $\widehat{\rho}(k)$ (exactly as for the theoretical relations Eq.(5.81))
- It turns out that $\pm 2/\sqrt{N}$ is also appropriate for deciding when the SPACF is zero (more in the next lecture)
- R: `acf(x, type="partial")`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

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 $\max(0, p - q)$

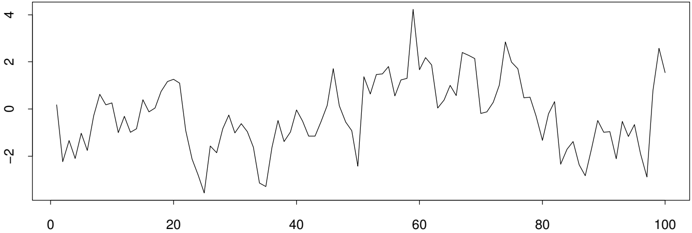
What would be an appropriate structure?

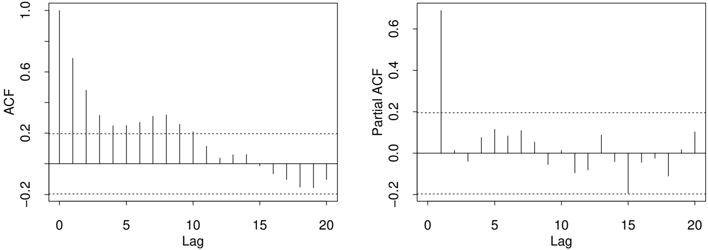


1: White noise 2: AR(1) 3: AR(2)

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

What would be an appropriate structure?

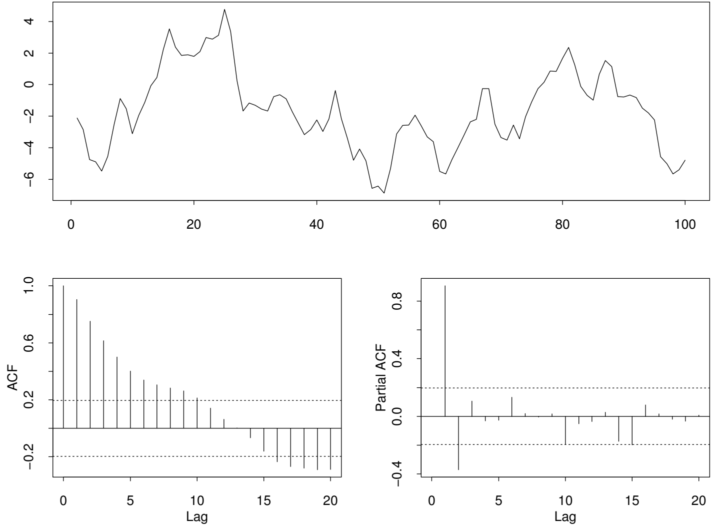




1: White noise   2: AR(1)   3: AR(2)

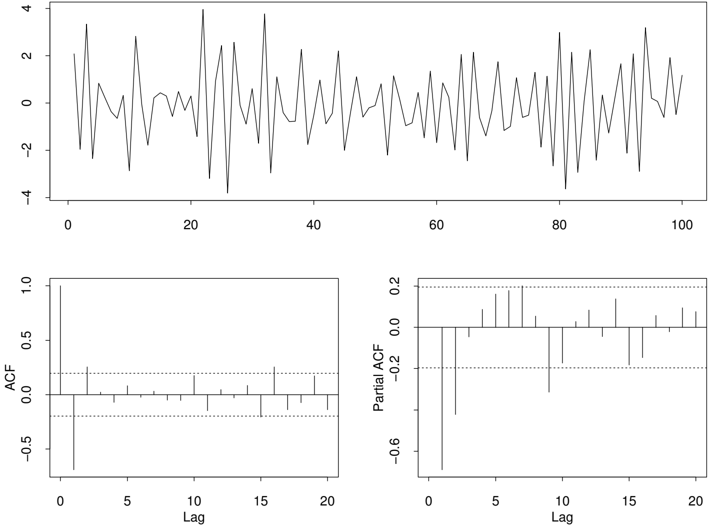
4: MA(1)   5: AR(8)   6: ARMA(1,1)

