Eco 5316 Time Series Econometrics
Spring 2016

Version: February 4, 2016, 3:08pm

Box-Jenkins methodology

1. *Indentifcation*

- graph data to determine if any transformations are necessary (logarithms, differencing, ...) to get weakly stationary time series, examine series for trend (linear/nonlinear), periods of higher volatility, outliers, seasonal patterns, structural breaks, missing data,...
- examine ACF and PACF of the transformed data to determine plausible models to be estimated
- use Q-statistics to test whether groups of autocorrelations are statistically significant

2. *Estimation*

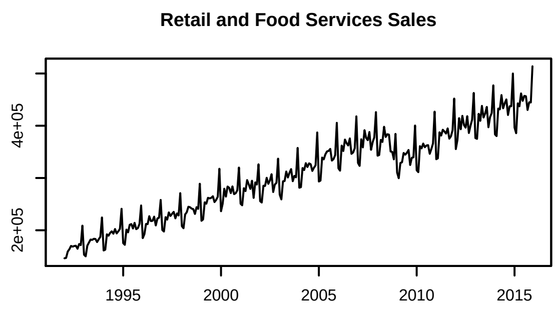
- model estimated using either conditional likelihood method or exact likelihhod method
- estimate the models considered and select the best - coefficients should be statistically significant, AIC and SBC should be low

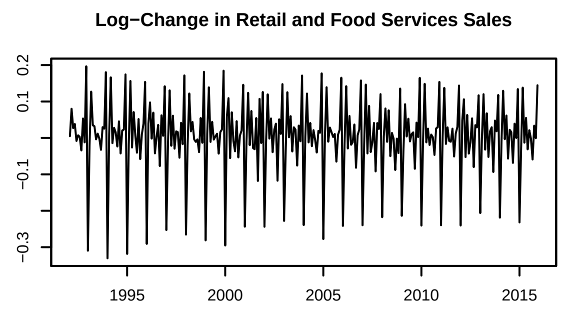
3. *Checking Model Adequacy*

- inspect residuals - if the model was well specified residuals should be very close to white-noise
  - plot residuals, look for outliers, periods in which the model does not fit the data well (evidence of structural change)
  - examine ACF and PACF of the residuals to check for significant autocorrelations
  - use Q-statistics to determine if groups of autocorrelations are statistically significant
- estimated coefficients should be consistent with the underlying assumption of stationarity
- check model for parameter instability and structural change
- out-of-sample forecasting of known data values

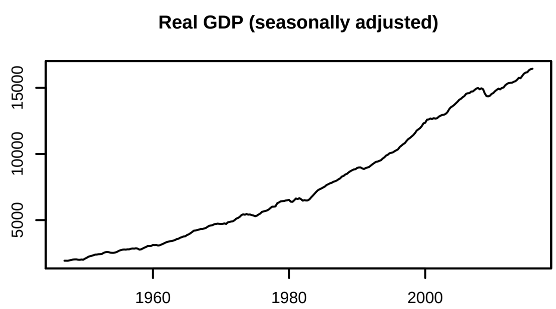
Trend, Volatility, Seasonality, Structural Change, Outliers

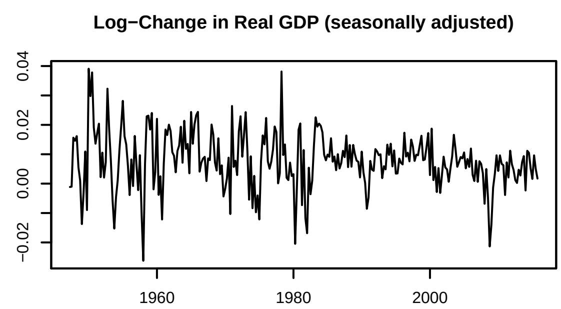
Retail and Food Services Sales https://www.quandl.com/data/FRED/RSAFSNA



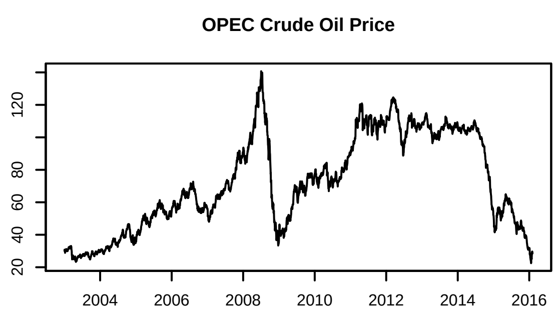


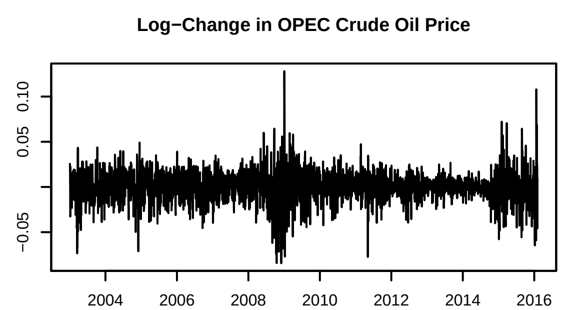
Real GDP https://www.quandl.com/data/FRED/GDPC1



