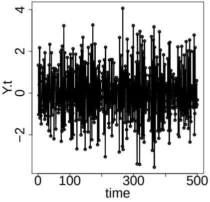
Ch 6: Model specification

- Suppose that we are interested in the time series below;
  - Can we use an ARIMA(p, d, q) model to forecast future outcomes?
  - What model(s) would be appropriate?
- These questions refer to model specification (Ch. 6).
- First (exploratory) step: Sample autocorrelation functions provide insight into what models may (not) be appropriate.
- Material:
  - Ch 6.1, 6.2 (PACF only, not the EACF), 6.4 (except unit root test); the sample autocorrelation function was defined in Ch 3.6, p.49.
  - We'll come back to the remaining material in Ch.6 after we discuss parameter estimation (ch.7).



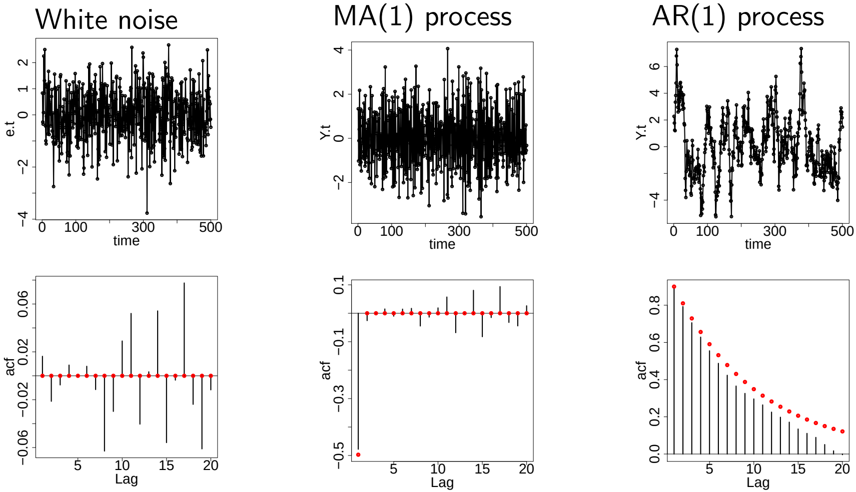
Sample autocorrelation function

- For an observed time series $Y_1, \ldots, Y_n$ with $\bar{Y} = 1/n \sum_{t=1}^{n} Y_t$, the sample autocorrelation function r_k (sample ACF) for time lag $k = 1, 2, \ldots$ is defined as

$$r_k = \frac{\sum_{t=k+1}^{n}(Y_t - \bar{Y})(Y_{t-k} - \bar{Y})}{\sum_{t=1}^{n}(Y_t - \bar{Y})^2}.$$

- For a stationary time series process $\{Y_t\}$, the sample autocorrelation function r_k provides information about the true autocorrelation function $\rho_k = Cor(Y_t, Y_{t-k})$.

Examples: r_k in black, ρ_k in red dots



Why is the sample autocorrelation function different from the true autocorrelation function?

Sampling properties of r_k

- For a stationary time series process $\{Y_t\}$, the sample autocorrelation function r_k is an estimator for the true autocorrelation function ρ_k.
- If r_k is based on a small sample, it may not correspond very closely to ρ_k.
  - For example, observations from a white noise process may result in large r_k's by chance.
- We need to study the properties of r_k to see how it compares to ρ_k for a given time series process and sample size.

Sampling distribution of r_k

- Expression for the sampling distribution for r_k is a bit complicated; we only discuss sampling distributions that would arise in large sample sizes (when we observe Y_t for a long period).
- For a stationary ARMA process, for large sample sizes, the sampling distribution of r_k can be *approximated* as follows:

$$\mathbf{r} \sim N_K(\boldsymbol{\rho}, \boldsymbol{\Sigma}),$$

  where $\mathbf{r} = (r_1, r_2, \ldots, r_K)'$, $\boldsymbol{\rho} = (\rho_1, \rho_2, \ldots, \rho_K)'$, $\Sigma_{i,j} = 1/n \cdot c_{i,j}$, with $c_{i,j} =$ $\sum_{k=-\infty}^{\infty}(\rho_{k+i}\rho_{k+j} + \rho_{k-i}\rho_{k+j} - 2\rho_i\rho_k\rho_{k+j} - 2\rho_j\rho_k\rho_{k+i} + 2\rho_i\rho_j\rho_k^2)$.
- What does this correspond to for white noise?

