Eco 5316 Time Series Econometrics

Lecture 2 Autoregressive (AR) processes

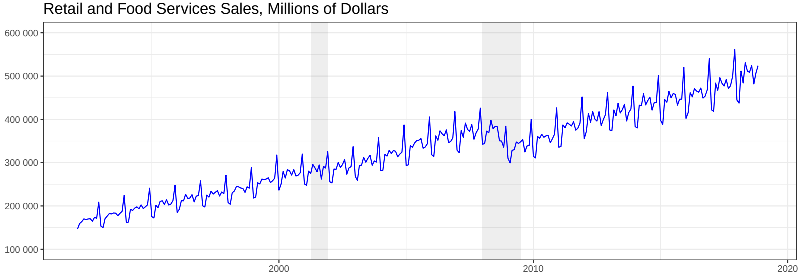
Outline

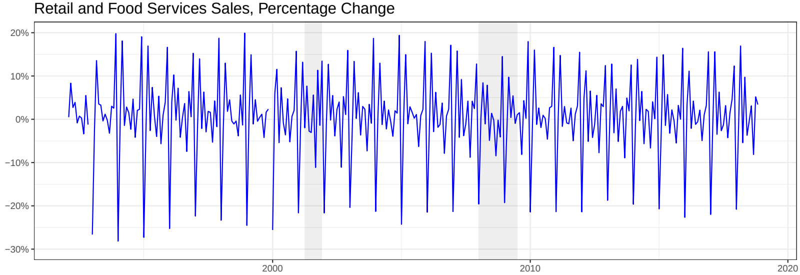
1. Features of Time Series
2. Box-Jenkins methodology
3. Autoregressive Model AR(p)
4. Autocorrelation Function (ACF)
5. Partial Autocorrelation Function (PACF)
6. Portmanteau Test - Box-Pierce test and Ljung-Box test
7. Information Criteria - Akaike (AIC) and Schwarz-Bayesian (BIC)
8. Example: AR model for Real GNP growth rate

Trend, Seasonality, Structural Change, Volatility, Outliers

- **trend** is a tendency of the time series to either grow or decline over the long term
- **seasonality** refers to regular patterns arising in economic activity due to calendar (on quarterly, monthly, day of week basis)
- **cycles** refer to patterns where the data rises and falls that are not of fixed period/duration (so while seasonal pattern has constant length cyclic pattern has variable length)
- timing of peaks and troughs is predictable with seasonal data, but unpredictable in the long term with cyclic data

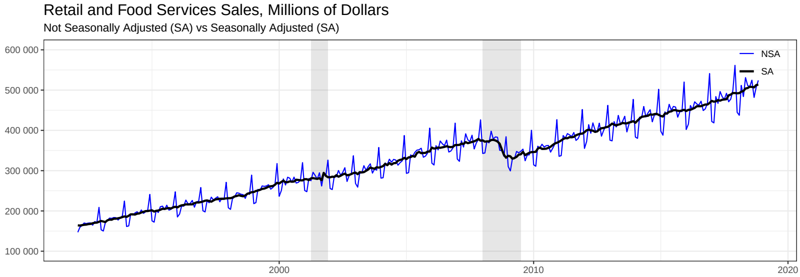
Trend, Seasonality, Structural Change, Volatility, Outliers

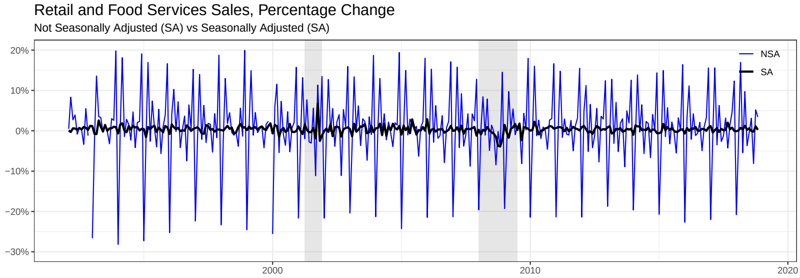




https://fred.stlouisfed.org/graph/?g=mHDh

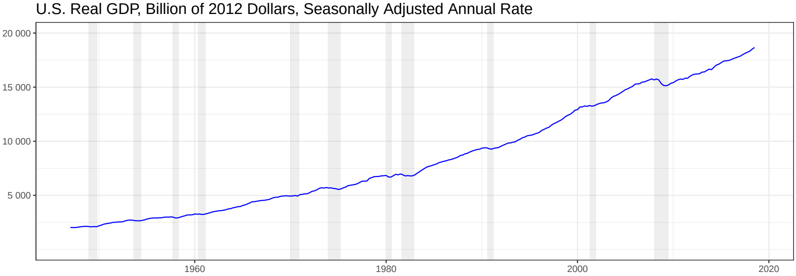
Trend, Seasonality, Structural Change, Volatility, Outliers