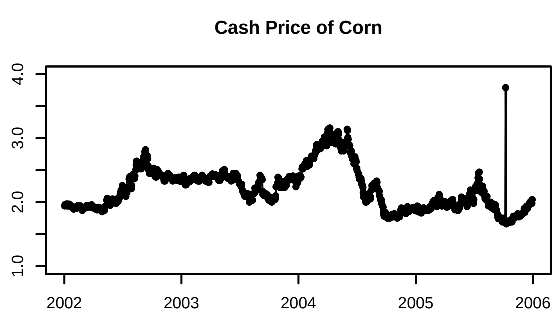


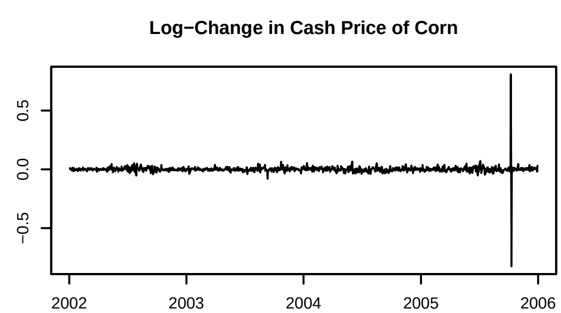
OPEC Crude Oil Price https://www.quandl.com/data/OPEC/ORB





Cash Price of Corn (October 8, 2005) https://www.quandl.com/data/TFGRAIN/CORN





AR(p) model

Suppose that $\{a_t\}$ is a white noise with $E(a_t) = 0$ and $Var(a_t) = \sigma_a^2$. Time series process $\{y_t\}$ then follows an autoregressive model of order 1, or AR(1), if

$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t$$

or equivalently, using the backshift operator

$$(1 - \phi_1 B) y_t = \phi_0 + a_t$$

More generaly, time series $\{y_t\}$ then follows an autoregressive model of order p, or AR(p), if

$$y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + a_t$$

or equivalently, using the backshift operator

$$(1 - \phi_1 B - \ldots - \phi_p B^p) y_t = \phi_0 + a_t$$

Autocorrelation function (ACF)

Linear dependence between y_t and y_{t-l} is measured by the correlation coefficient ρ_l. For a weakly stationary time series process$\{y_t\}$ we have

$$\rho_l = \frac{cov(y_t, y_{t-l})}{\sqrt{Var(y_t)Var(y_{t-l})}} = \frac{cov(y_t, y_{t-l})}{Var(y_t)} = \frac{\gamma_l}{\gamma_0}$$

and the *theoretical autocorrelation function* is then given by $\{\rho_1, \rho_2, \ldots\}$. Given a sample $\{y_t\}_{t=1}^T$ correlation coefficients ρ_l can be estimated as

$$\hat{\rho}_l = \frac{\sum_{t=l+1}^T (y_t - \bar{y})(y_{t-l} - \bar{r})}{\sum_{t=1}^T (y_t - \bar{y})^2}$$

where $\bar{y} = \sum_{t=1}^T y_t$, and the *sample autocorrelation function* is then given by $\{\hat{\rho}_1, \hat{\rho}_2, \ldots\}$

Autocorrelation function for AR(p) model

If $p = 1$ it can be shown that $\gamma_0 = Var(y_t) = \frac{\sigma_a^2}{1-\phi_1}$ and that in addition that $\gamma_l = \phi_1 \gamma_{l-1}$ for $l > 0$. Thus

$$\rho_l = \phi_1 \rho_{l-1} \tag{1}$$

and since $\rho_0 = 1$, we get $\rho_l = \phi_1^l$. Because for weakly stationary $\{y_t\}$ it has to hold that $|\phi_1| < 1$, the theoretical ACF thus decays exponentially, in either direct or oscilating way - see the two examples for AR(1) on the next page.

If $p = 2$ the theoretical ACF for AR(2) satisfies the following second order difference equation

$$\rho_l = \phi_1 \rho_{l-1} + \phi_2 \rho_{l-2} \tag{2}$$

or equivalently using the backshift operator $(1 - \phi_1 B - \phi_2 B^2)\rho_l = 0$. The solutions of the associated *characteristic equation*

$$1 - \phi_1 \omega - \phi_2 \omega^2 = 0$$

are $x_{1,2} = -\frac{\phi_1 \pm \sqrt{\phi_1^2 + 4\phi_2}}{2\phi_2}$. Their inverses $\omega_{1,2} = 1/x_{1,2}$ are called the **characteristic roots** of the AR(2) model. If the discriminant $D = \phi_1^2 + 4\phi_2$ is positive ω_1, ω_2 are real numbers, and ACF is a combination of two exponantial decays. If $D < 0$ characteristic roots are complex conjugantes, and ACF will resemble a dampened sine wave. For weak stationarity all characteristic roots need to lie inside the unit circle, that is $|\omega_i| < 1$ for $i = 1, 2$. Finally, from equation (2) we get $\rho_1 = \frac{\phi_1}{1-\phi_2}$ and $\rho_l = \rho_{l-1} + \phi_2 \rho_{l-2}$ for $l \geq 2$.

In general the theoretical ACF for AR(p) satisfies the following difference equation of order p

$$(1 - \phi_1 B - \ldots - \phi_p B^p)\rho_l = 0 \tag{3}$$

The characteristic equation of the AR(p) model is thus $1 - \phi_1 x - \ldots - \phi_p x^p = 0$, and the AR($p$) process is weakly stationary if the characteristic roots (i.e. inverses of the solutions of the characteristic equation) lie inside of the unit circle. The plot of the ACF of a weakly stationary AR(p) process will show a mixture of exponential decays and dampened sine waves.

