A lecture on time series

Bernt Arne Ødegaard

24 November 2021

Contents

1 Survey of time series issues

In this part we discuss some of the typical problems we run into when discussing time series. We start by giving an overview of the issues and methods.

1.1 Definition

A time series is a sequence of observations of economic variables,

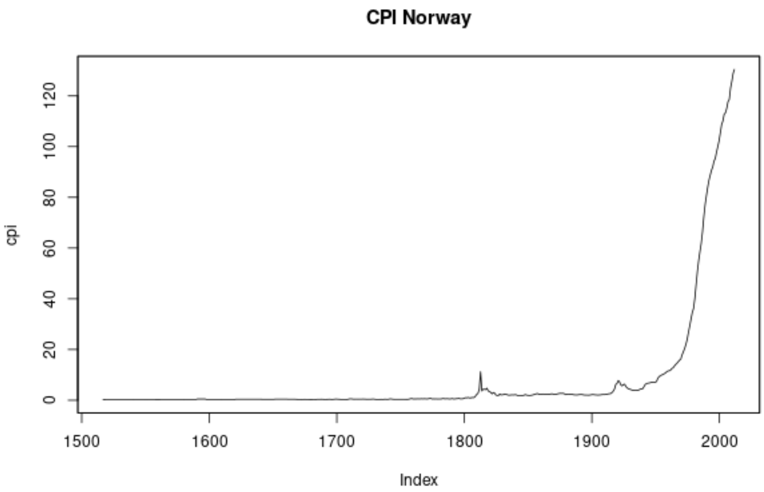
$$x_1, x_2 \dots x_t \dots x_T$$

usually sampled at fixed intervals. Typical intervals: Daily, weekly, monthly, quarterly, annual.

1.2 Examples

Example: Consumer Price Index. What is the relative value of one unit of currency. Figure 1 show the evolution of the CPI for Norway.

Figure 1 An example: Norwegian Consumer Price Index.



Another example: Stock market index. What is the total value of an investment in the stock market. Figure **??** shows the evolution of a stock market index for the Norwegian Stock market.
Time series can be either in the form of stocks and flows. Stocks measures the value of something at a specific time, a flow measures something between two time periods.
The price index is an example of a stock.
An example of a flow is the inflation during a period. This can be found by taking the "first differences" of the price indices. Figure 3 illustrates the time series of **monthly** inflation for Norway.
It is very hard to interpret this time series, so we will often look at transformations. An example is to look at **annual** observations.

1.3 Typical Problems in Time Series

1. Forecasting.
2. Trend Removal.
3. Seasonal Adjustment.
4. Detection of a Structural Break.

Figure 2 An example: Stock Market Index (SP500).

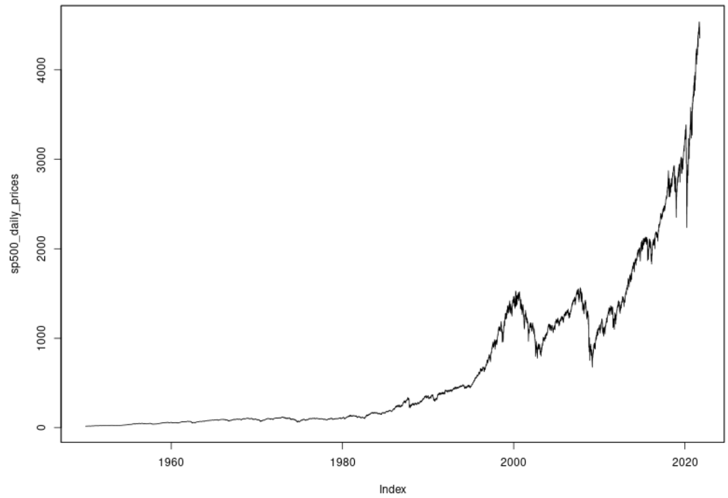


Figure 3 An example: Norwegian Consumer Price Index. First Differences

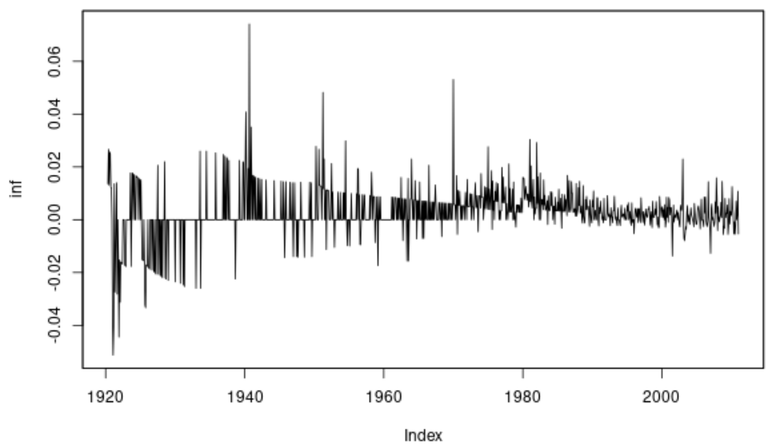
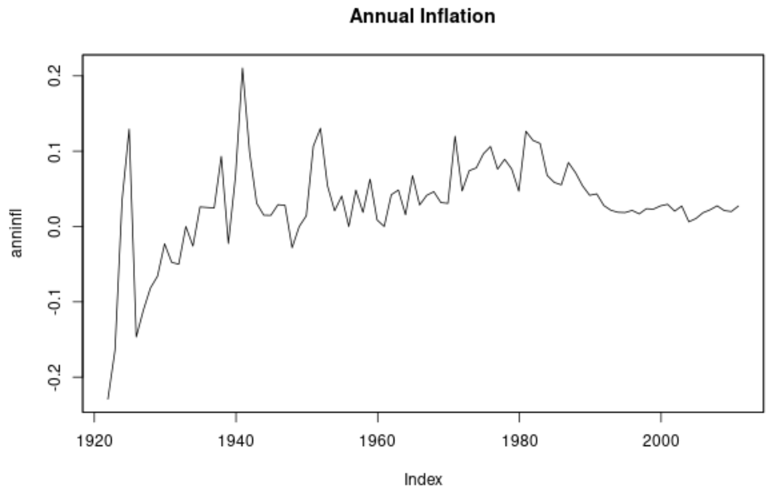


Figure 4 An example: Norwegian Consumer Price Index. First Differences of annual data



5. Causality–time lags.
6. Distinction between the short and long run
7. Study of agent's expectations.

1.3.1 Forecasting

Evaluation of *future* values x_{T+h} as a function of current and past values $x_1, x_2 \dots x_T$.

$$\hat{y}_{T+h} = f(y_1, y_2, \dots y_T)$$

Concerned about the quality of the forecast, depend on to what the degree the stochastic process governing x_t is "regular" as a function of time.
Also depend on the forecasting horizon. Usually get better forecasts for small values of h (short horizon).

1.3.2 Trend Removal.

If the series have a clear trend (see the examples of price indices and stock indices), may want to remove the trend and analyze the movements *around* the trend.

1.3.3 Seasonal Adjustment.

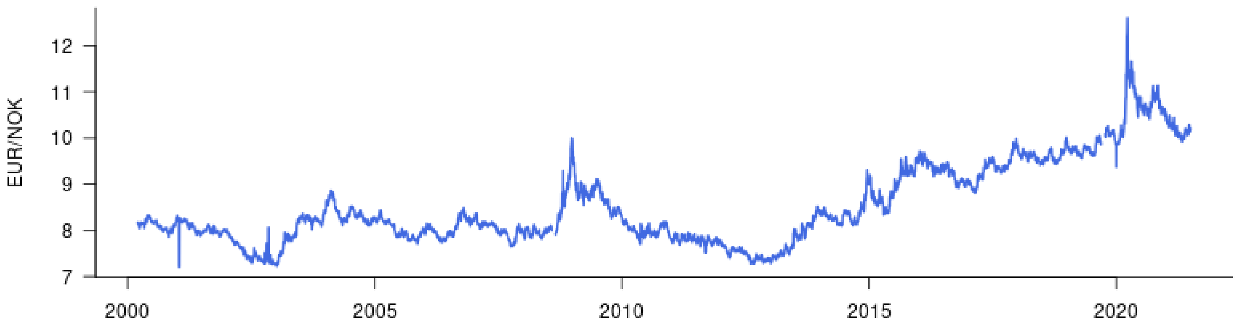
Many variable of economic interest will have strong seasonal components. For example heating cost will clearly be low in the summer and high in the winter. Depending on purpose may want to remove seasonal components.

1.3.4 Detection of a Structural Break.

One of the most important questions we face is determining whether the economic environment has changed *substantially*, such that we have moved into a new *regime*, which necessitates re–evaluation of the economic relationships

As an example, consider figure 5, which shows the currency exchange rate between NOK and EUR. Can we claim that the change in the time series is important enough such that this is a *structural break*, or is this just a random evolution of a stationary process.

Figure 5 NOK/EUR exchange rate



1.3.5 Causality–time lags.

Another economic question with obvious policy implications is the question of whether and how one economic variable influences another.

1.3.6 Distinction between the short and long run

To what degree are relationships *persistent*?

1.3.7 Study of agent's expectations.

If we have data on agent's forecasts, how accurate are these? Can agents forecast at all? Are expectations rational?

1.4 Categories of Time series Modelling

There are some important distinctions between the groups of tools used for time series modelling.

1. Adjustment models.
2. Autopredictive models.
3. Explanatory models.

1.4.1 Adjustment models.

Roughly, mechanical models for removing seasonal components and trends.

1.4.2 Autopredictive models.

Intuitively: "Model–free" forecasting models.

$$x_t = f(x_{t-1}, x_{t-2}, \ldots) + e_t$$

We work with simple models which merely *describe* the data. There is no theory which *explains* why a particular formulation describes the data, but for forecasting purposes such a "theoryless" formulation may often do better than a fully specified economic model.

1.4.3 Some Examples

Let y_t be the data of interest, and ϵ_t be white noise ($E[\epsilon_t] = 0,\ E[\epsilon_t^2 = \sigma_\epsilon^2],\ E[\epsilon_t \epsilon_s] = 0 \ \forall \ s \neq t$ The simplest time series specification is an AR(1) process, an *autoregressive* model of order 1:

$$y_t = \mu + \gamma y_{t-1} + \epsilon_t$$

More generally, can write autoregressive specifications with p terms, AR(p):

$$y_t = \mu + \sum_{k=1}^{p} \gamma y_{t-k} + \epsilon_t$$

An alternative specification is a *Moving Average* process (MA(1)):

$$y_t = \mu + \epsilon_t - \theta \epsilon_{t-1}$$

Again, we can add terms to produce a MA(q) process:

$$y_t = \mu + \epsilon_t - \sum_{k=1}^{q} \theta_k \epsilon_{t-k}$$

A generalization of these models is the autoregressive moving average (ARMA(p,q)) model

$$y_t = \mu + \sum_{k=1}^{p} \gamma y_{t-k} + \epsilon_t + \epsilon_t - \sum_{k=1}^{q} \theta_k \epsilon_{t-k}$$

I will not go into details about these models, but they are implemented in any statistics package you can think of, and they are important to know about for applied econometric work.