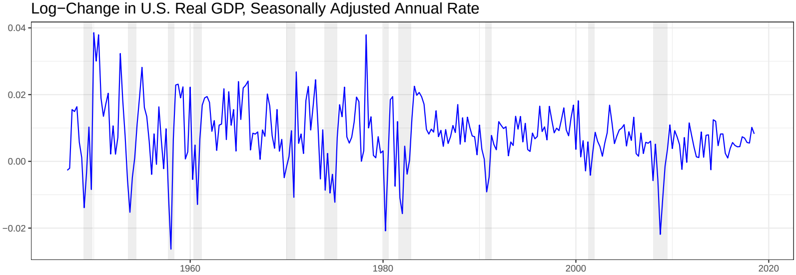




https://fred.stlouisfed.org/graph/?g=mHDh

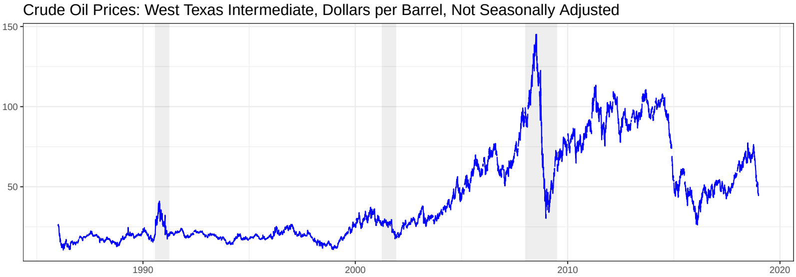
Trend, Seasonality, Structural Change, Volatility, Outliers

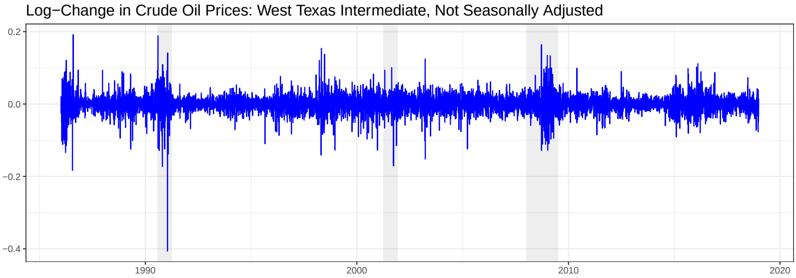




https://research.stlouisfed.org/fred2/series/GDPC1

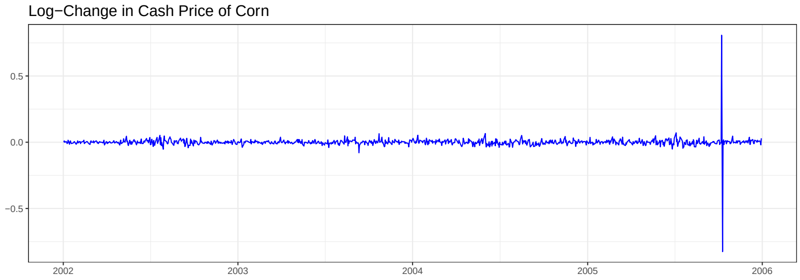
Trend, Seasonality, Structural Change, Volatility, Outliers





https://research.stlouisfed.org/fred2/series/DCOILWTICO

Trend, Seasonality, Structural Change, Volatility, Outliers



https://www.quandl.com/data/TFGRAIN/CORN

Trend, Seasonality, Structural Change, Volatility, Outliers

- decomposition of time series into trend, seasonal and irregular component

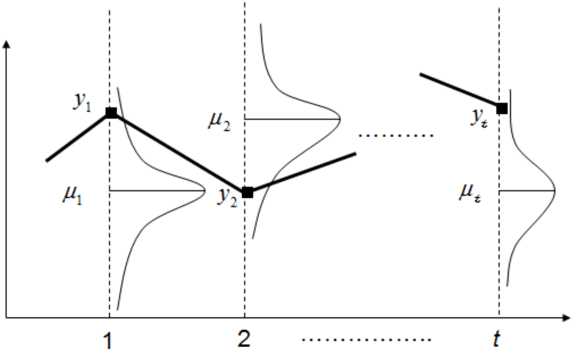
$$y_t = \mu_t + \gamma_t + \varepsilon_t$$

where

y_t is the observed data
μ_t is an slowly changing component (trend)
γ_t is periodic seasonal component
ε_t is irregular disturbance component