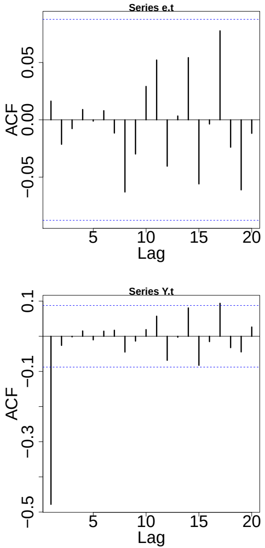
$$r_k \sim N(0, 1/n) \text{ (approximately)}.$$

Large-sample distributions of r_k for AR and MA processes

- Given the approximate distribution for r_k, we find $E(r_k) \approx \rho_k$ for any process.
- Expressions for the covariances for AR(1) and MA(1) are given in the book (p.111), as well as for MA(q) (p.112) (optional material).
- Just note the following take-away point of these derivations related to the variance of r_k:
  - $Var(r_k) \to 0$ as $n \to \infty$.
  - Standard deviations of r_k increase with k for AR(1) and MA(q) $\to$ r_k may be less precise at higher lags.
- When examining sample autocorrelation functions, we need to keep in mind that any suggested models are tentative (exploratory analysis), i.e.,
  - we need model diagnostics to verify whether a suggested model is appropriate,
  - alternative models can/should be explored too.

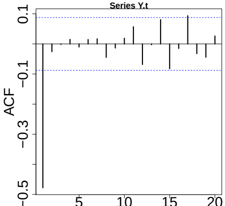
Back to some examples

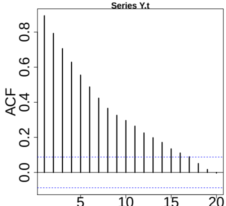
- For white noise, for large samples, we find that approximately $r_k \sim N(0, 1/n)$.
- In the sample acf plot (standard R function), the dashed lines represent $\pm 1.96/\sqrt{n}$ to roughly check whether the time series could be white noise.
- Do the two sample acf's (right) suggest that the underlying process is white noise? If not, what process could be appropriate?



Identifying MA and AR processes

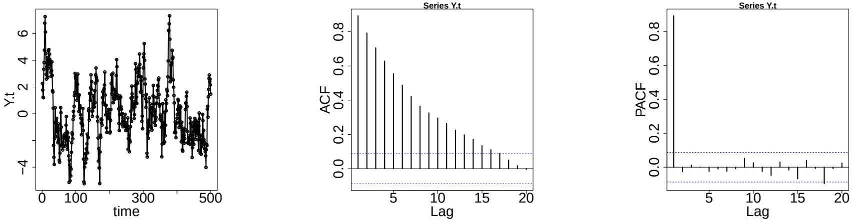
- Reminders:
  - In an MA(q) process, the ACF ρ_k cuts off at lag q.
  - $E(r_k) \approx \rho_k$ for large sample sizes and $Var(r_k) \to 0$ as $n \to \infty$.
- The sample ACF on left suggests an MA(1) model may be appropriate for that series.
  - Indeed, this is the sample ACF for the MA(1) simulation from slide 3 (given on slide 1 as well).
- The sample ACF plotted on the left suggests that an MA model is not appropriate for that series; it's not clear whether some AR(MA) model would be appropriate.
  - In an AR(p) process, ρ_k decays exponentially and may change signs/show cyclic behavior.





A more informative autocorrelation function for AR-processes

- Plots below show the simulated AR(1) series, r_k and a different sample autocorrelation function...
- Wow, what is that function???