1.4.4 The *Lag* operator

When reading the literature about these types of models, one confusing item is the heavy use of the *lag* operator. Let me just give you a few notes about this piece of notation.
Suppose our time series is y_t.
The lag operator L just shifts the time index one period back

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

The reason it is so useful written this way is that we can do algebra using the lag operator.
For example

$$L(Ly_t) = L(y_{t-1}) = y_{t-2}$$

This will be written

$$L^2 y_t$$

Think of the lag operator as a way of transforming time series following the rules of multiplication and addition.
With the lag notation we can very compactly express complicated time series models.
For example

$$x_t = (aL + bL^2)x_t = ax_{t-1} + bx_{t-2}$$

1.4.5 Unit Root tests

What is a unit root?
Consider the autoregressive relation

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

where ρ is a constant and ϵ_t an error term.
The parameter ρ is the parameter of interest. Note that it has implications for the type of process on y:

- If $\rho > 1$ then y is an explosive process.
- If $\rho = 1$ then y is a random walk.
- If $\rho < 1$ then y is a stationary process.

It is the case of $\rho = 1$ which is the borderline case. In this case ($\rho = 1$) the process y is said to have a unit root. Another terminology is to say that y is *integrated of order 1*.
When we test for an unit root, we want to test whether $\rho = 1$, because that has implications for how to estimate y. If $\rho = 1$, then we want to do estimation using $y_t - y_{t-1}$, (take first differences), since that is a stationary process, otherwise, if y is not integrated, we want to do inference on y directly.
Unit Root tests are designed to test the hypothesis $\rho = 1$, but they do not have a simple normal distribution, they are much more complicated.

1.4.6 Vector Autoregressions (VAR's)

We looked earlier at an example of testing for Granger causality in a model with two variables x and y:

$$y_t = c + \alpha_1 y_{t-1} + \alpha_2 y_{t-1} + \beta_1 x_{t-1} \beta_2 x_{t-2} + \epsilon_t$$

This is a simple example of an auto-regression, the current value of a variable is a (linear) function of past values of itself and other variables.
Already one of the characteristics of this type of estimation should be clear, it is "model-free" in the sense that we are not estimating an explicit economic model, we are exploring possible linear relationships between economic variables. A prime use of these investigations will be forecasting. It is also useful to generate "stylized facts" that our economic theories should be able to explain. But it should be clear that a VAR or some other time series model does not *explain* any causality or other regularities in the data.

1.5 Explanatory models.

Roughly: “Economic” models of relationships between variables.

$$y_t = f(x_t; b) + e_t$$

where y_t are endogenous variables and x_t are exogenous variables. b are parameters and e_t are random errors.

If the exogenous varibles x_t are only observed at time t and the errors e_t are independent of observations at other times, this is a *static* model.

But the model can be dynamic, both when

- The exogenous variables x_t include past (lagged) values of y, $y_{t-1}, y_{t-2}, \ldots$.
- The errors e_t depend on errors or variables at other times.

What we have covered so far in the course is of this type of model.

The typical issues we run into when it is time series data we are analyzing in a regression framework.

Consider a regression model of the standard type

$$y_t = X_t b + e_t$$

where y_t is the outcome of the dependent variable at time t, X_t the independent variables at time t, e_t the error term at time t, and b the parameters of interest. The independent variables X_t at time t may also include past values of y_t as well as other variables observed at earlier dates.

We then observe outcomes of y_t and X_t for a number of periods $t = 1, \ldots T$.

With this setup, the typical problem is that the error term e_t will not be independent. Think of a shock to the economy, like the recent stock market gyrations. The effects of this shock may very often persist for a time, affecting the error terms over this time in the same direction.

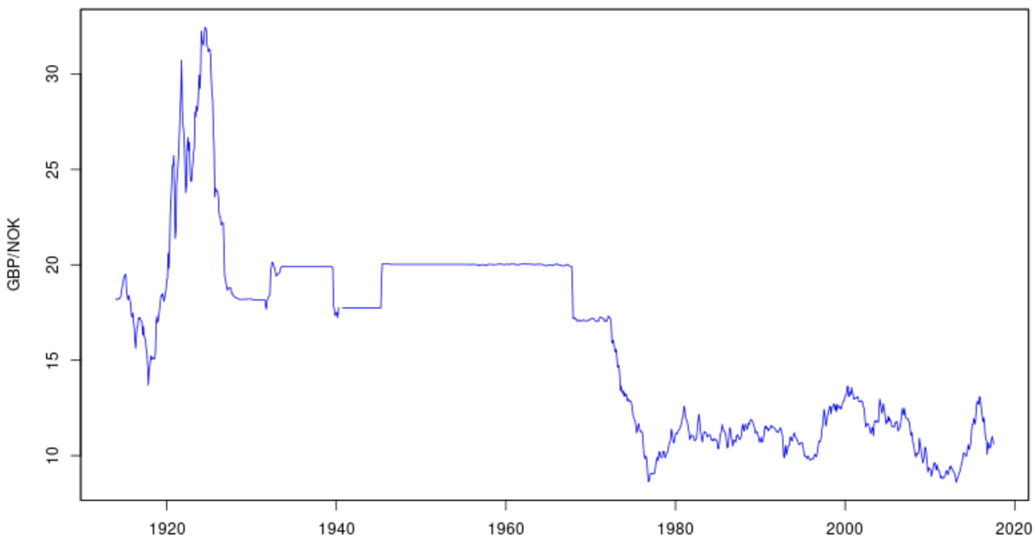
Because of the ordering that time imposes on the observations, we expect that errors that are “close” (in time) may well be correlated, but that errors that are “distant” there will have little correlation.

This imposes some structure on the form of the error covariance matrix.

The knowledge that we are dealing with a time series allow us to make assumptions about the *form* of the covariance matrix of the error terms.

2 Plotting

First thing to do with any time series you are using: Plot it to see if there are some special features. This is always a good idea, one need to both "get a feel for the data" and spot potential data problems, even though there is always the danger of seeing patterns that may not be there.



3 Univariate time series

The above is a motivation, let us now go for the abstract univariate time series formulation

$$x_t = f(x_{t-1}, \varepsilon_{t-1}, \dots)$$

When we write

$$x_t = f(x_{t-1}, \varepsilon_{t-1}, \dots)$$

we move to an atheoretical world, where we work with standard methods for modelling the time series relationships without explicitly thinking about an economic model which has generated these relationships. We will later map these "reduced form" specifications to actual economic models (typically called structural models).
Definition: Univariate time series modelling of process $\{x_t\}$. Modelling and predicting x_{t+1} as a function of past values $x_t, x_{t-1}, \cdots$ and past and current errors $u_t, u_{t-1}, \cdots$ and u_{t+1}.

3.0.1 Stationarity

First question is: Is a series *stationary*?
This is the first question that needs to be answered. The modelling is very different depending on whether a series is stationary or not.
Roughly, stationarity means that the stochastic relationship between one observations and the next does not depend on when the observation is made.
Stationarity can be formally defined in various ways.
Strict stationarity
Let x_{t_i} be the observation at time t_i. A time series is *strictly stationary* if the joint distribution of

$$x_{t_1}, x_{t_2}, x_{t_3}, \dots x_{t_n}$$

is the same as the distribution of

$$x_{t_1+k}, x_{t_2+k}, x_{t_3+k}, \dots x_{t_n+k}$$

for all possible n and k.
A weaker concept is called *weak stationarity*, usually defined in terms of first and second moments of $\{x_t\}$.
Define these moments
Mean:

$$E[y_t] = \mu \quad \forall t$$

Variance

$$E[(y_t - \mu)^2] = \sigma^2 < \infty \quad \forall t$$

Autocovariance

$$E\left[(y_{t_1} - \mu)(y_{t_2} - \mu)\right] = \gamma_{t_2 - t_1} < \infty \quad \forall t_1, t_2$$

A process satisfying these assumptions is said to be the weakly stationary, or covariance stationary.

Autocovariance function

$$\gamma_s = E\left[(y_t - E[y_t])(y_{t-s} - E[y_{t-s}])\right] \quad s = 1, 2, \cdots$$

Autocorrelation function (ACF)

$$\tau_s = \frac{\gamma_s}{\gamma_0} \quad s = 0, 1, 2, \cdots$$

Partial autocorrelation function (PACF)
No simple formula, roughly the change in autocorrelation from one step to the next.

3.0.2 Empirical ACF and PACF

Exercise 1.
Collect annual estimates of the Norwegian Consumers Price Index (CPI) starting in 1516. Estimate the annual inflation as

$$Inflation = \log(CPI_t) - \log(CPI_{t-1}$$

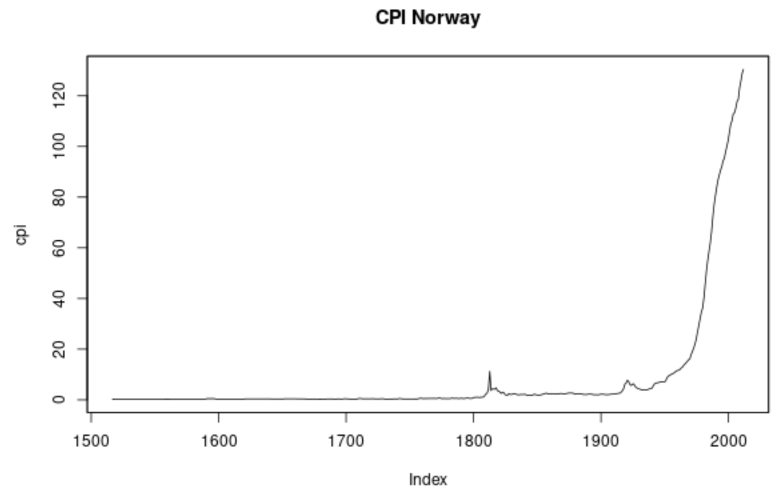
based on the CPI series. Let $q = 5$.