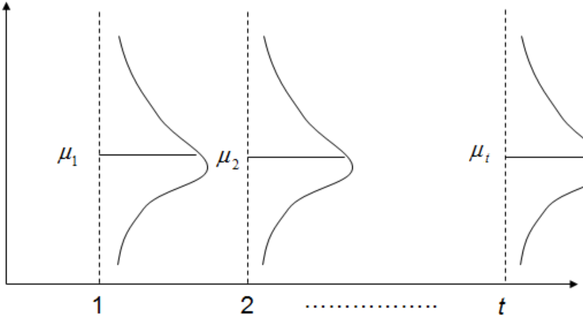
- classical approach - treat trend and seasonal components as deterministic functions
- modern approach - μ_t, γ_t, ε_t all contain stochastic components
- we will first look at the ways how to model the irregular component, and leave seasonal and trend components for later

Def: Stochastic process (or time series process) is a sequence of random variables. Observed time series is a particular realization of this process.



Preliminaries

Def: Stochastic process $\{y_t\}$ is **strictly stationary** if joint distributions $F(y_{t_1}, \ldots, y_{t_k})$ and $F(y_{t_1+l}, \ldots, y_{t_k+l})$ are identical for all l, k and all $t_1, \ldots, t_k$

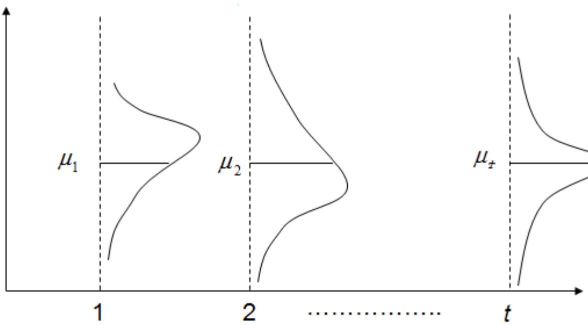


Preliminaries

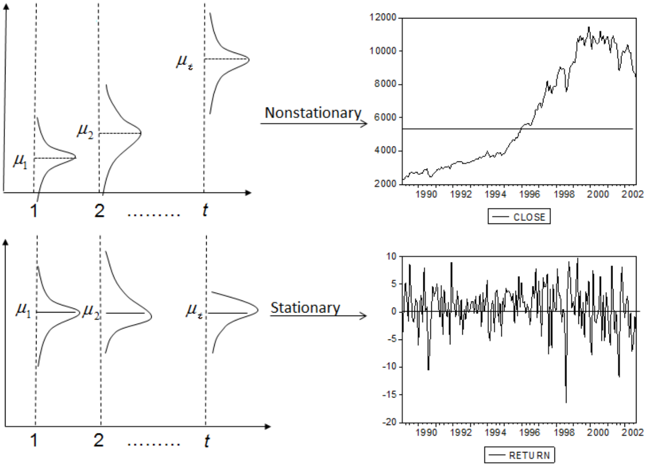
Def: Stochastic process $\{y_t\}$ is (second order) **weakly stationary** if
(i) $E(y_t) = \mu$ for all t
(ii) $cov(y_t, y_{t-l}) = \gamma_l$ for all t, l

Note: if (i) is satisfied but (ii) the process is first order weakly stationary

Note: for $l = 0$ we get that $var(y_t) = cov(y_t, y_t) = \gamma_0$ for all t, which means that variance is constant over time



Preliminaries



Preliminaries

- weak stationarity allows us to use sample moments to estimate population moments
- for example, given a weakly stationary time series $\{y_1, y_2, \ldots, y_t\}$ the first moment $E(y_t)$ can be estimated using $\frac{1}{t}\sum_{j=1}^{t} y_j$
- for nonstationary process $\frac{1}{t}\sum_{j=1}^{t} y_j$ is not a useful estimator, since $E(y_1) \neq E(y_2) \neq \ldots \neq E(y_t)$

Def: Stochastic process $\{\varepsilon_t\}$ is called a **white noise** if ε_t are independently identically distributed with zero mean and finite variance: $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma_\varepsilon^2 < \infty$, $cov(\varepsilon_t, \varepsilon_s) = 0$ for all $t \neq s$.

Box-Jenkins Methodology

Box-Jenkins methodology to modelling weakly stationary time series

1. Identification
2. Estimation
3. Checking Model Adequacy

Box-Jenkins Methodology

1. **Indentification**

- examine **time series plots** of the data to determine if any transformations are necessary (differencing, logarithms) to get weakly stationary time series, examine series for trend (linear/nonlinear), periods of higher volatility, seasonal patterns, structural breaks, outliers, missing data, ...
- examine **autocorrelation function (ACF)** and **partial autocorrelation function (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**

- estimate all models considered and select the best one - coefficients should be statistically significant, **information criteria (AIC, SBC)** should be low
- model can be estimated using either **conditional likelihood method** or exact **likelihood method**