Summary so far

- We are moving towards model specification:
  how to figure out what ARIMA model may be appropriate for the time series data at hand?
- We started with a relatively simple function, the sample autocorrelation function r_k, and found that it can provide information about the order of an MA(q) model for MA time series data.
- However, r_k does not provide information about the order p of an AR(p) process...
- The sample partial autocorrelation function does!

Partial autocorrelation function (PACF)

- For any stationary process Y_t, the partial autocorrelation between Y_t and Y_{t-k}, which is denoted by ϕ_{kk} with $-1 \leq \phi_{kk} \leq 1$, is the autocorrelation between Y_t and Y_{t-k}
  - "that is not explained by" $Y_{t-1}, \ldots, Y_{t-k+1}$,
  - after "controlling for" $Y_{t-1}, \ldots, Y_{t-k+1}$.
- Book gives explanation based on best linear predictors (Ch.9).
- Partial autocorrelation may be easiest to understand if we start with a familiar approach on how you could obtain the sample PACF $\hat{\phi}_{kk}$:
  - $\hat{\phi}_{kk}$ is an estimator for ϕ_{kk} based on an observed time series.
  - $\hat{\phi}_{kk}$ can be obtained using a multiple regression model.
- After discussing the multiple regression approach to better understand the sample partial autocorrelation function, we will discuss how to use the YW equations to obtain the PACF and sample PACF (which is the approach implemented in the R functions that we will use).

Sample partial autocorrelation function $\hat{\phi}_{kk}$ (sample PACF)

- $\hat{\phi}_{kk}$ can be obtained using a multiple regression model: $\hat{\phi}_{kk}$ is given by the least-squares estimate of the coefficient of Y_{t-k} in the multiple linear regression model where we regress Y_t on Y_{t-k}, as well as the in-between Y's:

$$Y_t = \phi_{k1} Y_{t-1} + \phi_{k2} Y_{t-2} + \ldots + \phi_{kk} Y_{t-k} + e_t,$$

  assuming $e_t \sim N(0, \sigma_e^2)$ (independent, even if that assumption doesn't hold true).
- The regression model corresponds to an AR(k) model: $\hat{\phi}_{kk}$ measures the excess correlation at lag k that is not yet accounted for by an AR($k-1$) model.
- If Y_t is an AR(p) process, do you expect that $|\hat{\phi}_{kk}|$ for $k > p$ is close to 0 or close to 1?

Calculating the (sample) PACF using the YW equations

- The definition (and easy method) that we will use for deriving the (sample) PACF is based on the Yule-Walker (YW) equations.
- Remember that for an AR(k) model with

$$Y_t = \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \ldots + \phi_k Y_{t-k} + e_t,$$

  we discussed that the following holds true (YW equations):

$$\rho_j = \phi_1 \rho_{j-1} + \phi_2 \rho_{j-2} + \phi_3 \rho_{j-3} + \ldots + \phi_k \rho_{j-k} \text{ for } j = 1, 2, 3, \ldots$$

- To obtain the PACF ϕ_{kk} at lag k via the YW equations, we consider the following (AR(k)) model:

$$Y_t = \phi_{k1} Y_{t-1} + \phi_{k2} Y_{t-2} + \ldots + \phi_{kk} Y_{t-k} + e_t,$$

  with corresponding YW equations:

$$\rho_j = \phi_{k1} \rho_{j-1} + \phi_{k2} \rho_{j-2} + \phi_{k3} \rho_{j-3} + \ldots + \phi_{kk} \rho_{j-k} \text{ for } j = 1, 2 \ldots$$

- Solving that set of equations for $j = 1, 2, \ldots, k$ gives ϕ_{kk} (as well as the other ϕ_{kj}'s, but we are only interested in ϕ_{kk}).
  - For an AR(p) process, it follows that $\phi_{kk} = 0$ for $k > p$.