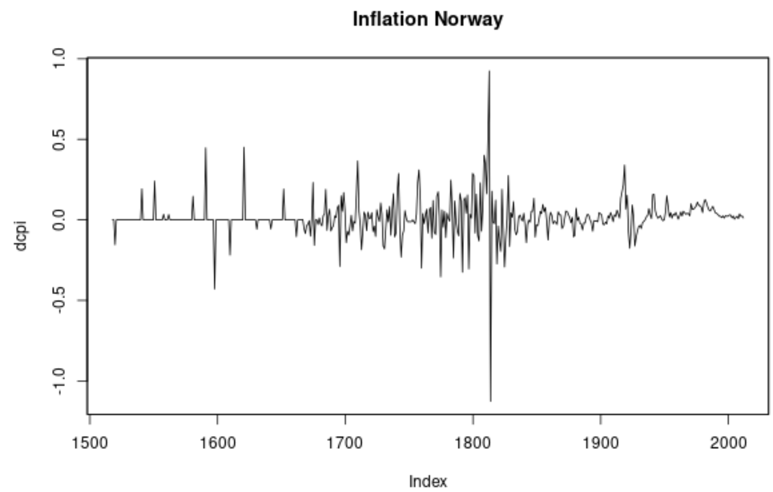
1. Plot the two series.
2. Estimate the autocorrelations (acf) of orders 1 through q for the CPI series.
3. Estimate the partial autocorrelations (pacf) of orders 1 through q for the CPI series.
4. Estimate the average inflation for the period.
5. Estimate the autocorrelations (acf) of orders 1 through q for the Inflation series.
6. Estimate the partial autocorrelations (pacf) of orders 1 through q for the Inflation series.

Solution to Exercise 1.

```
cpi <- read.zoo("../../../../data/norway/nb_historical_statistics/cpi_norway_1516_2011.csv",
                skip="2",format="%Y",header=TRUE,sep=",")
dcpi <- diff(log(cpi))
cpi <- as.matrix(cpi[,2])
dcpi <- as.matrix(dcpi[,2])
```

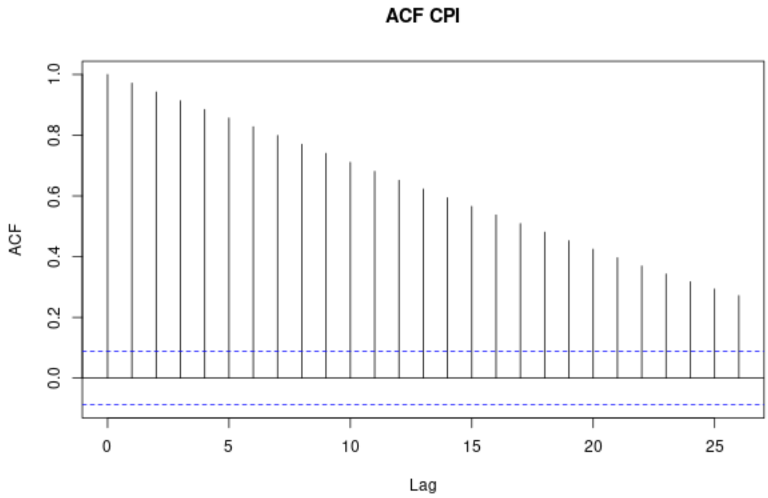
Time series plots of the series

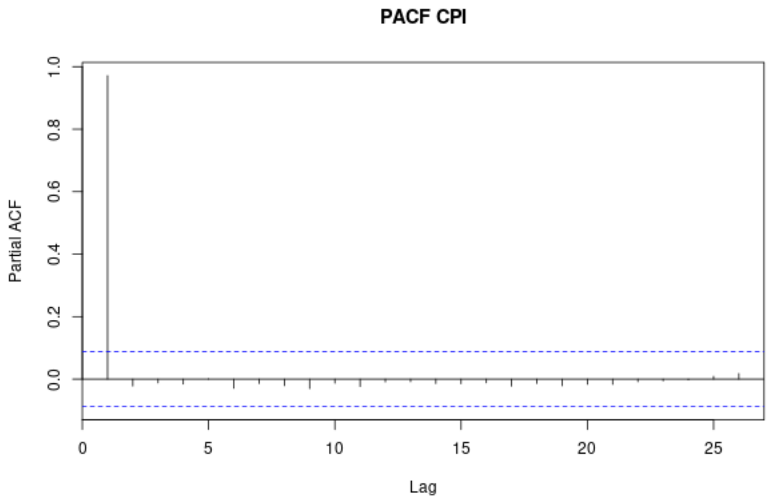




Describing the cpi series

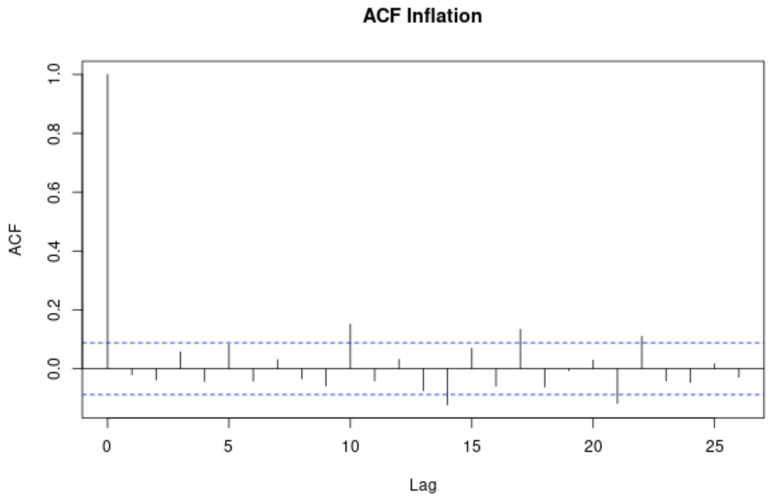
```
> acf(cpi,plot=FALSE,lag.max=5)

Autocorrelations of series ‘cpi’, by lag

    0     1     2     3     4     5 
1.000 0.971 0.942 0.914 0.885 0.857 
> pacf(cpi,plot=FALSE,lag.max=5)

Partial autocorrelations of series ‘cpi’, by lag

     1      2      3      4      5 
 0.971 -0.022 -0.012 -0.016  0.001 
```

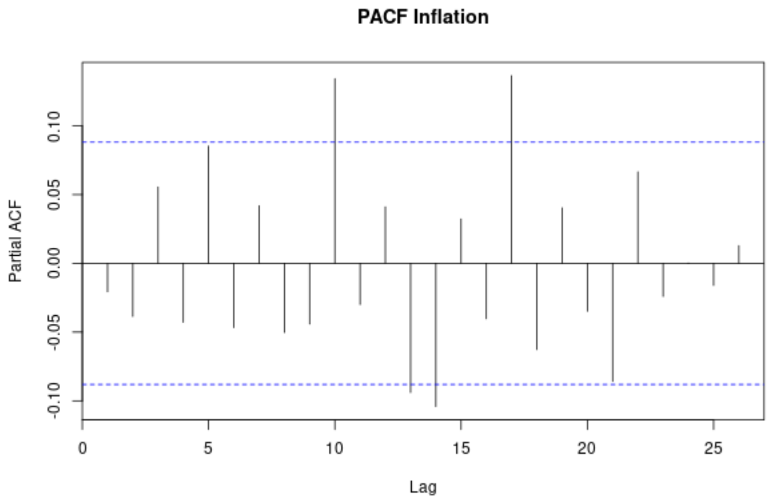




Describing the inflation series

```
> mean(dcpi)
[1] 0.01326274

> acf(dcpi,plot=FALSE,lag.max=5)

Autocorrelations of series ‘dcpi’, by lag

     0      1      2      3      4      5
 1.000 -0.021 -0.038  0.057 -0.044  0.082
> pacf(dcpi,plot=FALSE,lag.max=5)

Partial autocorrelations of series ‘dcpi’, by lag

     1      2      3      4      5
-0.021 -0.039  0.056 -0.043  0.085
```





3.0.3 White Noise Process

(Purely random process)
A process is white noise if it satisfies the following conditions

$$E[y_t] = \mu$$

$$\text{var}(y_t) = \sigma^2$$

$$\gamma_{t-r} = \begin{cases} \sigma^2 & \text{if } t = r \\ 0 & \text{otherwise} \end{cases}$$

Note that white noise is weakly stationary
The White Noise process is the natural null when testing for autocorrelation.
We typically assume one special case of white noise.
Gaussian White Noise

$$y \sim_{iid} N(\mu, \sigma^2)$$

with this as a null we can construct tests concerning autocorrelation.

3.1 Testing for autocorrelation

1) Test whether $\hat{\tau}_l = 0$ for one particular lag l.
A 95% confidence interval

$$\left[-1.96\frac{1}{\sqrt{T}}, +1.96\frac{1}{\sqrt{T}}\right]$$

under the null of a Gaussian White Noise process.
2) Test whether all correlation coefficients

$$\tau_1, \tau_2, \cdots, \tau_q$$

are jointly zero.
Two variants
Box-Pierce Q

$$Q_m = T\sum_{k=1}^{m} \hat{\tau_k}^2$$

Ljung-Box statistic (small sample correction)

$$Q_m^* = T(T+2)\sum_{k=1}^{m} \frac{\hat{\tau_k}^2}{T-k}$$

Both statistics have asymptotic χ^2 distributions.
Exercise 2.
Collect annual estimates of the Norwegian Consumers Price Index (CPI) starting in 1516. Estimate the annual inflation as

$$Inflation = \log(CPI_t) - \log(CPI_{t-1}$$

based on the CPI series.

1. Use the Ljung-Box Q statistic to test whether the first five autocorrelations are jointly zero.
2. Similary calculate the Box-Pierce Q statistic to test whether the first five autocorrelations are jointly zero.
3. What is the "best" AR model?