3. **Checking Model Adequacy**

- perform **in-sample evaluation** of the estimated model
  - estimated coefficients should be consistent with the underlying assumption of stationarity
  - 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 test whether autocorrelations of residuals are statistically significant
  - check model for parameter instability and structural change
- perform **out-of-sample evaluation** of the model forecast

Box-Jenkins Methodology

- we will now look at how the Box-Jenkins methodology works in case of a simple univariate time series model - an autoregressive model

AR(p) Model

- simple linear regression model with cross sectional data

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i$$

- suppose we are dealing with time series rather than cross sectional data, so that

$$y_t = \beta_0 + \beta_1 x_t + \epsilon_t$$

  and if the explanatory variable is the lagged dependent variable $x_t = y_{t-1}$ we get

$$y_t = \beta_0 + \beta_1 y_{t-1} + \epsilon_t$$

- main idea: past is prologue as it determines the present, which in turn sets the stage for future

AR(p) Model

- hourly time series for Akkoro Kamui's activities, before the fortress was built

$$\{y_1, y_2, \ldots, y_t\} = \{drink, drink, \ldots, drink\}$$

- lots of time dependence here:

$$y_t = y_{t-1}$$

AR(p) Model

- time series process $\{y_t\}$ follows autoregressive model of order 1, AR(1), if

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

or equivalently, using the lag operator

$$(1 - \phi_1 L) y_t = \phi_0 + \varepsilon_t$$

where $\{\varepsilon_t\}$ is a white noise with $E(\varepsilon_t) = 0$ and $Var(\varepsilon_t) = \sigma_\varepsilon^2$

- more generally, time series $\{y_t\}$ follows an autoregressive model of order p, AR(p), if

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

or equivalently, using the lag operator

$$(1 - \phi_1 L - \ldots - \phi_p L^p) y_t = \phi_0 + \varepsilon_t$$

AR(p) Model

tools to determined the order p of the autoregressive model given $\{y_t\}$

- Autocorrelation Function (ACF)
- Partial Autocorrelation Function (PACF)
- Portmanteau Test - Box-Pierce test and Ljung-Box test
- Information Criteria - Akaike (AIC) and Schwarz-Bayesian (BIC)

Autocorrelation Function (ACF)

- linear dependence between y_t and y_{t-l} is given by correlation coefficient ρ_l
- for a weakly stationary time series process$\{y_t\}$ we have

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

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

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

  where $\bar{y} = \frac{1}{T}\sum_{t=1}^{T} y_t$
- **sample autocorrelation function** is $\{\hat{\rho}_1, \hat{\rho}_2, \ldots\}$

Autocorrelation function for AR(p) model

- if $p = 1$ then $\gamma_0 = Var(y_t) = \frac{\sigma_\varepsilon^2}{1-\phi_1^2}$ and also $\gamma_l = \phi_1 \gamma_{l-1}$ for $l > 0$, thus

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

  and since $\rho_0 = 1$, we get $\rho_l = \phi_1^l$
- for weakly stationary $\{y_t\}$ it has to hold that $|\phi_1| < 1$, theoretical ACF of a stationary AR(1) thus decays exponentially, in either direct or oscillating way

Autocorrelation function for AR(p) model

- if $p = 2$ theoretical ACF for AR(2) satisfies second order difference equation

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

  or equivalently using the lag operator $(1 - \phi_1 L - \phi_2 L^2)\rho_l = 0$

- solutions of the associated **characteristic equation**

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

  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 $D = \phi_1^2 + 4\phi_2 > 0$ then ω_1, ω_2 are real numbers, and theoretical ACF is a combination of two exponential decays
- if $D < 0$ characteristic roots are complex conjugates, and theoretical 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 $i = 1, 2$
- 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 $l \geq 2$

Autocorrelation function for AR(p) model

- in general, theoretical ACF for AR(p) satisfies the difference equation of order p
$$(1-\phi_1 L-\ldots-\phi_p L^p)\rho_l = 0 \tag{3}$$
- characteristic equation of the AR(p) model is thus $1-\phi_1 x-\ldots-\phi_p x^p = 0$
- 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
- plot of the theoretical 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

$$y_t = \phi_{0,1} + \phi_{1,1} y_{t-1} + e_{1,t} \quad (4)$$

$$y_t = \phi_{0,2} + \phi_{1,2} y_{t-1} + \phi_{2,2} y_{t-2} + e_{2,t} \quad (5)$$