PACF calculation method

- The YW equations for the PACF can be solved recursively using equations:

$$
\begin{aligned}
\phi_{kk} &= \frac{\rho_k - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_{k-j}}{1 - \sum_{j=1}^{k-1} \phi_{k-1,j}\rho_j}, && (1) \\
\phi_{k,j} &= \phi_{k-1,j} - \phi_{k,k} \cdot \phi_{k-1,k-j}. && (2)
\end{aligned}
$$

- The steps are as follows:
  1. Set $\phi_{11} = \rho_1$ (note typo in book).
  2. Find ϕ_{22} using Eq.(1), which for $k = 2$ corresponds to $\phi_{22} = \frac{\rho_2 - \phi_{11}\rho_1}{1-\phi_{11}\rho_1} = \frac{\rho_2 - \rho_1^2}{1-\rho_1^2}$.
  3. Find $\phi_{2,j}$ for $j = 1$ using Eq.(2), which for $k = 2$ corresponds to $\phi_{21} = \phi_{11} - \phi_{22}\phi_{11}$.
  4. Repeat steps for $k = 3, 4, \ldots$ to obtain $\phi_{33}, \phi_{44}, \ldots$:
     - 4.1 Find ϕ_{kk} using Eq.(1),
     - 4.2 Find $\phi_{k,j}$ for $j = 1, \ldots, k-1$ using Eq.(2).

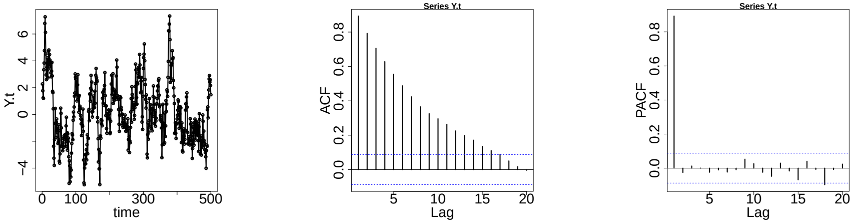
PACF and sample PACF

- PACF ϕ_{kk} refers to
  - partial autocorrelation for lag k.
  - autocorrelation between Y_t and Y_{t-k} that is not explained by in-between Y's.

  We have an recursive method for obtaining ϕ_{kk} using ρ_i for $i = 1, 2, \ldots, k$.
- The *sample PACF* function $\hat{\phi}_{kk}$ estimates ϕ_{kk} based on an observed time series and can be obtained using the PACF recursive method, replacing ϕ_{kj} by $\hat{\phi}_{kj}$ and ρ_k by sample ACF r_k for all k, j.
- Note that for an AR(p) process:
  - $\phi_{kk} = 0$ for $k > p$,
  - for large sample sizes n, $\hat{\phi}_{kk} \sim N(0, 1/n)$ approximately for $k > p$.
- Ah, so we can use the sample PACF to get information on the order of an AR(p) model for a given time series:
  we expect that, approximately, $|\hat{\phi}_{kk}| < 2/\sqrt{n}$ for $k > p$.

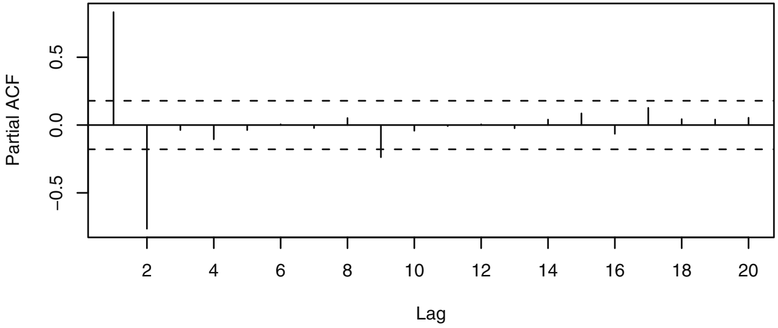
From earlier slide: A more informative autocorrelation function for AR-processes

- Plots below show the simulated AR(1) series, r_k and the sample PACF $\hat{\phi}_{kk}$.
- The ACF and PACF suggest an AR(1) model.