Solution to Exercise 2.

```
cpi <- read.zoo("../../../../data/norway/nb_historical_statistics/cpi_norway_1516_2011.csv",
                skip="2",format="%Y",header=TRUE,sep=",")
dcpi <- diff(log(cpi))
cpi <- as.matrix(cpi[,2])
dcpi <- as.matrix(dcpi[,2])
```

Calculating the test statistics
Box Pierce

```
> Box.test(dcpi,type="Box-Pierce",lag=5)
Box-Pierce test
data:  dcpi
X-squared = 6.8142, df = 5, p-value = 0.2348
```

Ljung Box

```
> Box.test(dcpi,type="Ljung-Box",lag=5)
Box-Ljung test
data:  dcpi
X-squared = 6.8967, df = 5, p-value = 0.2284
```

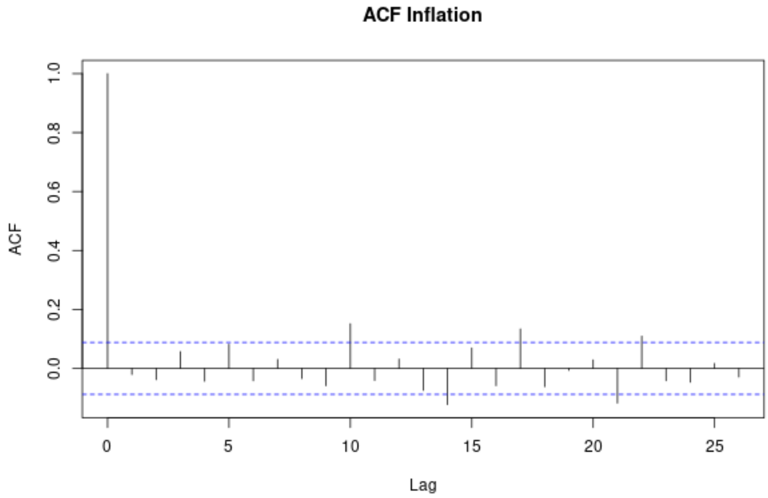
Not specifying a lag in the ar command says to choose the optimal lag length using the AIC criterion.

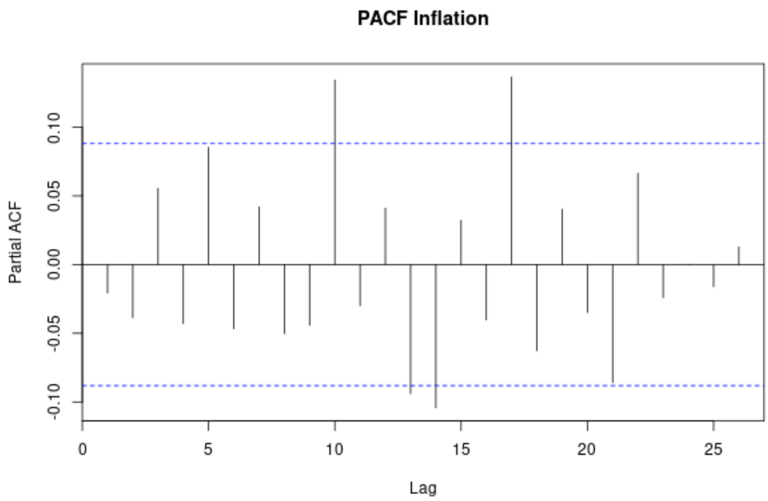
```
> ar(dcpi,AIC=true)
Call:
ar(x = dcpi, AIC = true)
Coefficients:
      1        2        3        4        5        6        7        8
 0.0079  -0.0538   0.0788  -0.0109   0.0544  -0.0270   0.0066  -0.0340
      9       10       11       12       13       14       15       16
-0.0353   0.1307  -0.0198   0.0247  -0.0857  -0.1147   0.0391  -0.0406
     17
 0.1365
Order selected 17  sigma^2 estimated as  0.01315
```

So the optimal autoregressive model involves 17 autoregressive terms.
However, when one tries to do this with slightly lower number of autoregressive terms, get some surprising effects. Try specying the ar command with `order.max` specifiction, which says the maximal lag length is `order.max`.

```
> ar(dcpi,AIC=true,order.max=5)
Call:
ar(x = dcpi, order.max = 5, AIC = true)

Order selected 0  sigma^2 estimated as  0.01382
> ar(dcpi,AIC=true,order.max=10)

Call:
ar(x = dcpi, order.max = 10, AIC = true)

Coefficients:
      1        2        3        4        5        6        7        8
-0.0016  -0.0429   0.0595  -0.0382   0.0758  -0.0399   0.0306  -0.0439
      9       10
-0.0432   0.1343

Order selected 10  sigma^2 estimated as  0.01354
> ar(dcpi,AIC=true,order.max=15)

Call:
ar(x = dcpi, order.max = 15, AIC = true)

Coefficients:
      1        2        3        4        5        6        7        8
-0.0022  -0.0482   0.0692  -0.0252   0.0644  -0.0371   0.0253  -0.0372
      9       10       11       12       13       14
-0.0429   0.1376  -0.0271   0.0360  -0.0932  -0.1042

Order selected 14  sigma^2 estimated as  0.01335
```

If the maximal lag length is five, the model selected has zero autoregressive terms. Once the max is above 10, the order increases. One can figure out why by looking at the ACF and PACF plots.





Observe that the estimated AR coeffisients for lags below five are small in magnitude. Longer lag lengths have higher estimated AR coefficients.

3.2 Moving Average Processes

MA(q)

$$y_t = \mu + \sum_{i=1}^{q} \theta u_{t-i} + u_t$$

q'th order moving average, where

$$u_t \sim_{iid} (0, \sigma^2)$$

Linear combination of White Noise processes.
Properties of a moving average MA(q) process

$$E[y_t] = \mu$$

$$\text{var}(y_t) = \gamma_0 = (1 + \theta_1^2 + \theta_2^2 + \theta_3^2 + \cdots + \theta_q^2)\sigma^2$$

$$\gamma_s = \begin{cases} (\theta_s + \theta_{s+1}\theta_1 + \theta_{s+2}\theta_2 + \cdots + \theta_q\theta_{q-s})\sigma^2 & \text{if } s \leq q \\ 0 & \text{if } s > q \end{cases}$$

MA process:

- Constant mean
- Constant variance
- Autocovariances zero after q lags.

3.3 The Lag operator

Useful piece of notation, the lag operator L. (Also called the backshift operator B.)

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

$$L^k y_t = y_{t-k}$$

3.4 Autoregressive processes

$$y_t = \mu + \sum_{i=1}^{p} \phi_i y_{t-i} + u_t$$

in lag operator notation

$$y_t = \mu + \sum_{i=1}^{p} \phi_i L^i y_t + u_t$$

or

$$\phi(L) = (1 - \phi_1 L - \phi_2 L^2 - \cdots - \phi_p L^p)$$

Stationarity condition for autoregressive processes. Consider

$$(1 - \phi_1 z - \phi_2 z^2 - \cdots - \phi_p z^p)$$

All solutions of this “characteristic equation” lie outside unit circle.
Example

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

Is it stationary?

$$1 - \phi_1 z = 0$$

$$z = 1$$

on the unit circle.
This is not a stationary process.

3.5 Wold decomposition theorem

Roughly: Any stationary autoregressive process of finite order p can be expressed as an infinite order Moving Average Process.

3.6 ARMA processes

$ARMA(p, q)$ Combination of a AR and MA process.

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \cdots + \phi_p y_{t-p} + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \cdots + \theta_q u_{t-q} + u_t$$

3.7 Typical ACF functions

Recognising the different suspects
$AR(p)$

- Geometrically declining ACF
- Number of non-zero PACF equals p.

$MA(q)$

- Number of Non-zero ACF equals q.
- Geometrically declining PACF

$ARMA(p, q)$.

- Geometrically declining ACF
- Geometrically declining PACF

Exercise 3.
Suppose the stochastic proces $\{y_t\}$ has the following structure

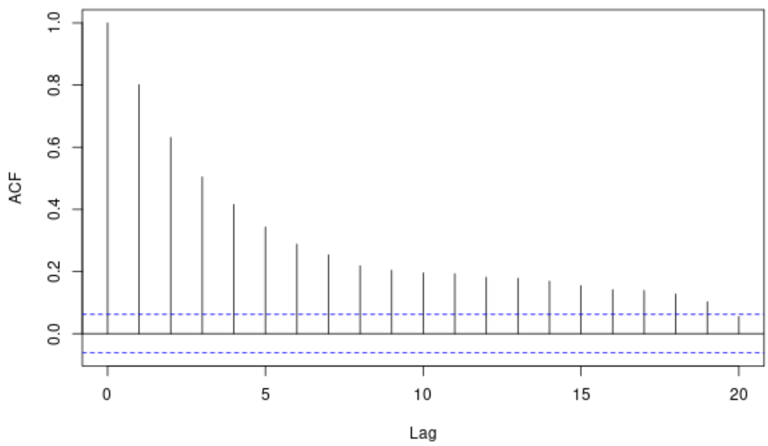
$$y_t = \rho y_{t-1} + u_t$$

where u_t is Gaussian White Noise with $\sigma_u^2 = 1$. Setting $\rho = 0.8$ and $y_0 = 0$ simulate $T = 1000$ realizations of this process, and plot the ACF and PACF of the resulting data series for lags 1-20.
Solution to Exercise 3.
Doing the simulation

```
series <- arima.sim(n=1000,list(ar=c(0.8,0),ma=c(0,0)))
```

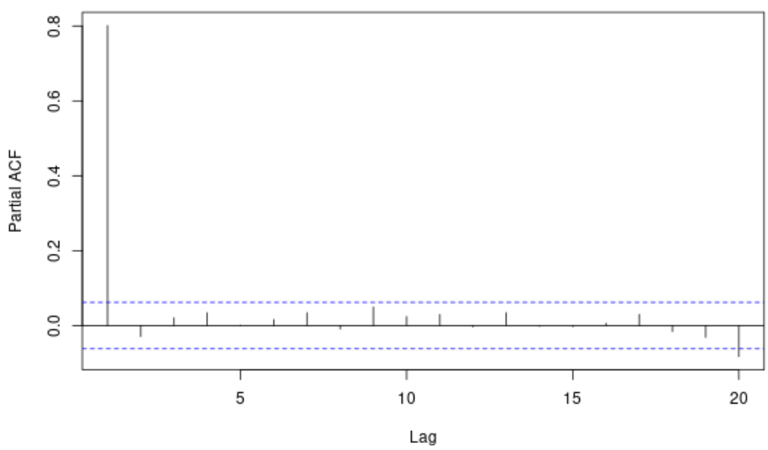
ACF for process $y_t = 0.8 y_{t-1} + u_t$

```
> acf(series,plot=FALSE,max.lag=20)

Autocorrelations of series ‘series’, by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.791  0.654  0.505  0.398  0.305  0.223  0.168  0.124  0.079  0.063
    11     12     13     14     15     16     17     18     19     20     21
 0.048  0.057  0.068  0.088  0.108  0.115  0.104  0.079  0.060  0.054  0.046
    22     23     24     25     26     27     28     29     30
 0.048  0.047  0.046  0.045  0.027  0.020 -0.007 -0.026 -0.042
```



PACF for process $y_t = 0.8y_{t-1} + u_t$

```
> pacf(series,plot=FALSE,max.lag=20)

Partial autocorrelations of series ‘series’, by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.791  0.077 -0.091  0.012 -0.013 -0.033  0.015  0.001 -0.037  0.042  0.005
    12     13     14     15     16     17     18     19     20     21     22
 0.043  0.028  0.040  0.026 -0.003 -0.034 -0.040  0.003  0.031  0.002  0.020
    23     24     25     26     27     28     29     30
 0.011  0.000  0.007 -0.042  0.003 -0.055 -0.023 -0.007
```



Exercise 4.