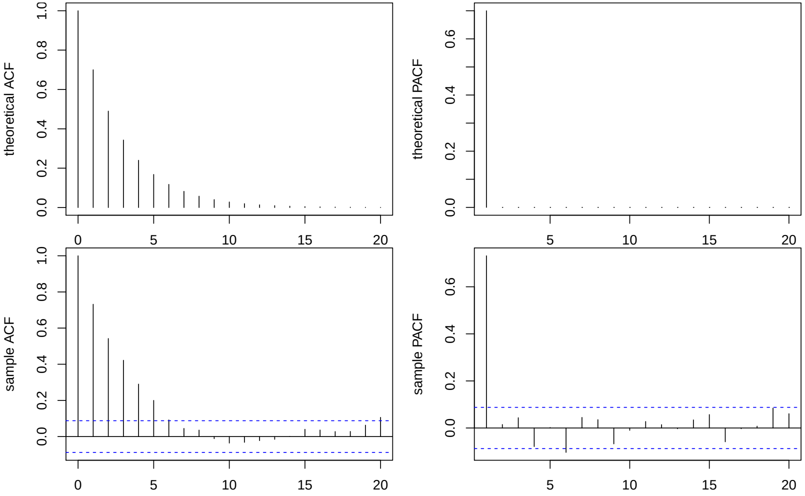
$$y_t = \phi_{0,3} + \phi_{1,3} y_{t-1} + \phi_{2,3} y_{t-2} + \phi_{3,3} y_{t-3} + e_{3,t} \quad (6)$$

$$\vdots \quad (7)$$

- 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 asymptotic 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 $0 \pm 2/\sqrt{T}$
- order of the AR process can thus be determined by finding the lag after which PACF cuts off to zero