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)

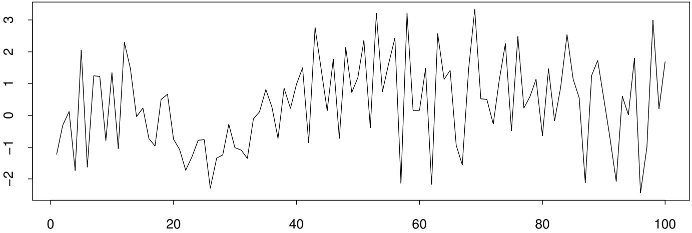
What would be an appropriate structure?

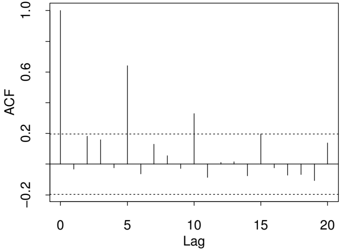


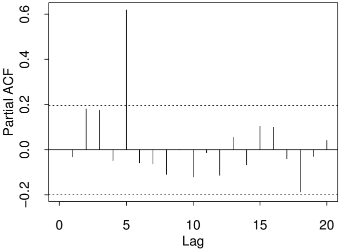
1: White noise   2: AR(1)   3: AR(2)

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

What would be an appropriate structure?



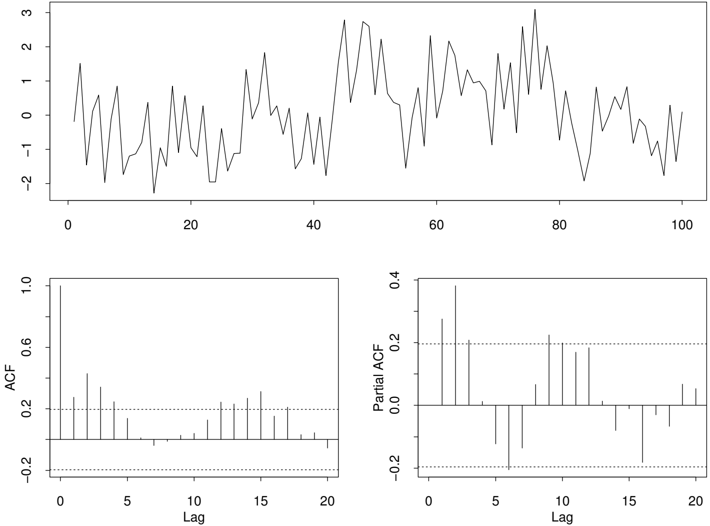




1: AR(1) 2: AR(5) 3: Seasonal AR(2) S=5

4: Seasonal AR(1) S=4 5: Seasonal AR(1) S=5 6: AR(10)

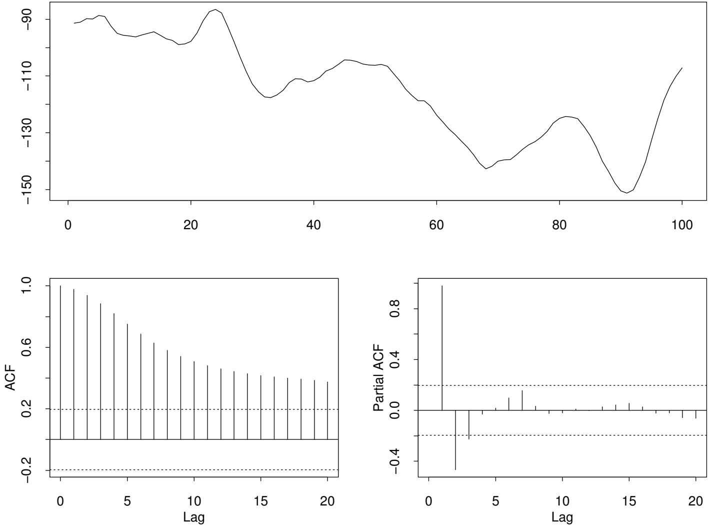
What would be an appropriate structure?