Suppose the stochastic proces $\{y_t\}$ has the following structure

$$y_t = \rho y_{t-1} + u_t$$

where u_t is Gaussian White Noise with $\sigma_u^2 = 1$. Setting $\rho = 0.2,\ y_0 = 0$ and $T = 1000$, simulate T realizations of this process, and plot the ACF and PACF of the resulting data series for lags 1-20.

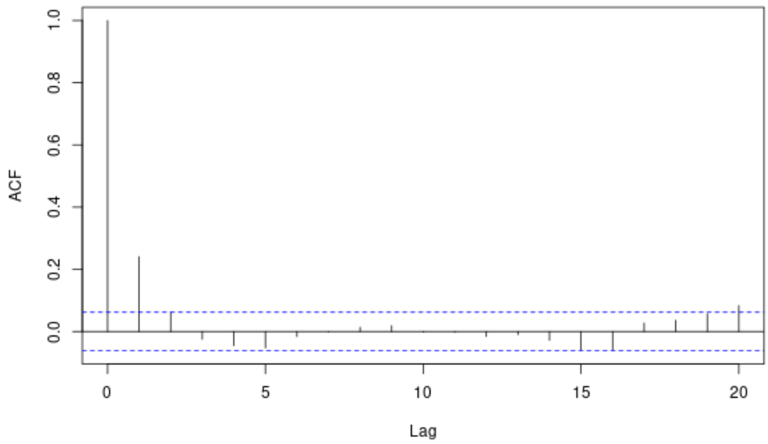
Solution to Exercise 4.

Performing the simulation

```
> series <- arima.sim(n=1000,list(ar=c(0.2,0),ma=c(0,0)))
```

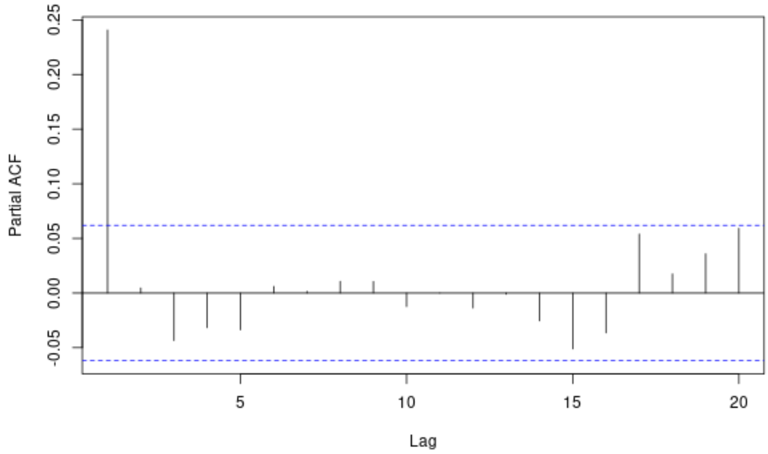
ACF for process $y_t = 0.2y_{t-1} + u_t$

```
> acf(series,plot=FALSE,lag.max=20)
Autocorrelations of series ‘series’, by lag
     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.188  0.062  0.012 -0.007  0.017  0.026  0.040 -0.019  0.031 -0.025
    11     12     13     14     15     16     17     18     19     20
-0.036 -0.041 -0.008 -0.021 -0.039 -0.067 -0.065 -0.016 -0.019 -0.012
```



PACF for process $y_t = 0.2y_{t-1} + u_t$

```
> pacf(series,plot=FALSE,lag.max=20)
Partial autocorrelations of series ‘series’, by lag
     1      2      3      4      5      6      7      8      9     10     11
 0.188  0.028 -0.004 -0.010  0.020  0.021  0.031 -0.035  0.040 -0.036 -0.029
    12     13     14     15     16     17     18     19     20
-0.030  0.009 -0.021 -0.032 -0.057 -0.035  0.008 -0.011 -0.007
```



Exercise 5.

Suppose the stochastic proces $\{y_t\}$ has the following structure

$$y_t = \theta u_{t-1} + u_t$$

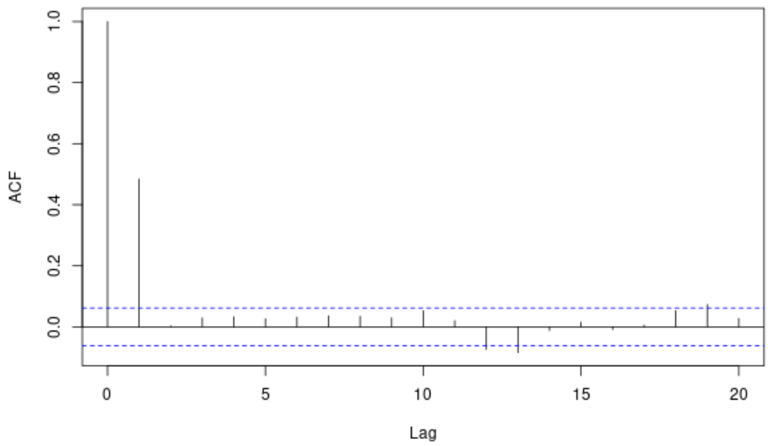
where u_t is Gaussian White Noise with $\sigma_u^2 = 1$. Setting $\theta = 0.8$, simulate 1000 realizations of this process, and plot the ACF and PACF of the resulting data series for lags 1-20.

Solution to Exercise 5.

```
> series <- arima.sim(n=1000,list(ar=c(0,0),ma=c(0.8,0)))
```

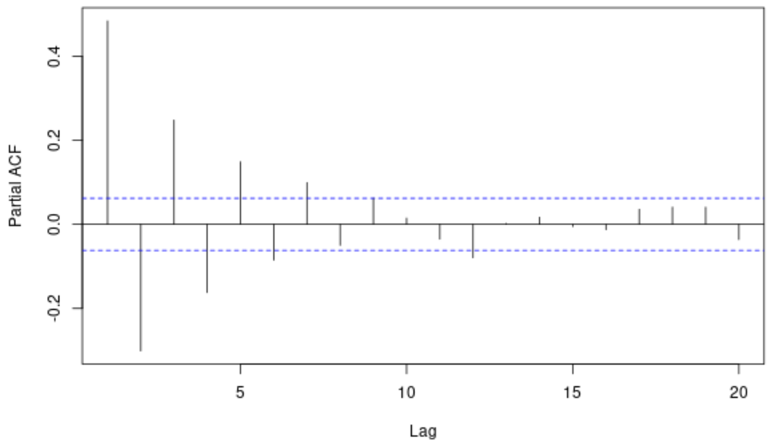
ACF for process $y_t = 0.8u_{t-1} + u_t$

```
> acf(series,plot=FALSE,lag.max=20)

Autocorrelations of series ‘series’, by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.453 -0.058 -0.045 -0.040 -0.019  0.031  0.038  0.052  0.080  0.047
    11     12     13     14     15     16     17     18     19     20
 0.012 -0.030 -0.044 -0.002  0.033 -0.003 -0.025 -0.017 -0.027  0.012
```



PACF for process $y_t = 0.8u_{t-1} + u_t$

```
> pacf(series,plot=FALSE,lag.max=20)

Partial autocorrelations of series ‘series’, by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.453 -0.331  0.198 -0.185  0.133 -0.057  0.066  0.016  0.069 -0.025  0.039
    12     13     14     15     16     17     18     19     20
-0.076  0.033 -0.011  0.035 -0.064  0.026 -0.043  0.001  0.037
```



Exercise 6.

Suppose the stochastic proces $\{y_t\}$ has the following structure

$$y_t = \theta u_{t-1} + u_t$$

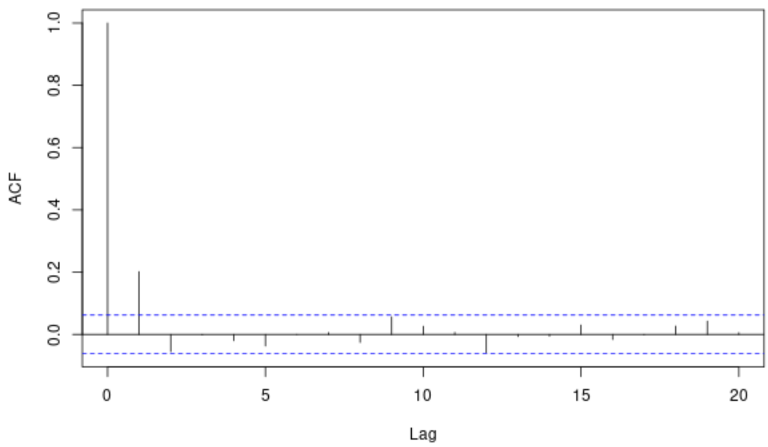
where u_t is Gaussian White Noise with $\sigma_u^2 = 1$. Setting $\theta = 0.2$, simulate 1000 realizations of this process, and plot the ACF and PACF of the resulting data series for lags 1-20.