Partial autocorrelation function (PACF)

Consider the following system of AR models that can be estimated by OLS

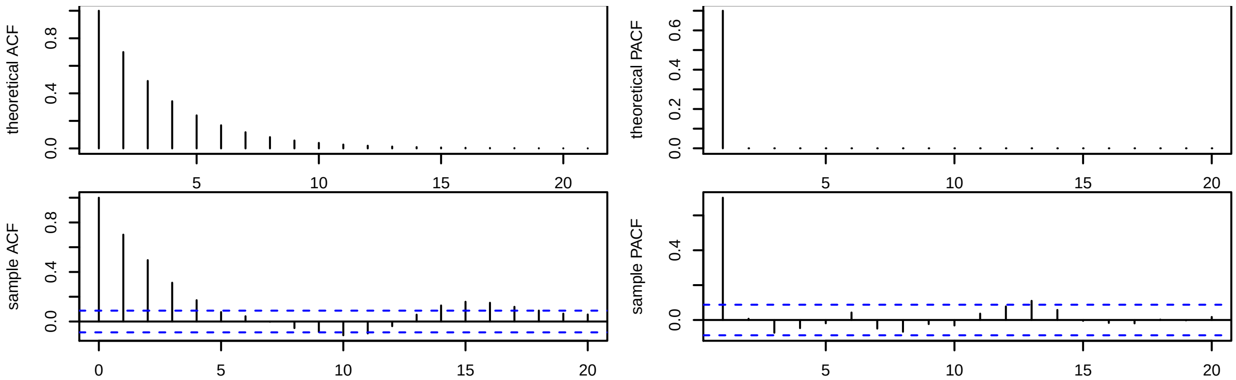
$$\begin{aligned}
y_t &= \phi_{0,1} + \phi_{1,1} y_{t-1} + e_{1,t} \\
y_t &= \phi_{0,2} + \phi_{1,2} y_{t-1} + \phi_{2,2} y_{t-2} + e_{2,t} \\
y_t &= \phi_{0,3} + \phi_{1,3} y_{t-1} + \phi_{2,3} y_{t-2} + \phi_{3,3} y_{t-3} + e_{3,t} \\
&\vdots
\end{aligned}$$

Estimated coefficients $\hat{\phi}_{1,1}, \hat{\phi}_{2,2}, \hat{\phi}_{3,3}, \ldots$ form the sample *partial autocorrelation function* (PACF). If the time series process $\{y_t\}$ comes from an AR(p) process, sample PACF should have $\hat{\phi}_{j,j}$ close to zero for $j > p$.

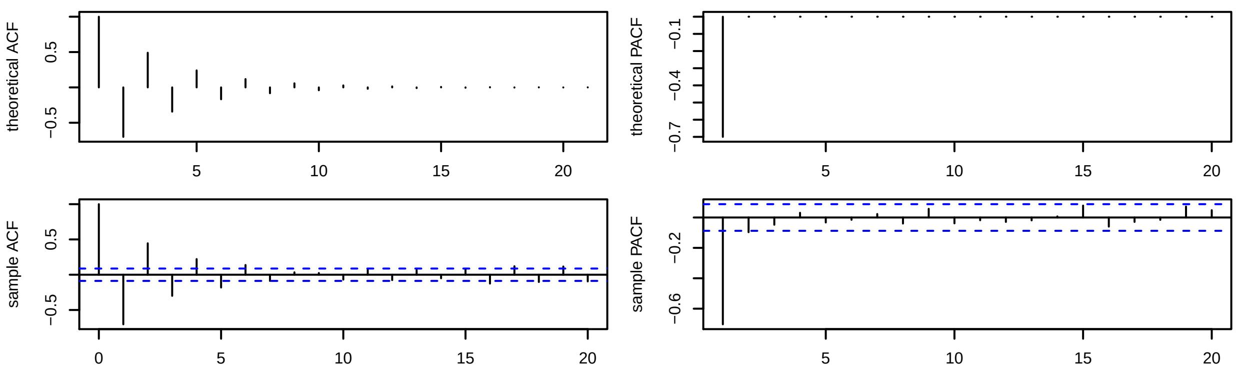
For an AR(p) with Gaussian white noise as T goes to infinity $\hat{\phi}_{p,p}$ converges to ϕ_p and $\hat{\phi}_{l,l}$ converges to 0 for $l > p$, in addition the assymptotic variance of $\hat{\phi}_{l,l}$ for $l > p$ is $1/T$. This is the reason why the interval plotted by R in the plot of PACF is $0 \pm 2/\sqrt{T}$. The order of the AR process can thus be determined by finding the lag after which PACF cuts off to zero.

An overview of ACF and PACF for simulated AR(p) model can be found here.