1: AR(4) 2: AR(5) 3: MA(4)

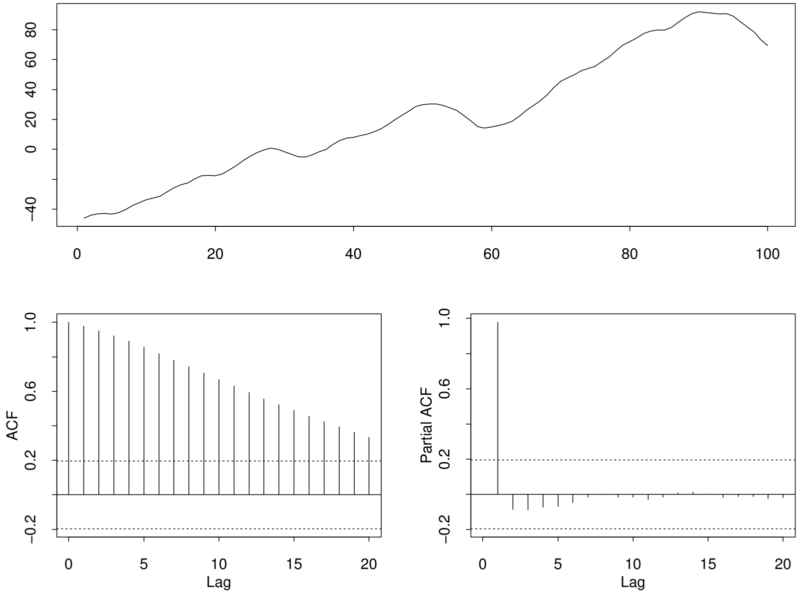
4: ARMA(1,1) 5: AR(2) 6: ARMA(2,2)

What would be an appropriate structure?

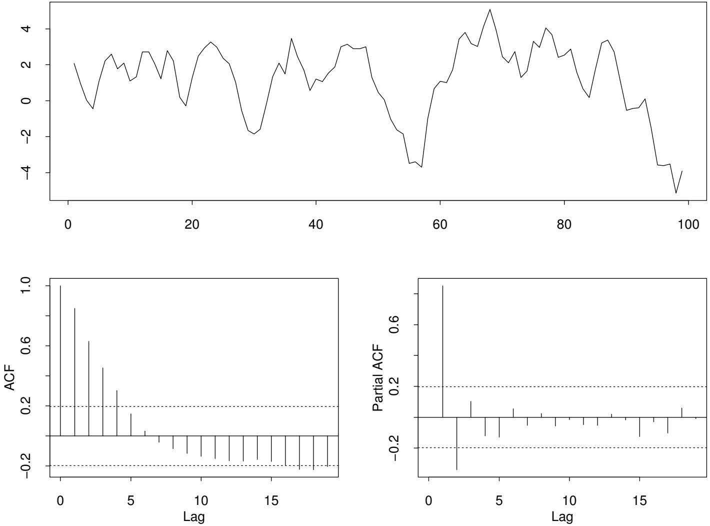


1: AR(3)	2: AR(2)	3: ARIMA(0,1,0)
4: ARMA(1,1)	5: ARIMA(2,1,0)	6: ARIMA(0,2,2)

Example of data from a non-stationary process



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



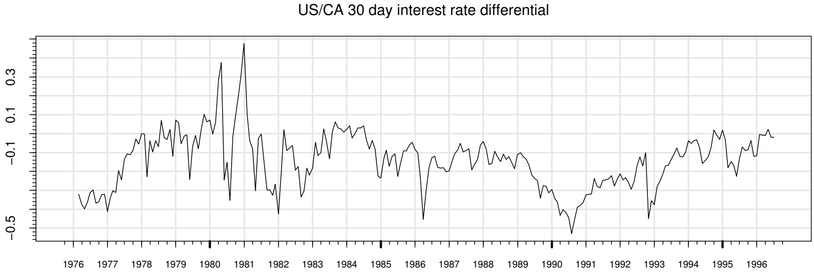
1: AR(3) 2: AR(2) 3: ARIMA(0,1,0)