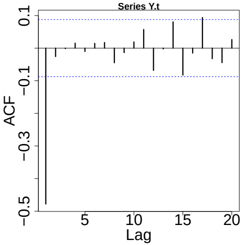
Example of sample PACF for AR(2) simulation

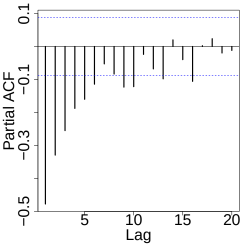
Exhibit 6.13 Sample PACF for an AR(2) Process with $\phi_1 = 1.5$ and $\phi_2 = -0.75$



What about MA processes?

- Sample ACF and PACF for MA(1) simulation (from earlier slides) shown below.
- For MA processes ϕ_{kk} never equals zero but decays to 0 exponentially fast.
- So how to find out about p and q in ARMA(p, q) processes?
  - The extended autocorrelation function can be used (see book) but it doesn't seem to work very well in simulation settings, so we won't discuss it.
  - We will focus on model selection procedures for ARMA processes based on information criteria (e.g., a la AIC).

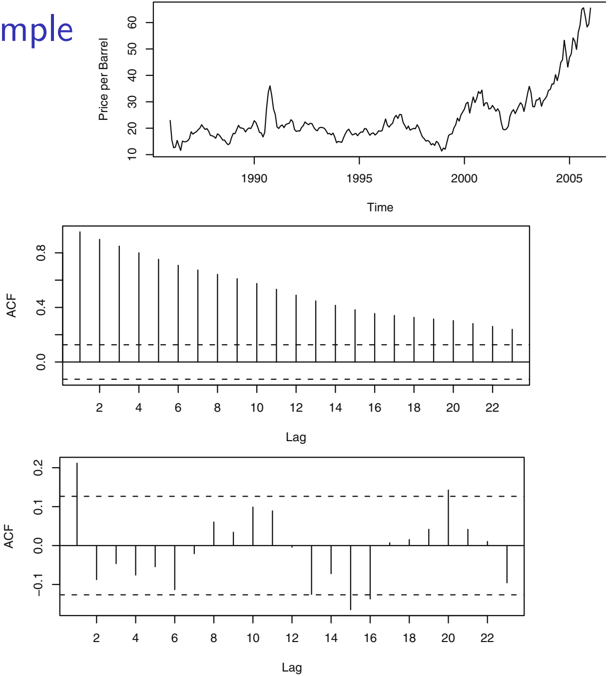




Non-stationarity

- What if we're dealing with a non-stationary time series?
- Not clear what ACF and PACF are estimating (only sensible when γ_k is constant with time).
- However, the ACF can still be informative to flag non-stationarity: autocorrelation tends to be high, thus the sample ACF typically fails to die out for increasing lags.
- When detecting possible non-stationarity, one approach is to revert to differencing as discussed in Ch.5.

Oil price example



Overdifferencing

- When detecting possible non-stationarity, one approach is to revert to differencing as discussed in Ch.5.
- Watch out for overdifferencing though: differencing a series that is already stationary can introduce additional autocorrelations and complicate modeling the time series.
- Example:
  - Suppose Y_t is a random walk, $Y_t = Y_{t-1} + e_t$.
  - $\nabla Y_t = e_t$ (white noise) but

$$\nabla^2 Y_t = e_t - e_{t-1},$$

    which is an MA(1) model with $\theta = 1$.
- Keep in mind the principle of model parsimony: try to find the simplest model that describes the data.

Summary

- You now know of 2 functions (tools) to explore whether an AR or MA process may be suitable, and what values for p and q to attempt.
- Overview:

Model	Features
MA(q)	ACF cuts off at lag q, PACF tails off
AR(p)	PACF cuts off at lag p, ACF tails off
ARMA(p, q)	ACF/PACF tail off

  The series could be non-stationary if ACF does not die out rapidly.
- We will discuss AIC-ish model selection criteria (an alternative way to do model selection) and the unit root test (to check stationarity more formally) later.
- First: parameter estimation (Ch. 7)!