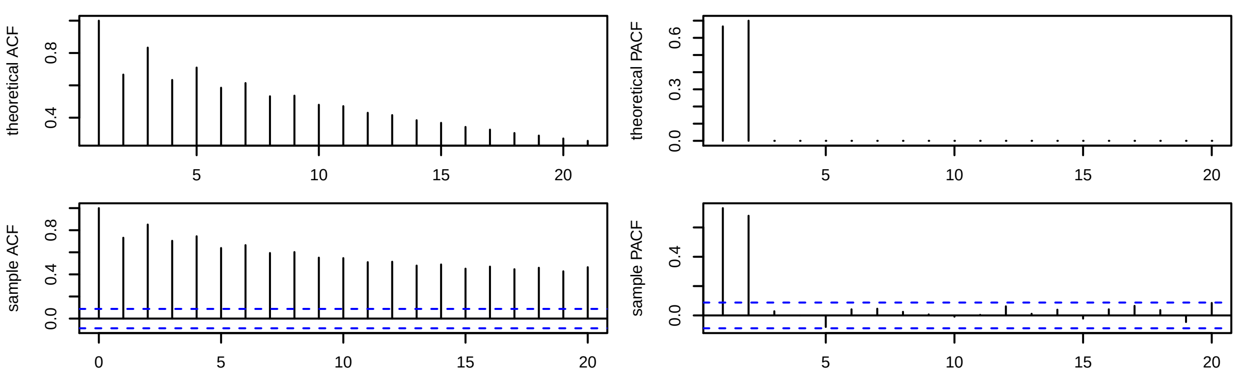
AR(1) with $\phi_1 = 0.7$



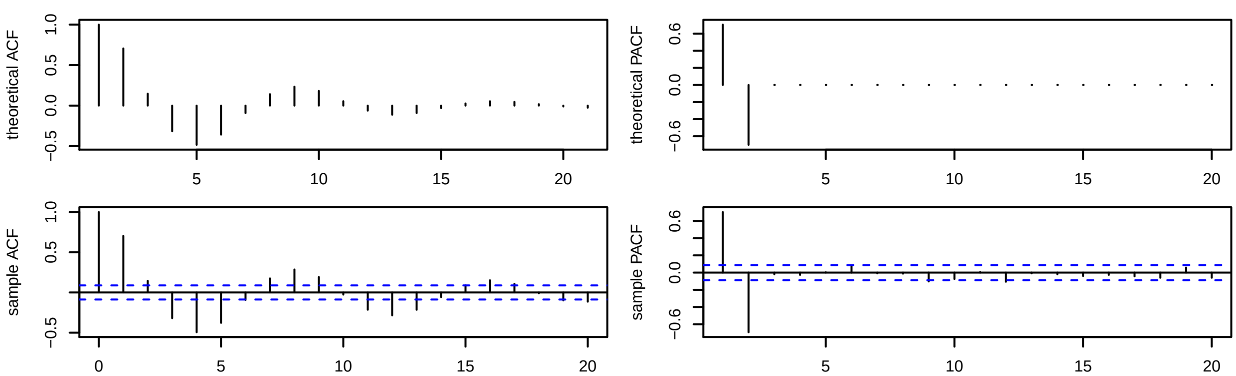
AR(1) with $\phi_1 = -0.7$



AR(2) with $\phi_1 = 0.2$, $\phi_2 = 0.7$



AR(2) with $\phi_1 = 1.2$, $\phi_1 = -0.7$



MA(q) model

Time series process $\{y_t\}$ follows a moving average model of order 1, or MA(1) model, if

$$y_t = c_0 + a_t + \theta_1 a_{t-1}$$

where $\{a_t\}$ is a white noise. Note that using the backshift operator we can write $y_t = c_0 + (1 + \theta_1 B)a_t$.

In general, for $q > 0$ a moving average model of order q, or MA(q) model, is given by

$$y_t = c_0 + a_t + \theta_1 a_{t-1} + \ldots + \theta_q a_{t-q}$$

Equivalently, using the backshift operator $y_t = c_0 + (1 + \theta_1 B + \ldots + \theta_q B^q)a_t$.

Autocorrelation function for MA(q) model

If $q = 1$ it can be shown that $\gamma_0 = Var(y_t) = (1 + \theta_1^2)\sigma_a^2$, and also that $\gamma_1 = \theta_1 \sigma_a^2$ and $\gamma_l = 0$ for $l \geq 2$. Thus

$$\rho_1 = \frac{\theta_1}{1 + \theta_1^2} \qquad \rho_l = 0 \text{ for } l > 1 \tag{4}$$

The theoretical ACF for MA(1) thus cuts off to zero after lag 1. See the top two examples on the next page.

If $q = 2$ we have $\gamma_0 = Var(y_t) = (1 + \theta_1^2 + \theta_2^2)\sigma_a^2$, and the theoretical ACF for MA(2) satisfies

$$\rho_1 = \frac{\theta_1 + \theta_1\theta_2}{1 + \theta_1^2 + \theta_2^2} \qquad \rho_2 = \frac{\theta_2}{1 + \theta_2^2 + \theta_1^2} \qquad \rho_l = 0 \text{ for } l > 2 \tag{5}$$

The theoretical ACF for MA(1) thus cuts off to zero after lag 2. See the bowttom two examples on the next page.

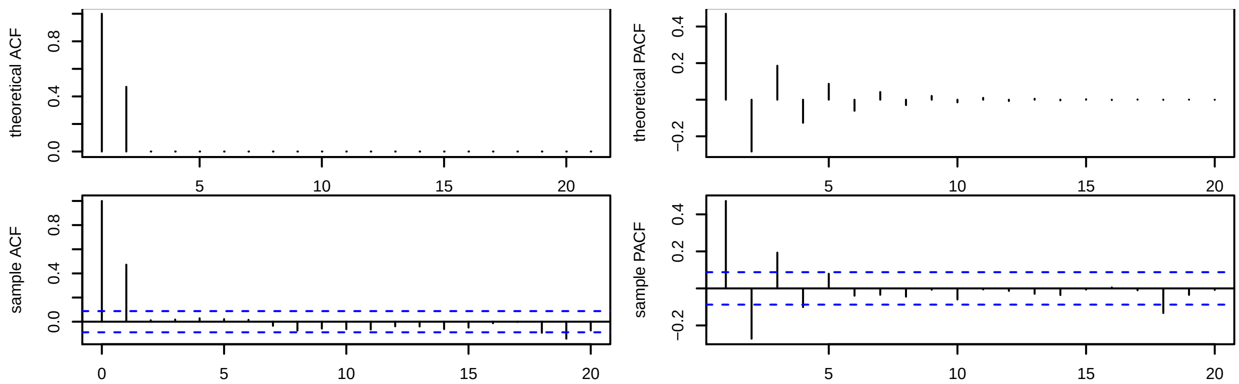
In general, for for MA(q) model $\gamma_0 = Var(y_t) = (1 + \theta_1^2 + \ldots + \theta_q^2)\sigma_a^2$ and its theoretical ACF satisfies $\rho_l = 0$ for $l > q$.