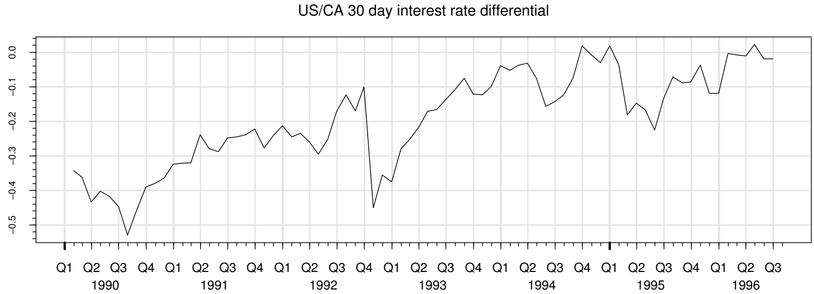
4: ARMA(1,1) 5: ARIMA(2,1,0) 6: ARIMA(0,2,2)

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 … 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
- If that method fails to identify (p, q) because the process:
  - Is not a standard AR-proces (the table should work directly);
  - is not a standard MA-proces (the table should work directly);
  - is not a directly identifiable ARMA proces;
- Then one must do something else...
- Try a small model and reconsider
- Consider transformations ...
  Typically sqrt, log, square, inverse ...

Iterative model building

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

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

2. (Estimation step): Estimate the parameters and calculate the model residuals $W(\hat{\phi}, \hat{\theta})$
3. (Model checking step):
    - Are the estimated parameters significant?
    - Does $W(\hat{\phi}, \hat{\theta})$ 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

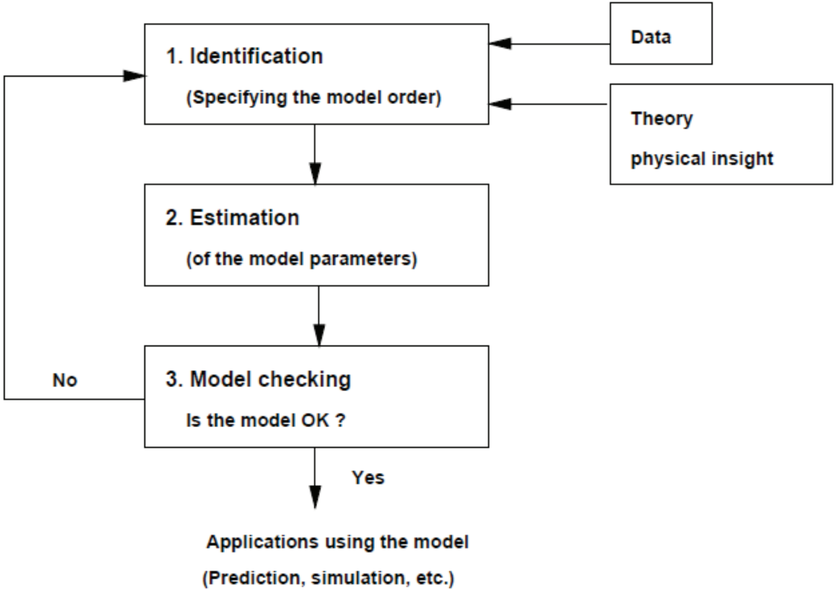
- $W(\hat{\phi}, \hat{\theta})$ will often have a simpler behavior than Y, if the original model $\phi(B)\, Y_t = \theta(B)\, W_t$ captures the essential terms of Y's behavior.

1. Construct an ARMA description for $W(\hat{\phi}, \hat{\theta})$: $\phi^*(B)\, W_t = \theta^*(B)\varepsilon_t$.
2. Insert $W_t = \phi^{*-1}(B)\theta^*(B)\varepsilon_t$ 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.