Solution to Exercise 6.

```
> series <- arima.sim(n=1000,list(ar=c(0,0),ma=c(0.2,0)))
```

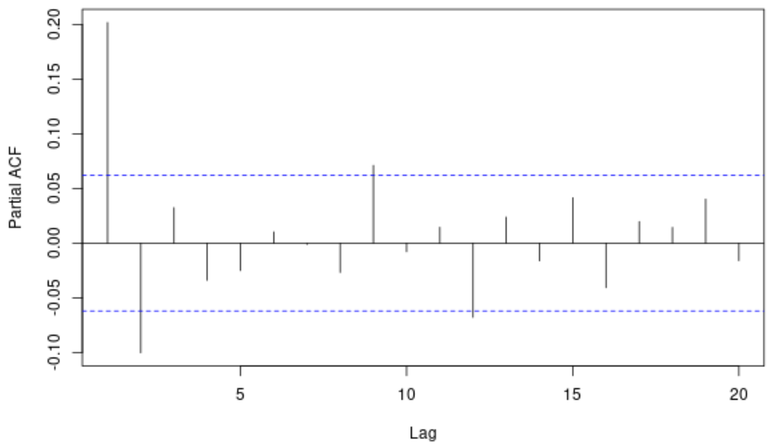
ACF for process $y_t = 0.2u_{t-1} + u_t$

```
> acf(series,plot=FALSE,lag.max=20)

Autocorrelations of series 'series', by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.196  0.012 -0.052 -0.020 -0.001 -0.038 -0.074 -0.026 -0.019  0.006
    11     12     13     14     15     16     17     18     19     20
 0.054  0.080  0.044  0.040 -0.036 -0.083  0.036  0.001  0.030  0.014
```



PACF for process $y_t = 0.2u_{t-1} + u_t$

```
> pacf(series,plot=FALSE,lag.max=20)

Partial autocorrelations of series ‘series’, by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.196 -0.028 -0.051  0.001  0.003 -0.043 -0.062  0.001 -0.019  0.005  0.053
    12     13     14     15     16     17     18     19     20
 0.060  0.013  0.031 -0.046 -0.071  0.075 -0.015  0.035  0.017
```



Exercise 7.
Suppose the stochastic proces $\{y_t\}$ has the following structure

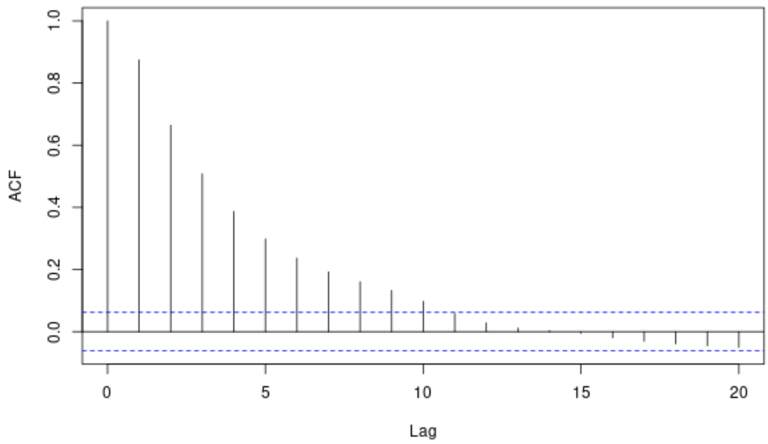
$$y_t = \rho y_{t-1} + \theta u_{t-1} + u_t$$

where u_t is Gaussian White Noise with $\sigma_u^2 = 1$. Setting $\rho = 0.8$ and $\theta = 0.8$, simulate $T = 1000$ realizations of this process, and plot the ACF and PACF of the resulting data series for lags 1-20.
Solution to Exercise 7.

```
> series <- arima.sim(n=1000,list(ar=c(0.8,0),ma=c(0.8,0)))
```

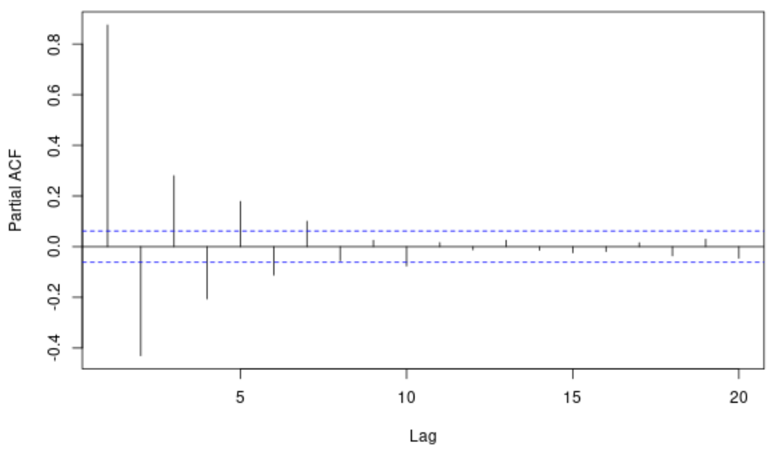
ACF for process $y_t = 0.8y_{t-1} + 0.8u_{t-1} + u_t$

```
> acf(series,plot=FALSE,lag.max=20)

Autocorrelations of series ‘series’, by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.896  0.699  0.521  0.364  0.236  0.141  0.077  0.036  0.010 -0.002
    11     12     13     14     15     16     17     18     19     20
 0.002  0.010  0.018  0.034  0.044  0.045  0.041  0.022 -0.004 -0.026
```



PACF for process $y_t = 0.8y_{t-1} + 0.8u_{t-1} + u_t$

```
> pacf(series,plot=FALSE,lag.max=20)

Partial autocorrelations of series 'series', by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.896 -0.525  0.248 -0.252  0.162 -0.124  0.105 -0.084  0.054 -0.008  0.045
    12     13     14     15     16     17     18     19     20
-0.043  0.055  0.017 -0.051  0.036 -0.044 -0.049  0.017 -0.008
```



Exercise 8.

Suppose the stochastic proces $\{y_t\}$ has the following structure

$$y_t = \rho y_{t-1} + u_t$$

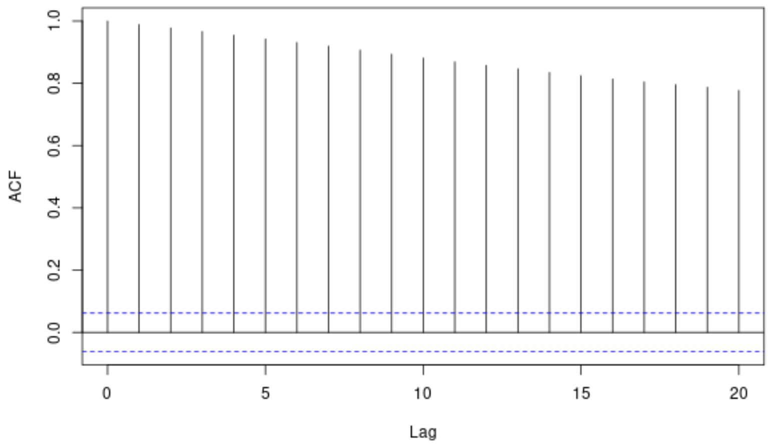
where u_t is Gaussian White Noise with $\sigma_u^2 = 1$. Setting $\rho = 0.99$, simulate 1000 realizations of this process, and plot the ACF and PACF of the resulting data series for lags 1-20. (This is close to a unit root)

Solution to Exercise 8.

```
> series <- arima.sim(n=1000,list(ar=c(0.99,0),ma=c(0,0)))
```

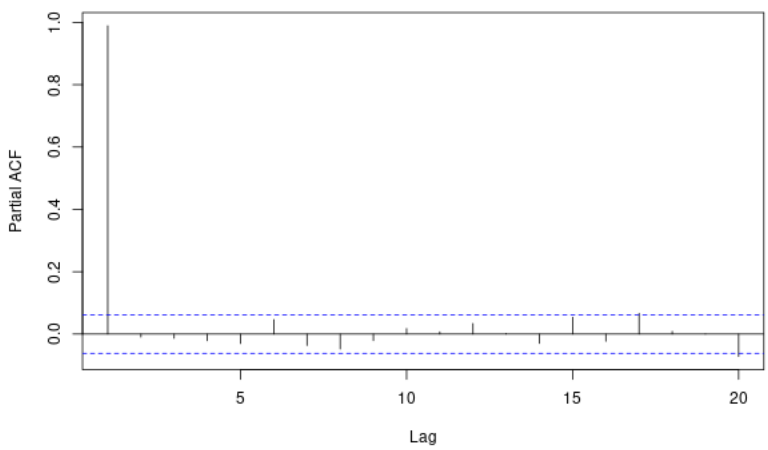
ACF for process $y_t = 0.99y_{t-1} + u_t$

```
> acf(series,plot=FALSE,lag.max=20)

Autocorrelations of series ‘series’, by lag

    0     1     2     3     4     5     6     7     8     9    10    11    12
1.000 0.986 0.971 0.955 0.938 0.922 0.905 0.889 0.874 0.859 0.845 0.833 0.821
   13    14    15    16    17    18    19    20
0.809 0.797 0.785 0.773 0.762 0.751 0.740 0.730
```



PACF for process $y_t = 0.99y_{t-1} + u_t$

```
> pacf(series,plot=FALSE,lag.max=20)
Partial autocorrelations of series ‘series’, by lag
     1      2      3      4      5      6      7      8      9     10     11
 0.986 -0.031 -0.063 -0.006 -0.009 -0.006  0.004  0.007  0.017  0.038  0.025
    12     13     14     15     16     17     18     19     20
 0.024 -0.036 -0.022  0.012  0.009  0.018  0.001 -0.011  0.013
```

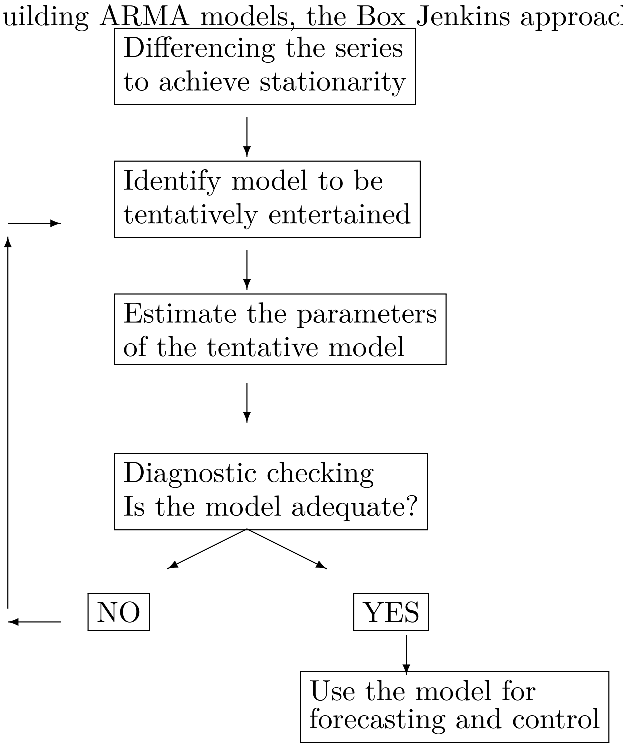


3.8 Estimation of a given model

Given an AR/MA/ARMA specification, how is it estimated?
Usually, since for MA specifications the errors are not observed, the estimation is done with Maximum Likelihood.
One exception is $AR(p)$ specifications, without MA terms. Here OLS gives consistent estimates of the AR parameters, although Maximum Likelihood may be better also here.

3.9 Box Jenkins model selection

Building ARMA models, the Box Jenkins approach.



The classical approach: Looking at the ACF and PACF functions to determine a reasonable structure.

Exercise 9.