The ACF is thus useful for identifying the order of an MA model in the sama way the PACF is useful in identifying the order of an AR model.

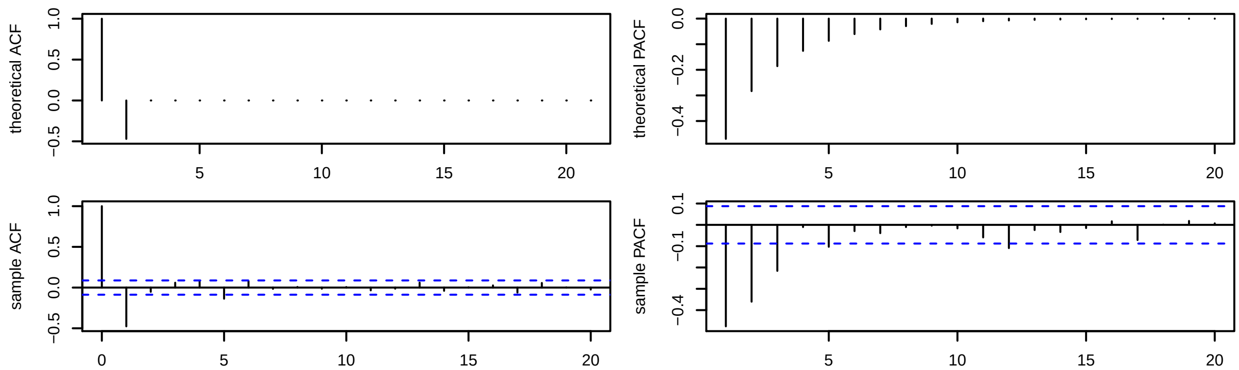
Note also that MA(q) model is always weakly stationary, unlike AR(p) model, which is only weakly stationary if its characteristic roots lie inside the unit circle.

An overview of ACF and PACF for simulated MA(q) models can be found here.

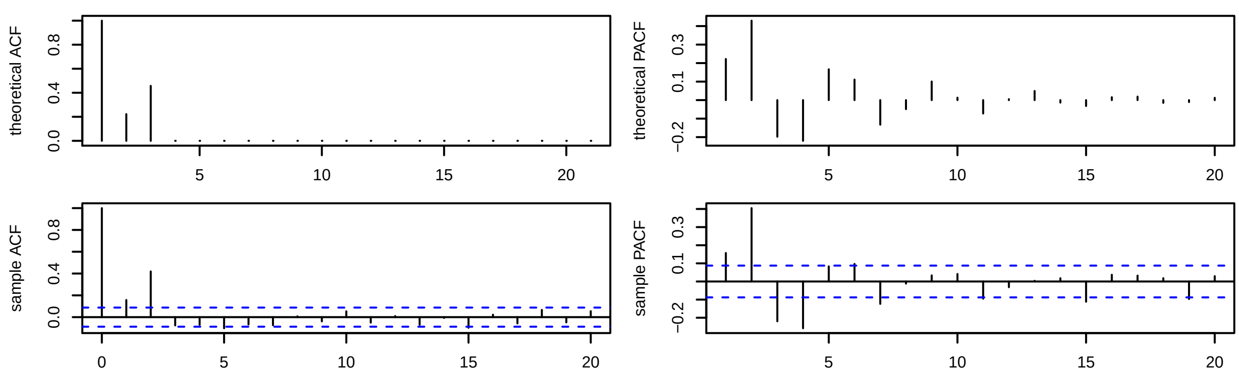
MA(1) with $\theta_1 = 0.7$



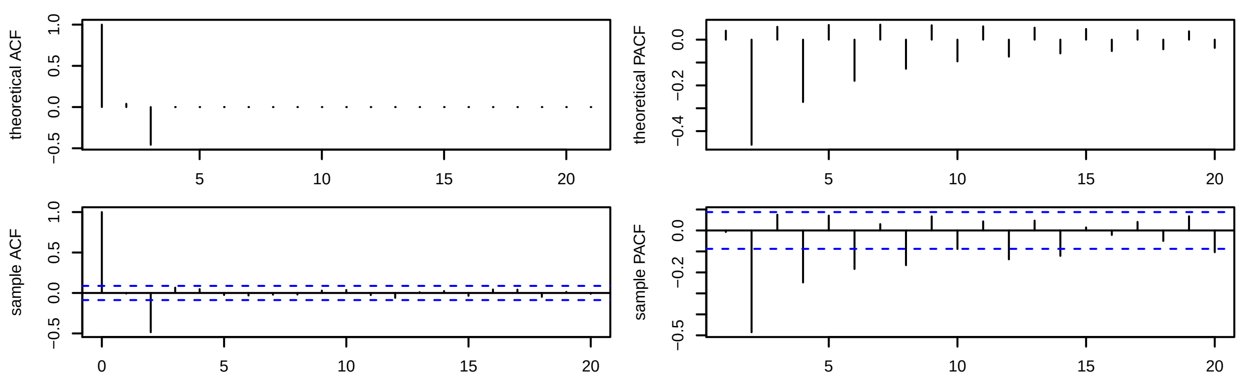
MA(1) with $\theta_1 = -0.7$



MA(2) with $\theta_1 = 0.2$, $\theta_2 = 0.7$



MA(2) with $\theta_1 = 0.2$, $\theta_1 = -0.7$



ARMA(p,q) model

AR or MA models may require a high-order model and thus many parameters to adequately describe the dynamic structure of the data. Autoregressive moving-average (ARMA) models allow to overcome this and allow parsimonious model specification with a small number of parameters.