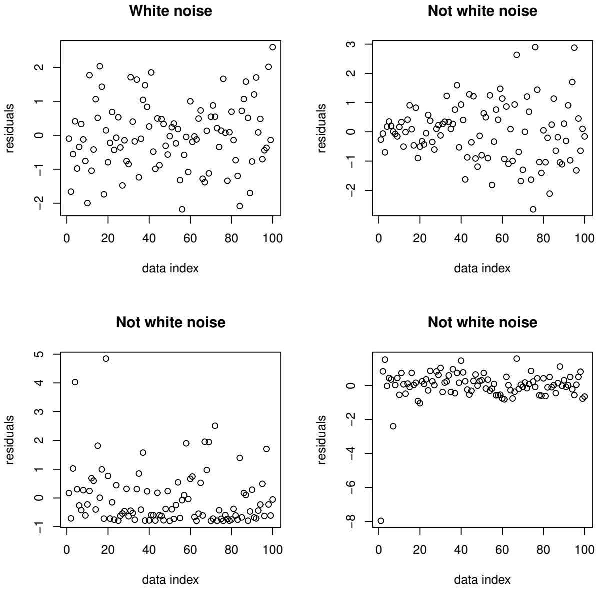
Model building in general



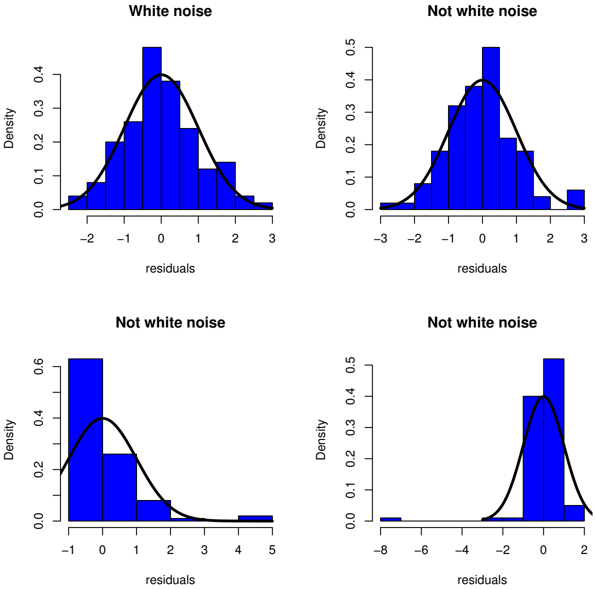
Residual Analysis

- The order of the model is decided when the model errors resemble white noise.
- Is the model order a uniquely determined set of numbers (p,q)?
  NO !
- How can we check that the model errors resemble white noise?
- First and most important - plot the data.