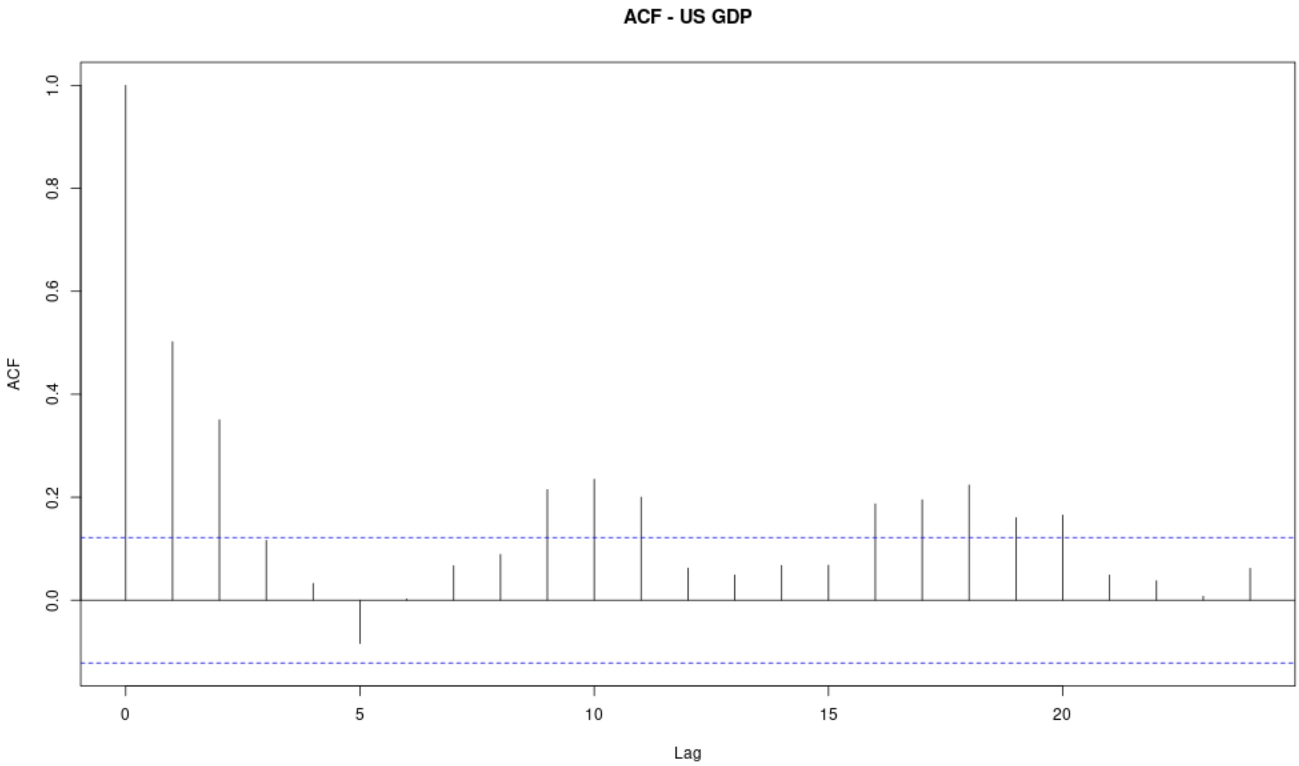
Collect quarterly data from the US on the aggregate GDP (Gross Domestic Product) for the period 1947:I to 2007:II. Calculate the log difference of the GDP series ($\ln(Y_t) - \ln(Y_{t-1})$). You want to model this series using time series, and apply the Box Jenkins methodology to select a reasonable representation.

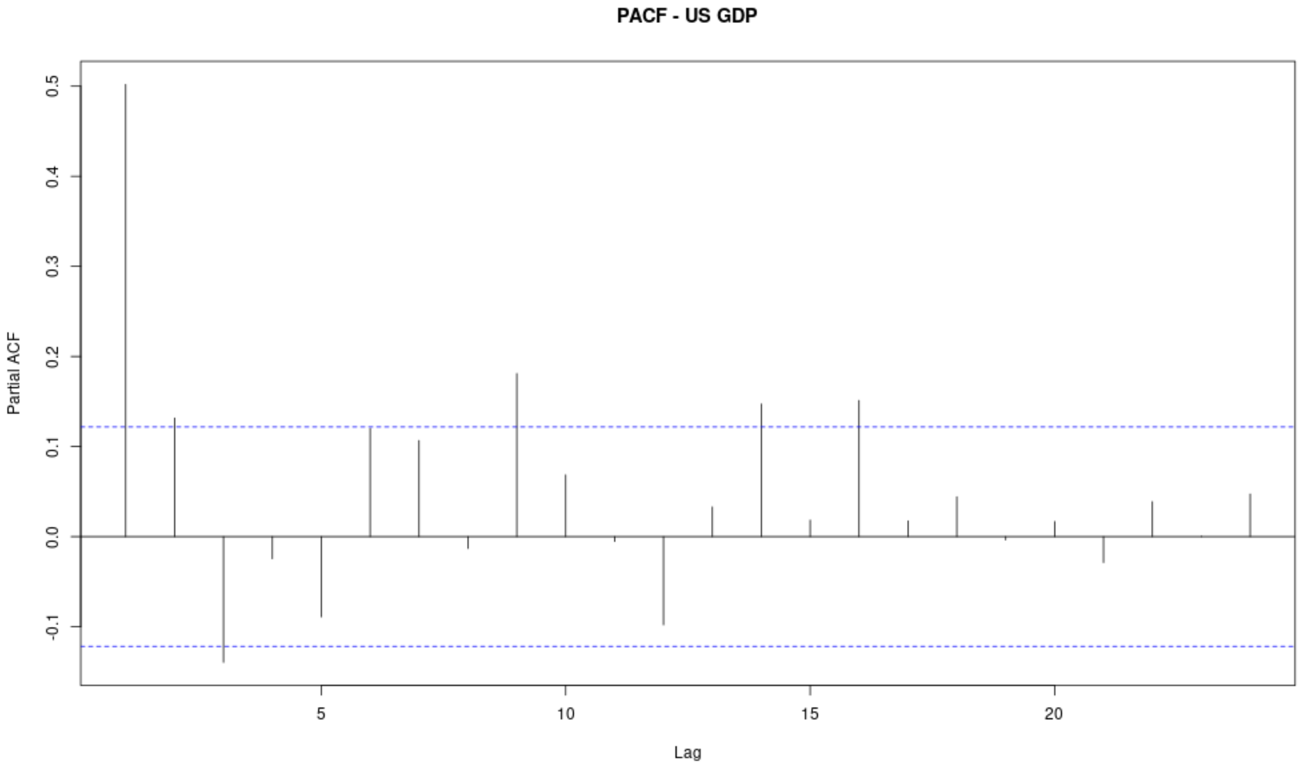
1. Plot the series.
2. Calculate the ACF and PACF for the series.
3. Use the ACF and PACF to select a model specifaction (AR/MA/ARMA).

Solution to Exercise 9.

```
> GDP <- read.table("../../../../data/usa/us_macro_data/us_gdp.csv",
```

```
         skip=3,header=TRUE)
> gdp <- ts(GDP[,3],frequency=4,start=c(1947,3))
> dgdp <- diff(log(gdp))
> acf(as.matrix(dgdp),plot=FALSE)

Autocorrelations of series ‘as.matrix(dgdp)’, by lag

     0      1      2      3      4      5      6      7      8      9     10
 1.000  0.502  0.350  0.116  0.033 -0.084  0.002  0.067  0.089  0.215  0.235
    11     12     13     14     15     16     17     18     19     20     21
 0.200  0.062  0.049  0.067  0.068  0.187  0.195  0.224  0.160  0.165  0.049
    22     23     24
 0.038  0.008  0.062

> pacf(as.matrix(dgdp),plot=FALSE)

Partial autocorrelations of series ‘as.matrix(dgdp)’, by lag

     1      2      3      4      5      6      7      8      9     10     11
 0.502  0.132 -0.140 -0.025 -0.089  0.120  0.107 -0.013  0.181  0.069 -0.005
    12     13     14     15     16     17     18     19     20     21     22
-0.098  0.033  0.147  0.018  0.151  0.017  0.044 -0.004  0.017 -0.029  0.039
    23     24
 0.000  0.047
```





3.10 Using information criteria to choose model

The more modern approach - Information criterion
For a large number of possible model specifications, estimate the model, and then calculate a measure of fit.
The information critera is based on the theory of *non-nested tests*, for the interested.
Akaike's information criterion.

$$AIC = \ln\left(\hat{\sigma}^2\right) + \frac{2k}{T}$$

where k is the lag length.

Exercise 10.

Collect quarterly data from the US on the aggregate GDP (Gross Domestic Product) for the period staring 1947:3. Calculate the log difference of the GDP series ($\ln(Y_t) - \ln(Y_{t-1})$). You want to model this series using time series. Preliminary plots of the ACF suggestes an AR representation. Compare various AR representations using Akaike's information criterion.

$$AIC = \ln\left(\hat{\sigma}^2\right) + \frac{2k}{T}$$

Solution to Exercise 10.

```
> GDP <- read.table("../../../../data/usa/us_macro_data/us_gdp.csv",skip=3,header=TRUE)
> gdp <- ts(GDP[,3],frequency=4,start=c(1947,3))
> dgdp <- diff(log(gdp))
> ar(as.matrix(dgdp),aic=TRUE,plot=FALSE)
```

```
Call:
ar(x = as.matrix(dgdp), aic = TRUE, plot = FALSE)

Coefficients:
      1        2        3        4        5        6        7        8
 0.4307   0.1888  -0.1029   0.0190  -0.1704   0.0926   0.0345  -0.0975
      9       10       11       12       13       14       15       16
 0.1490   0.0776   0.0716  -0.1384  -0.0195   0.1076  -0.0472   0.1512

Order selected 16  sigma^2 estimated as  8.731e-05
```

3.11 Forecasting

Forecasting in econometrics.
Forecasting: Predicting the values a series is likely to take.
Chief worry: Forecasting accuracy. If you get accurate forecasts, who cares where they come from?
Two approaches to forecasting:

- Econometric (structural) forecasting. (Comes from a given *economic* model.)
- Time series forecasting. (General functions of past data and errors).

Difference between

- In-sample forecasts.

  Generated for the same sample as was used to estimate the model's parameters.

- Out-of-sample forecasts.

  Using estimated parameters on "fresh data," data not used to generate parameter estimates.

What do you forecasts?

- Tomorrow/next period only – one step ahead forecast.
- Several periods forward – multistep ahead forecasts.