Suppose that $\{a_t\}$ is a white noise. Time series process $\{y_t\}$ follows an ARMA(1,1) if

$$y_t = \phi_0 + \phi_1 y_{t-1} + a_t + \theta_1 a_{t-1}$$

or equivalently, using the backshift operator if

$$(1 - \phi_1 B)y_t = \phi_0 + (1 + \theta_1 B)a_t$$

More generally, time series process $\{y_t\}$ follows an ARMA(p,q) if

$$y_t = \phi_0 + \sum_{i=1}^{p} \phi_i y_{t-i} + a_t + \sum_{i=i}^{q} \theta_q a_{t-q}$$

or, using the backshift operator $(1 - \phi_1 B - \ldots - \phi_p B^p)y_t = \phi_0 + (1 + \theta_1 B + \ldots + \theta_q B^q)a_t$

Autocorrelation function for ARMA(p,q) model

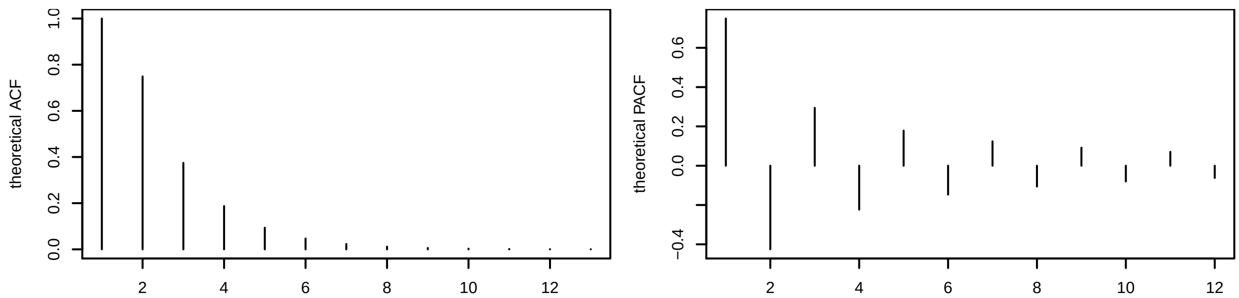
ACF and PACF are useful tools to determine whether we should consider an AR, or an MA model and what the order of that model should be:
- for AR(p): ACF dies out slowly, PACF drops to zero suddenly after lag p
- for MA(q): ACF drops to zero immediately after lag q, PACF dies out slowly

If neither ACF nor PACF drop to zero abruptly we are dealing with and ARMA model. In this case both ACF and PACF die out slowly in exponential, oscilating exponential of dampened sine wave pattern.

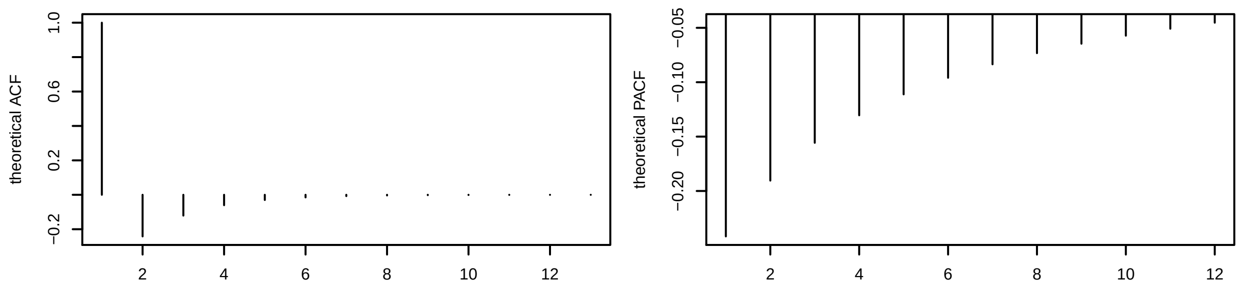
process		ACF	PACF
white noise		$\rho_l = 0$ for all $l > 0$	$\phi_{l,l} = 0$ for all l
AR(1)	$\phi_1 > 0$	exponential decay, $\rho_l = \phi_1^l$	$\phi_{l,l} = \phi_1$, $\phi_{l,l} = 0$ for $l > 1$
	$\phi_1 < 0$	oscillating decay, $\rho_l = \phi_1^l$	$\phi_{l,l} = \phi_1$, $\phi_{l,l} = 0$ for $l > 1$
AR(2)	$\phi_1^2 + 4\phi_2 > 0$	mixture of two exponential decays	$\phi_{1,1} \neq 0$, $\phi_{2,2} \neq 0$, $\phi_{l,l} = 0$ for $l > 2$
	$\phi_1^2 + 4\phi_2 < 0$	dampened since wave	$\phi_{1,1} \neq 0$, $\phi_{2,2} < 0$, $\phi_{l,l} = 0$ for $l > 2$
AR(p)		decays toward zero, may oscillate or have a shape of a dampened sine wave	$\phi_{l,l} = 0$ for $l > p$
MA(1)	$\theta_1 > 0$	$\rho_1 > 0$, $\rho_l = 0$ for all $l > 1$	oscillating decay, $\phi_{1,1} > 0$, $\phi_{2,2} < 0$, ...
	$\theta_1 < 0$	$\rho_1 < 0$, $\rho_l = 0$ for all $l > 1$	exponential decay, $phi_{l,l} < 0$ for all l
ARMA(1,1)	$\phi_1 > 0$, $\theta_1 > 0$	exponential decay	oscilating exponential decay
	$\phi_1 > 0$, $\theta_1 < 0$	exponential decay after lag 1	exponential decay
	$\phi_1 < 0$, $\theta_1 > 0$	oscillating exponential decay	oscillating exponential decay
	$\phi_1 < 0$, $\theta_1 < 0$	oscillating exponential decay	exponential decay
ARMA(p,q)		decay (direct or oscillatory) after lag p or dampened since wave	decay (direct or oscillatory) after lag q or dampened since wave

An overview of ACF and PACF for simulated ARMA(p,q) models can be found here.