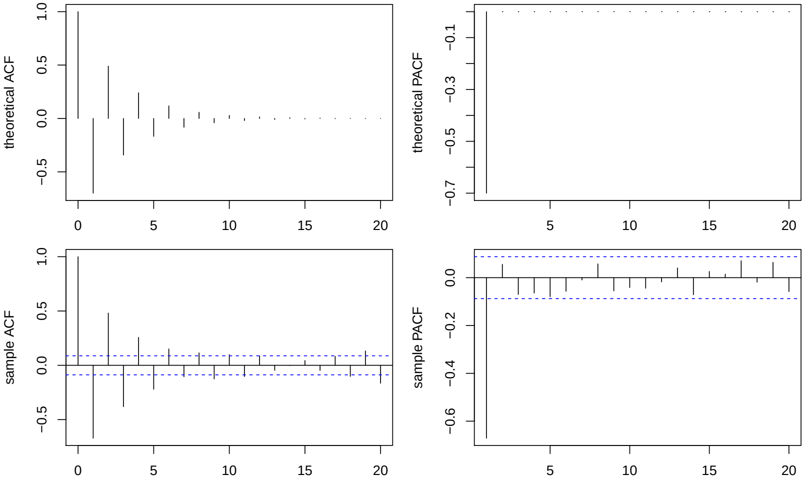
ACF and PACF for AR(1) model

AR(1) with $\phi_1 = 0.7$



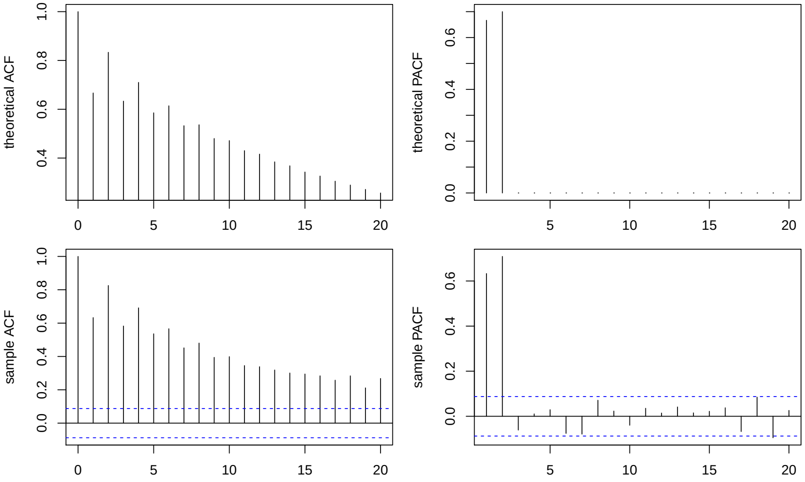
ACF and PACF for AR(1) model

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



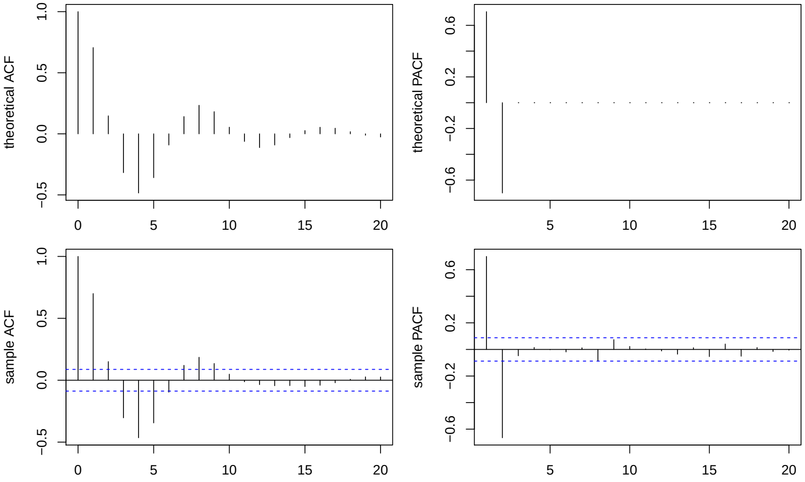
ACF and PACF for AR(2) model

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



ACF and PACF for AR(2) model

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



ACF and PACF for AR(p) model

- interactive overview of ACF and PACF for simulated AR(p) models is here

Portmanteau Test

- to test $H_0 : \rho_1 = \ldots = \rho_m = 0$ against an alternative hypothesis $H_a : \rho_j \neq 0$ for some $j \in \{1, \ldots, m\}$ following two statistics can be used:

  Box-Pierce test

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

  Ljung-Box test

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

- the null hypothesis is rejected at α% 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
- the general recommendation is to use $m \approx \ln T$, but this depends on application
- e.g.: 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

Portmanteau Test

- these tests are also used for in-sample evaluation of model adequacy
- if the model was correctly specified Ljung-Box $Q(m)$ statistics for the residuals of the estimated model follows chi-squared distribution with $m-g$ degrees of freedom where g is the number of estimated parameters
- for AR(p) that includes a constant $g = p+1$

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
- main idea: information criteria combine the goodness of fit with a penalty for using more parameters

Information Criteria

- two commonly used information criteria:

  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

- 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

Information Criteria

- fundamental difference - AIC tries to select the model that most adequately approximates unknown complex data generating process with infinite number of parameters
- this true process is never in the set of candidate models that are being considered
- BIC assumes that the true model is among the set of considered candidates and tries to identify it
  - 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

Information Criteria

- some software packages report other information criteria in addition to AIC and BIC
- **Hannan-Quinn information criterion (HQ)**

$$HQ = -\frac{2}{T}\log L + \frac{2\log(\log T)}{T}n$$

- **corrected Akaike Information Criterion (AICc)** which is AIC with a correction for finite sample sizes to limit overfitting; for a univariate linear model with normal residuals

$$AICc = AIC + \frac{2(n+1)(n+2)}{T-n-2}$$

where T is the sample size and n is the number of estimated parameters

Example: AR model for Real GNP growth rate

```
# load magrittr package (pipe operators)
library(magrittr)
```

```
# import the data on the growth rate of GDP, convert it into time series xts object
y <- scan(file = "http://faculty.chicagobooth.edu/ruey.tsay/teaching/fts3/q-gnp4791.txt") %>%
       ts(start = c(1947,2), frequency = 4)
```

```
str(y)
```

```
##  Time-Series [1:176] from 1947 to 1991: 0.00632 0.00366 0.01202 0.00627 0.01761 ...
```

```
head(y)
```

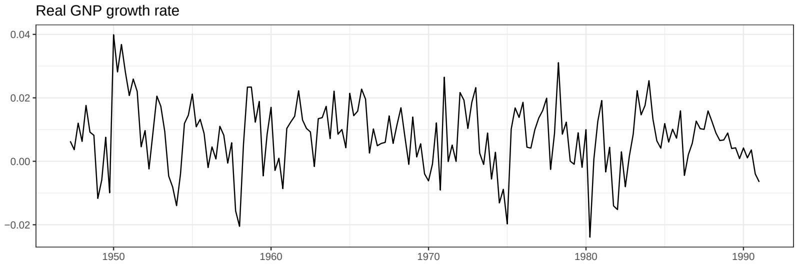
```
## [1] 0.00632 0.00366 0.01202 0.00627 0.01761 0.00918
```

```
tail(y)
```

```
## [1]  0.00085  0.00420  0.00108  0.00358 -0.00399 -0.00650
```

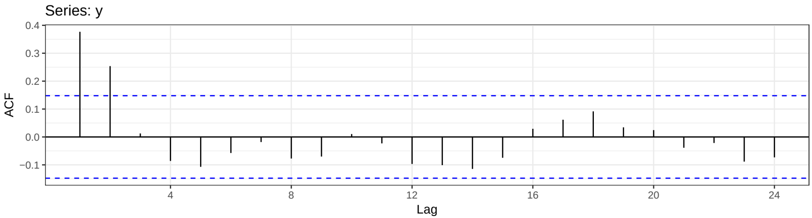
Example: AR model for Real GNP growth rate

```
# load ggplot2, ggfortify and forecast packages
library(ggplot2)
library(ggfortify)
library(forecast)
# define default theme to be B&W
theme_set(theme_bw())
# plot
autoplot(y) +
    labs(x = "", y = "", title = "Real GNP growth rate")
```