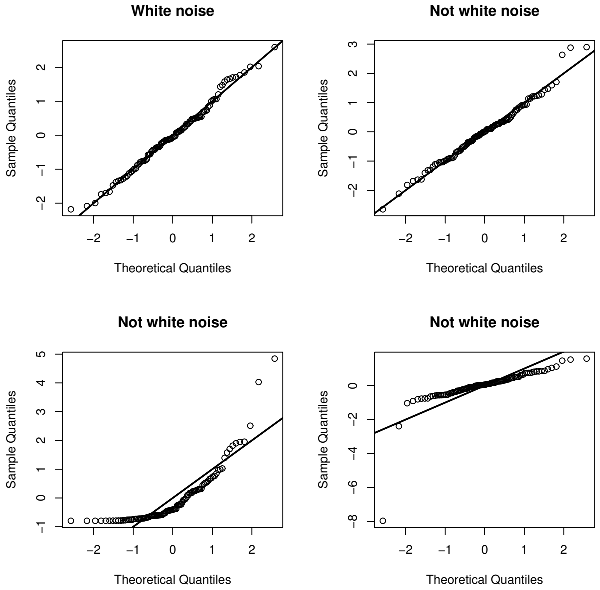
Residual analysis – Plot the data



Residual analysis – Plot the data II

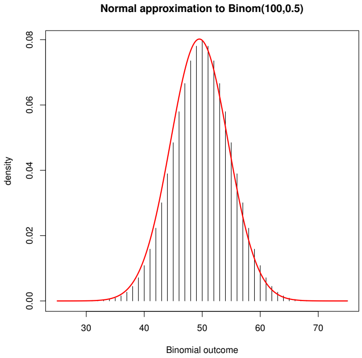


Residual analysis – Plot the data III

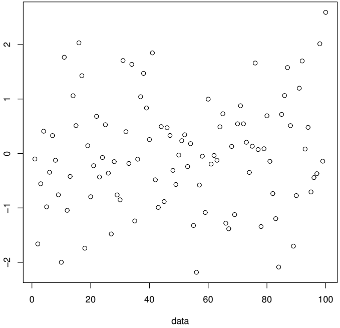


Residual analysis – sign test

- 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.

- 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 – autocorrelation test

- If (ε_t) is white noise, we have seen that $\hat{\rho}_\varepsilon(k) \sim N(0, \frac{1}{N})$ for all k (approx).
  So : $Q^2 = \sum_{i=1}^{m}(\sqrt{N}\hat{\rho}_\varepsilon(i))^2 \sim \chi^2(m)$ (approx).
- If we instead consider the model errors $\varepsilon_t(\hat{\theta})$, $\frac{1}{N}$ is still an upper limit for the variance. However, we obtain less degrees of freedom:

$$Q^2 = \sum_{k=1}^{m}\left(\sqrt{N}\rho_{\varepsilon(\hat{\theta})}(k)\right)^2$$

is approximately distributed as $\chi^2(m-n)$, where n is the number of parameters - IF the residuals are white noise.

Residual analysis – summary

- Plot $\{\varepsilon_t(\hat{\theta})\}$; do the residuals look stationary?
- Tests in the autocorrelation. If $\{\varepsilon_t(\hat{\theta})\}$ is white noise then $(\rho_\varepsilon(k))$ is approximately Gaussian distributed with mean 0 and variance $1/N$. If the model fails, calculate the SPACF also and see if an ARMA-structure for the residuals can be derived (Sec. 6.5.1)
- Since $\hat{\rho}_\varepsilon(k_1)$ and $\hat{\rho}_\varepsilon(k_2)$ are approximately independent for $k_1 \neq k_2$ (Eq. 6.4), the test statistic $Q^2 = \sum_{k=1}^{m} \left(\sqrt{N}\hat{\rho}_{\varepsilon(\hat{\theta})}(k)\right)^2$ is approximately distributed as $\chi^2(m-n)$, where n is the number of parameters.
- R: `tsdiag( output.from.arima )`

Residual analysis – summary II

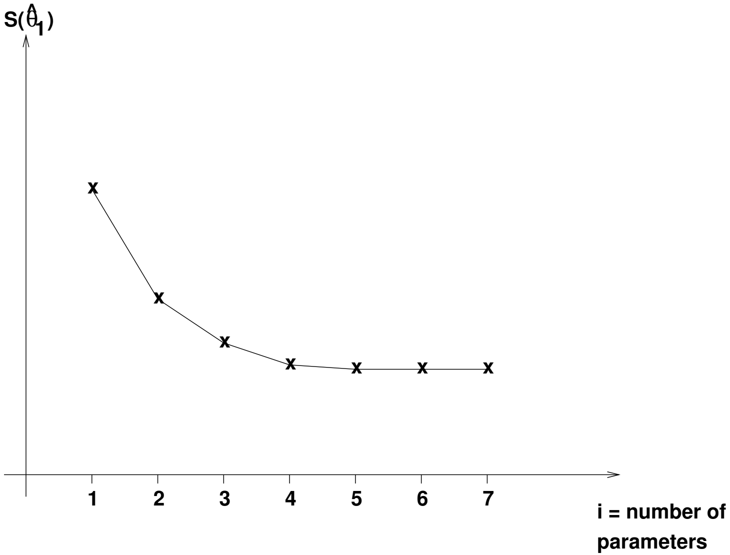
- Test for the number of changes in sign.
  - In a series of length N there is $N-1$ possibilities for changes in sign.
  - If the series is white noise (with mean zero) the probability of change is $1/2$ and the changes will be independent.
  - Therefore the number of changes is distributed as $Bin(N-1, 1/2)$
  - R: `binom.test(No. of changes,N-1).`
- Test in the scaled cumulated periodogram of the residuals is done by plotting it and adding lines at $\pm K_\alpha/\sqrt{q}$, where $q = (N-2)/2$ for N even and $q = (N-1)/2$ for N odd.
  - For $1-\alpha$ confidence limits, K_α can be found in Table 6.2.
  - R (95% confidence interval):