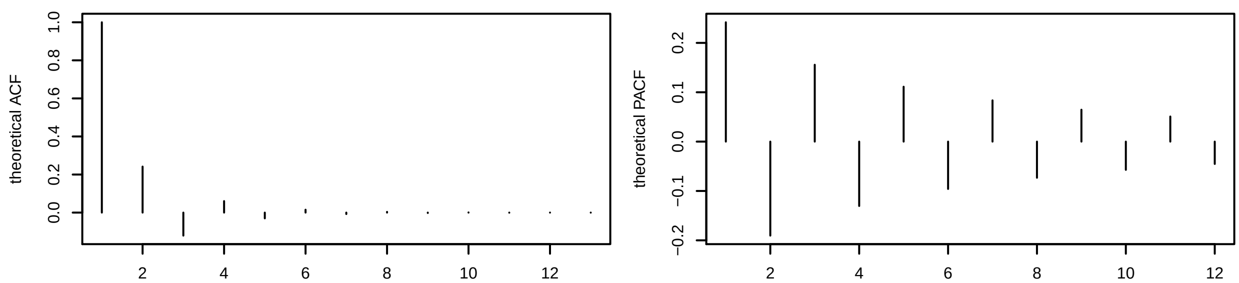
ARMA(1,1) with $\phi_1 = 0.5$, $\theta_1 = 0.9$



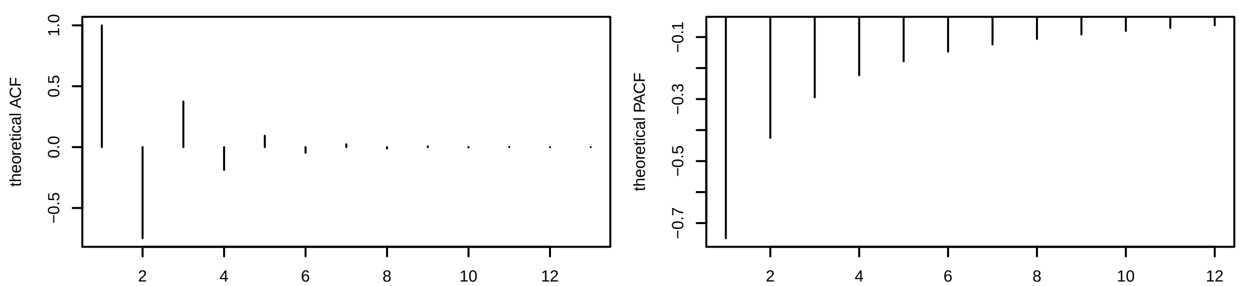
ARMA(1,1) with $\phi_1 = 0.5$, $\theta_1 = -0.9$



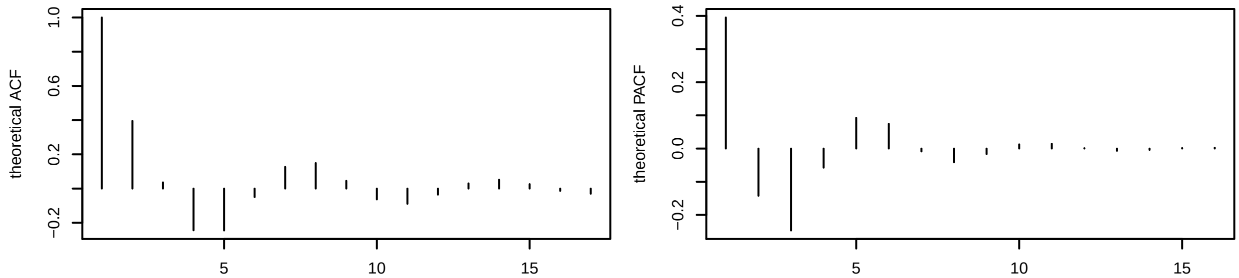
ARMA(1,1) with $\phi_1 = -0.5$, $\theta_1 = 0.9$



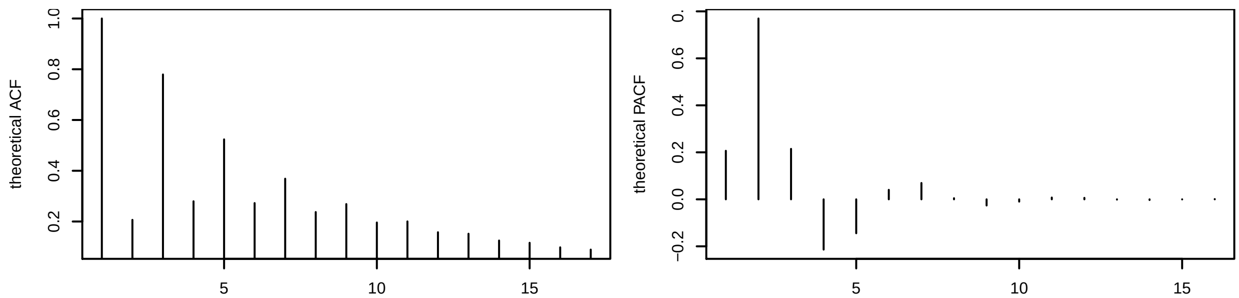
ARMA(1,1) with $\phi_1 = -0.5$, $\theta_1 = -0.9$