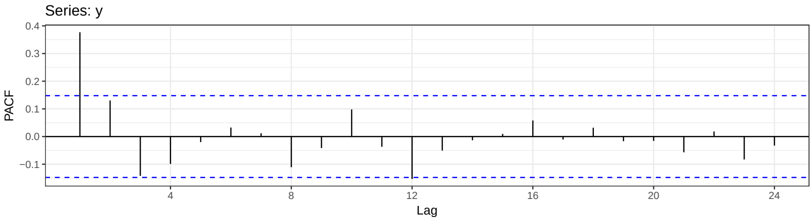


Example: AR model for Real GNP growth rate

```r
# plot ACF and PACF for y up to lag 24
ggAcf(y, lag.max = 24)
```




```r
ggPacf(y, lag.max = 24)
```



Example: AR model for Real GNP growth rate

```
# estimate an AR(1) model - there is only one significant coefficient in the PACF plot for y
m1 <- Arima(y, order = c(1,0,0))
# show the structure of object m1
str(m1)
```

```
## List of 18
##  $ coef     : Named num [1:2] 0.37865 0.00769
##   ..- attr(*, "names")= chr [1:2] "ar1" "intercept"
##  $ sigma2   : num 9.91e-05
##  $ var.coef : num [1:2, 1:2] 4.88e-03 -1.12e-06 -1.12e-06 1.44e-06
##   ..- attr(*, "dimnames")=List of 2
##   .. ..$ : chr [1:2] "ar1" "intercept"
##   .. ..$ : chr [1:2] "ar1" "intercept"
##  $ mask     : logi [1:2] TRUE TRUE
##  $ loglik   : num 562
##  $ aic      : num -1119
##  $ arma     : int [1:7] 1 0 0 0 4 0 0
##  $ residuals: Time-Series [1:176] from 1947 to 1991: -0.00126 -0.00351 0.00586 -0.00306 0.01046 ...
##  $ call     : language Arima(y = y, order = c(1, 0, 0))
##  $ series   : chr "y"
##  $ code     : int 0
##  $ n.cond   : int 0
##  $ nobs     : int 176
##  $ model    :List of 10
##   ..$ phi  : num 0.379
##   ..$ theta: num(0)
##   ..$ Delta: num(0)
##   ..$ Z    : num 1
##   ..$ a    : num -0.0142
##   ..$ P    : num [1, 1] 0
##   ..$ T    : num [1, 1] 0.379
##   ..$ V    : num [1, 1] 1
##   ..$ h    : num 0
##   ..$ Pn   : num [1, 1] 1
##  $ aicc     : num -1119
##  $ bic      : num -1109
##  $ x        : Time-Series [1:176] from 1947 to 1991: 0.00632 0.00366 0.01202 0.00627 0.01761 ...
##  $ fitted   : Time-Series [1:176] from 1947 to 1991: 0.00758 0.00717 0.00616 0.00933 0.00715 ...
##  - attr(*, "class")= chr [1:2] "ARIMA" "Arima"
```

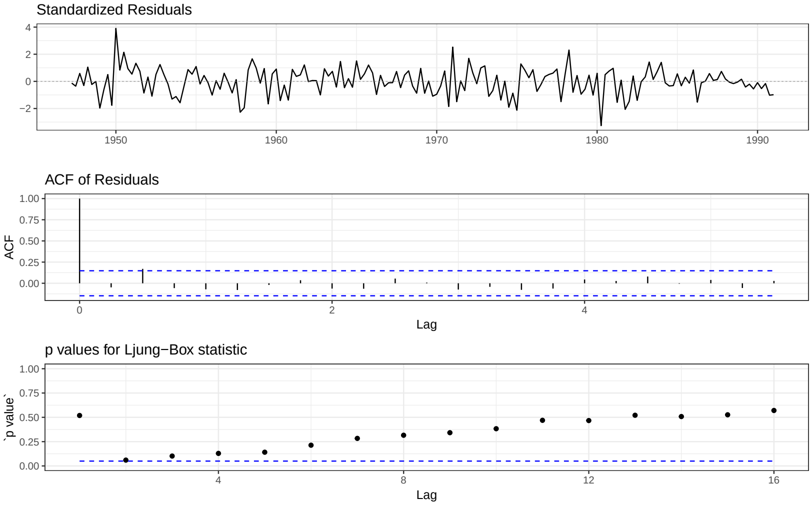
Example: AR model for Real GNP growth rate

```r
# print out results for m1
m1
```

```
## Series: y
## ARIMA(1,0,0) with non-zero mean
##
## Coefficients:
##          ar1    mean
##       0.3787  0.0077
## s.e.  0.0698  0.0012
##
## sigma^2 estimated as 9.913e-05:  log likelihood=562.47
## AIC=-1118.94   AICc=-1118.8   BIC=-1109.43
```