    ```
    > cpgram('residuals')
    ```

Model validation summary: Extensions/Reductions

- Residual analysis (Sec. 6.6.2): Is it possible to detect problems with residuals? (the 1-step prediction errors using the estimates, i.e. $(\varepsilon_t(\hat{\theta}))$ should be white noise).
- If the SACF or the SPACF of $(\varepsilon_t(\hat{\theta}))$ points towards a particular ARMA-structure we can derive how the original model should be extended (Sec. 6.5.1)
- If the model pass the residual analysis it makes sense to test null hypotheses about the parameters (Sec. 6.5.2)

Sum of squared residuals and model size



(It is assumed that the models are nested)

Test for model extension/reduction

- The test essentially checks if the reduction in SSE ($S_1 - S_2$) is large enough to justify the extra parameters in model 2 (n_2 parameters) as compared to model 1 (n_1 parameters). The number of observations used is called N.
- If vector θ_{extra} is used to denote the extra parameters in model 2 as compared to model 1, then the test is formally:

$$H_0 : \theta_{\mathrm{extra}} = 0 \; vs. \; H_1 : \theta_{\mathrm{extra}} \neq 0$$

- If H_0 is true, it (approximately) holds that

$$\frac{(S_1 - S_2)/(n_2 - n_1)}{S_2/(N - n_2)} \sim F(n_2 - n_1, N - n_2)$$

The likelihood ratio test is also a possibility, which may coincide with the above.

Testing one parameter for significance

$$H_0 : \theta_i = 0 \quad \text{against} \quad H_1 : \theta_i \neq 0$$

- Can be done as described on the previous frame.
- Alternatively we can use a t-test based on the estimate and its standard error: $\hat{\theta}_i / \sqrt{V(\hat{\theta}_i)}$
- Under H_0 and for an $ARMA(p, q)$-model this follows a $t(N - p - q)$ distribution (or $t(N - 1 - p - q)$ if we estimated an overall mean of the series)
- Often N is so large compared to the number of parameters that we can just use the standard normal distribution

Information criteria

For models that are not nested, the significance of a model extension cannot be tested.

- Select the model which minimizes some information criterion.
- Akaike's Information Criterion:
  $AIC = -2log(L(Y_N; \hat{\theta}, \hat{\sigma}_\varepsilon^2)) + 2n_{par}$
- Bayesian Information Criterion:
  $BIC = -2log(L(Y_N; \hat{\theta}, \hat{\sigma}_\varepsilon^2)) + log(N)n_{par}$
- Except for an additive constant this can also be expressed as

$$AIC = Nlog(\hat{\sigma}_\varepsilon^2) + 2n_{par}$$

$$BIC = Nlog(\hat{\sigma}_\varepsilon^2) + log(N)n_{par}$$

- AIC is most commonly used, but BIC yields a consistent estimate of the model order.

Highlights

- Estimating the sample ACF
- Table for identification of ARMA models.
- Iterative model building by making model for residuals.
- Residual analysis - several methods
- Testing significance of individual parameters
- Use information criteria when models are not nested. (Typically AIC)