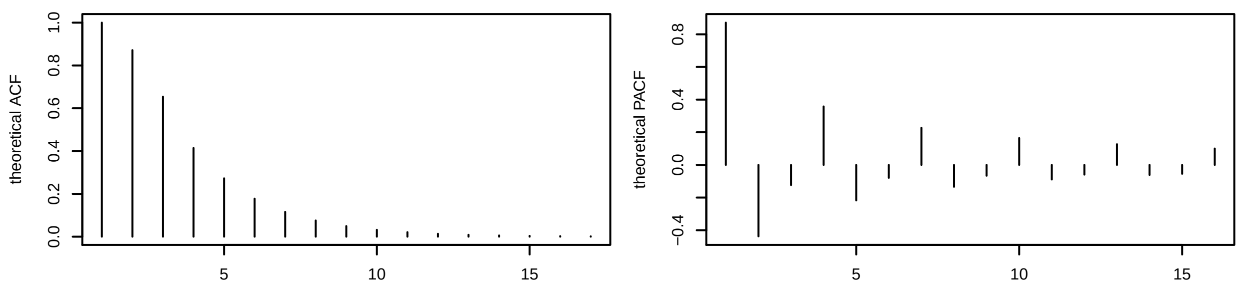
ARMA(2,2) with $\phi = (0.9, -0.7)'$, $\theta_1 = (-0.5, 0.5)'$



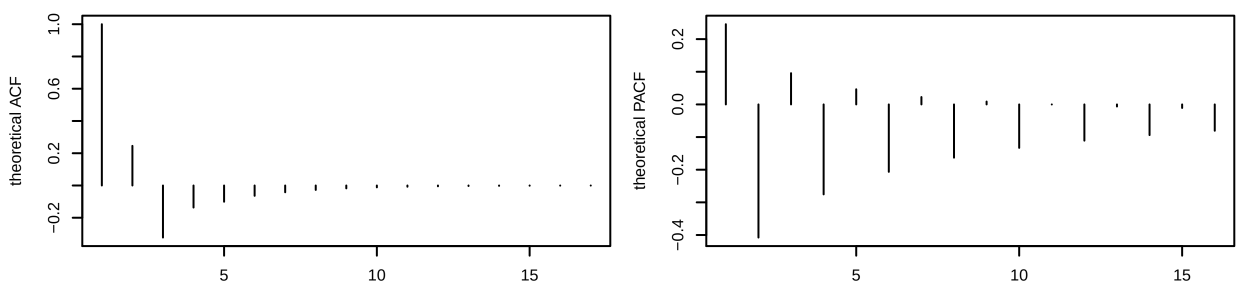
ARMA(2,2) with $\phi = (0.2, 0.6)'$, $\theta_1 = (-0.3, -0.4)'$



ARMA(2,2) with $\phi = (0.5, 0.1)'$, $\theta_1 = (0.9, 0.9)'$



ARMA(2,2) with $\phi = (0.5, 0.1)'$, $\theta_1 = (0.1, -0.9)'$



Portmanteau Test

To test $H_0 : \rho_1 = \ldots = \rho_m = 0$ againgst an alternative hypothesis $H_a : \rho_j \neq 0$ for some $j \in \{1, \ldots, m\}$ the follwoing two statistics can be used:

Box-Pierce

$$Q^*(m) = T \sum_{l=1}^{m} \hat{\rho}_l^2$$

Ljung-Box test

$$Q(m) = T(T+2) \sum_{l=1}^{m} \frac{\hat{\rho}_l^2}{T-l}$$

and the null hypothesis is rejected at $\alpha\%$ level if the above statistics are larger than the $100(1-\alpha)$th percentile of chi-squared distribution with m degrees of freedom. Note: Ljung-Box statistics tends to perform better in smaller samples. In general the recommendation is to use $m \approx \ln T$, but this depends on application; for example for monthly data with a seasonal pattern it makes sense to set m to 12, 24 or 36, and for quarterly data with a seasonal pattern m to 4, 8, 12.

Checking Adequacy of an estimated model - if the model is correctly specified Ljung-Box $Q(m)$ statistics follows chi-squared distribution with $m - g$ degress of freedom where g is the number of estimated parameters (e.g. for ARMA(p, q) that includes a constant $g = p + q + 1$)

Information Criteria

In practice, there will be often several competing models that would be considered. If these models are adequate and with very similar properties based on ACF, PACF, and Q statistics for residuals, information criteria can help decide which one is preferred. Information Criteria combine the goodness of fit with a penalty for using more parameters.

The two commonly used criteria are:

Akaike Information Criterion (AIC)

$$AIC = -\frac{2}{T} \log L + \frac{2}{T} n$$

Schwarz-Bayesian information criterion (BIC)

$$BIC = -\frac{2}{T} \log L + \frac{\log T}{T} n$$

in both expressions above T is the sample size, n is the number of parameters in the model, L is the value of the likelihood function, and log is the natural logarithm.

The AIC or BIC of competing models can be compared and the model that has the smallest AIC or BIC value is preferred. BIC will always select a more parsimonious model with fewer parameters than the AIC because $\log T > 2$ and each additional parameter is thus penalized more heavily. BIC performs better than AIC in large samples - it is asymptotically consistent while AIC is biased toward selecting an overparameterized model. In small samples AIC can perform better than BIC.

An examle showing the steps of estimating and checking a model for the growth rate of GNP can be found here: lec03GNP.zip.