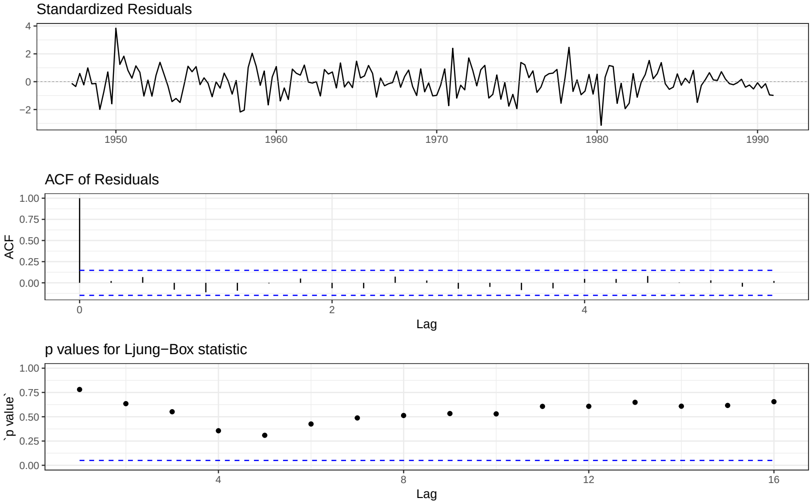
Example: AR model for Real GNP growth rate

```
# diagnostics for AR(1) model - there seems to be a problem with remaining serial correlation at lag 2
ggtsdiag(m1, gof.lag = 16)
```



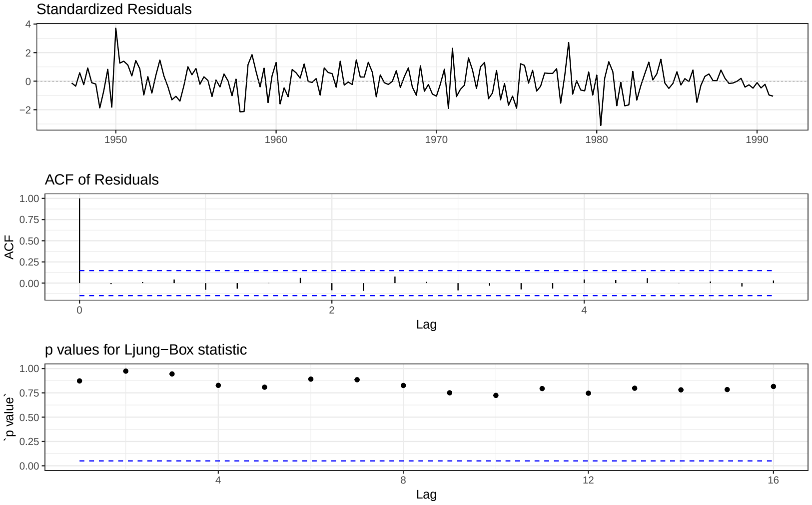
Example: AR model for Real GNP growth rate

```r
# estimate an AR(2) model to deal with the problem of remaining serial correlation at lag 2
m2 <- Arima(y, order = c(2,0,0))
# diagnostics for AR(2) model shows that problem with remaining serial correlation at lag 2 is gone
ggtsdiag(m2, gof.lag = 16)
```



Example: AR model for Real GNP growth rate

```r
# estimate an AR(3) model since PACF for lag 2 and 3 are comparable in size
m3 <- Arima(y, order = c(3,0,0))
# diagnostics for the AR(3) model
ggtsdiag(m3, gof.lag = 16)
```



Example: AR model for Real GNP growth rate

```
# Ljung-Box test - for residuals of a model adjust the degrees of freedom m
# by subtracting the number of parameters g
# this adjustment will not make a big difference if m is large but matters if m is small

m2.LB.lag8 <- Box.test(m2$residuals, lag = 8, type = "Ljung")
m2.LB.lag8
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 7.2222, df = 8, p-value = 0.5129
```

```
1-pchisq(m2.LB.lag8$statistic, df = 6)
```

```
## X-squared
## 0.3007889
```

```
m2.LB.lag12 <- Box.test(m2$residuals, lag = 12, type = "Ljung")
m2.LB.lag12
```

```
##
##  Box-Ljung test
##
## data:  m2$residuals
## X-squared = 10.098, df = 12, p-value = 0.6074
```

```
1-pchisq(m2.LB.lag12$statistic, df = 10)
```

```
## X-squared
## 0.4319577
```