Time series forecasting.
Do not cover forecasting with structural models, since they require forecasts for explanatory variables. Therefore, of more interest is forecasting with the usual time-series models.

3.12 Forecasting with ARMA models

Want: $E[y_{t+s}|\Omega_t]$: Expectation of process at time $t+s$ conditional on information at time t.
In particular, want $E[y_{t+1}|\Omega_t]$, the one step ahead forecast.
Suppose we have an $ARMA(p,q)$

$$y_t = \sum_{t=1}^{p} a_i y_{t-i} + \sum_{j=1}^{q} b_j u_{t-j} + u_t$$

Note that

$$E[u_{t+s}|\Omega_t] = 0 \;\; \forall \; s > 0$$

(Since this is an indepedent process, our best guess is the unconditional expectation)
Let $f_{t,s}$ be the forecast at time t for s steps into the future

$$f_{t,1} = E[y_{t+1}|\Omega_t]$$

for a general $ARMA(p,q)$.

$$f_{t,s} = \sum_{i=1}^{p} a_i f_{t,s-i} + \sum_{j=1}^{q} b_j u_{t+s-j}$$

Note that in this function $u_{t+s} = 0$ if $s > 0$ and equal to the realized error u_{t+s} if $s \leq 0$.

Exercise 11.

Suppose you have a MA(3) process

$$y_t = \mu + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \theta_3 u_{t-3} + u_t$$

where u_t is White Noise.
What are the one step, two step, three step and four step ahead forecasts?

Solution to Exercise 11.

One step ahead forecasts

$$E[y_{t+1}|y_t]$$

$$y_{t+1} = \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + u_{t+1}$$

$$\begin{aligned} E[y_{t+1}|y_t] &= \mu + \theta_1 E[u_t|y_t] + \theta_2 E[u_{t-1}|y_t] + \theta_3 E[u_{t-2}|y_t] + E[u_{t+1}|y_t] \\ &= \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} + 0 \\ &= \mu + \theta_1 u_t + \theta_2 u_{t-1} + \theta_3 u_{t-2} \end{aligned}$$

Two step ahead forecasts

$$E[y_{t+2}|y_t]$$

$$y_{t+2} = \mu + \theta_1 u_{t+1} + \theta_2 u_t + \theta_3 u_{t-1} + u_{t+2}$$

$$\begin{aligned} E[y_{t+2}|y_t] &= \mu + \theta_1 E[u_{t+1}|y_t] + \theta_2 E[u_t|y_t] + \theta_3 E[u_{t-1}|y_t] + E[u_{t+2}|y_t] \\ &= \mu + \theta_1 0 + \theta_2 u_t + \theta_3 u_{t-1} + 0 \\ &= \mu + \theta_2 u_t + \theta_3 u_{t-1} \end{aligned}$$

Three step ahead forecasts

$$E[y_{t+3}|y_t]$$

$$y_{t+3} = \mu + \theta_1 u_{t+2} + \theta_2 u_{t+1} + \theta_3 u_t + u_{t+3}$$

$$\begin{aligned} E[y_{t+3}|y_t] &= \mu + \theta_1 E[u_{t+2}|y_t] + \theta_2 E[u_{t+1}|y_t] + \theta_3 E[u_t|y_t] + E[u_{t+3}|y_t] \\ &= \mu + 0 + 0 + \theta_3 u_t + 0 \\ &= \mu + \theta_3 u_t \end{aligned}$$

Four step ahead forecasts

$$E[y_{t+4}|y_t]$$

$$y_{t+4} = \mu + \theta_1 u_{t+3} + \theta_2 u_{t+2} + \theta_3 u_{t+1} + u_{t+4}$$

$$\begin{aligned} E[y_{t+4}|y_t] &= \mu + \theta_1 E[u_{t+3}|y_t] + \theta_2 E[u_{t+2}|y_t] + \theta_3 E[u_{t+1}|y_t] + E[u_{t+4}|y_t] \\ &= \mu + 0 + 0 + 0 + 0 \\ &= \mu \end{aligned}$$

Exercise 12.
Suppose you have a AR(2) process

$$y_t = \mu + \phi_1 y_{t-1} + \phi_2 y_{t-2} + u_t$$

where u_t is White Noise.
What are the one step, two step and three step ahead forecasts?
Solution to Exercise 12.
One step ahead forecast

$$E[y_{t+1}|y_t]$$

$$y_{t+1} = \mu + \phi_1 y_t + \phi_2 y_{t-1} + u_{t+1}$$

$$\begin{aligned} E[y_{t+1}|y_t] &= \mu + \phi_1 E[y_t|y_t] + \phi_2 E[y_{t-1}|y_t] + E[u_{t+1}|y_t] \\ &= \mu + \phi_1 y_t + \phi_2 y_{t-1} + 0 \\ &= \mu + \phi_1 y_t + \phi_2 y_{t-1} \end{aligned}$$

Two step ahead forecast

$$E[y_{t+2}|y_t]$$

$$y_{t+2} = \mu + \phi_1 y_{t+1} + \phi_2 y_t + u_{t+2}$$

$$\begin{aligned} E[y_{t+2}|y_t] &= \mu + \phi_1 E[y_{t+1}|y_t] + \phi_2 E[y_t|y_t] + E[u_{t+2}|y_t] \\ &= \mu + \phi_1 E[y_{t+1}|y_t] + \phi_2 y_t + 0 \\ &= \mu + \phi_1 (\mu + \phi_1 y_t + \phi_2 y_{t-1}) + \phi_2 y_t \\ &= \mu + \phi_1 \mu + \phi_1 \phi_1 y_t + \phi_1 \phi_2 y_{t-1} + \phi_2 y_t \\ &= \mu + \phi_1 \mu + \phi_1^2 y_t + \phi_1 \phi_2 y_{t-1} + \phi_2 y_t \\ &= \mu + \phi_1 \mu + \left(\phi_1^2 + \phi_2\right) y_t + \phi_1 \phi_2 y_{t-1} \end{aligned}$$

Three step ahead forecast

$$E[y_{t+3}|y_t]$$

$$y_{t+3} = \mu + \phi_1 y_{t+2} + \phi_2 y_{t+1} + u_{t+3}$$

$$\begin{aligned} E[y_{t+3}|y_t] &= \mu + \phi_1 E[y_{t+2}|y_t] + \phi_2 E[y_{t+1}|y_t] + E[u_{t+3}|y_t] \\ &= \mu + \phi_1 \left(\mu + \phi_1 \mu + \left(\phi_1^2 + \phi_2\right) y_t + \phi_1 \phi_2 y_{t-1}\right) + \phi_2 (\mu + \phi_1 y_t + \phi_2 y_{t-1}) + 0 \\ &= \left(1 + \phi_1 + \phi_1^2 + \phi_1\right) \mu + \left(\phi_1^2 + \phi_2 + \phi_1 \phi_2\right) y_t + \left(\phi_1^2 \phi_2 + \phi_2^2\right) y_{t-1} \end{aligned}$$

3.13 Comparing forecasts.

This is relevant for out-of-sample work, where we use the forecast model to predict values, and then compare the forecasts to the realizations.
Want to have the forecasts as "close" to the realized values as possible. The close, the better forecast quality.
Need a metric for asking "how close" the forecasts are to the realizations.
Metrics for evaluating forecast performance
Mean Squared Error

$$MSE = \frac{1}{T-(T_1-1)}\sum_{t=1}^{T}(y_{t+s}-f_{t,s})^2$$

Mean Absolute Error

$$MAE = \frac{1}{T-(T_1-1)}\sum_{t=1}^{T}|y_{t+s}-f_{t,s}|$$

Mean Absolute Percentage Error

$$MAPE = \frac{1}{T-(T_1-1)}\sum_{t=1}^{T}\left|\frac{y_{t+s}-f_{t,s}}{y_{t+s}}\right|$$

Adjusted AMAPE

$$MAPE = \frac{1}{T-(T_1-1)}\sum_{t=1}^{T}\left|\frac{y_{t+s}-f_{t,s}}{y_{t+s}+f_{t,s}}\right|$$

Theils U-statistic

$$U = \sqrt{\sum_{t=T_1}^{T}\frac{\left(\frac{y_{t+s}-f_{t,s}}{x_{t+s}}\right)^2}{\left(\frac{y_{t+s}-fb_{t,s}}{x_{t+s}}\right)^2}}$$

where fb is a *benchmark* forecast.
Alternative, closer to economic penalty function:
Count number of successful predictions of right sign.
Test for whether you can do better than pure chance.

Literature Most of this lecture is taken rather directly from Brooks (2002), with additional input from Hamilton (1994).

4 Stock prices

However, when we in finance (and economics) talk about "time series analysis" we typically have in mind the relationship between *past* realizations of a variable, and the *next* realization, i.e. *prediction.*

However, to put this in the finance perspective, we don't always try to predict something, we often try to establish a *lack* of predictive ability.

To put this seemingly strange statement in perspective, let us talk a bit about finance theory.

The theory of efficient markets states (roughly) that the current price of a financial asset is the markets best evaluation of what the assets value *is.* Any alternative prediction than what is done by the market can not do better than the market.

Thinking about this, this statement needs to be formalized in some way to make it testable.

The simplest possible such formalization is the

4.1 Random Walk Model

$$P_t = P_{t-1} + \varepsilon_t$$

where P_t is the stock price at time t, P_{t-1} the price at time $t-1$, and ε_t is a random term with expectation zero.

The name Random Walk betrays the model's origin, which was to describe the path of a drunk left in the middle of a field.

Expanding this

$$P_t = P_{t-1} + \varepsilon_t$$

$$= P_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$= P_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$= P_{t-4} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t$$

$$= P_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j}$$

Observe immediately that in the Random Walk model, the effect of a shock (ε_t) is permanent.

While the Random Walk model is simple, it does not suffice as a model of stock price behaviour. If it is one thing that we know about stock returns, it is that holding stock had better promise higher expected return than risk free investments, otherwise who would bother?

Writing stock returns

$$R_t = \frac{P_t - P_{t-1} + D_t}{P_{t-1}}$$

and assuming they are constant expectation μ

$$E[R_t] = \mu$$

$$\mu = \frac{P_t - P_{t-1} + D_t}{P_{t-1}}$$

$$\mu P_{t-1} = P_t + D_t - P_{t-1}$$

$$P_t + D_t = P_{t-1}(1 + \mu)$$

Typically which suggests that modelling stock prices as

$$P_t + D_t = P_{t-1}(1 + \mu) + \varepsilon_t$$

Typically we will add dividends into the price, so that P_t now includes dividend paid out at time t.
Then we can write

$$P_t = P_{t-1}(1 + \mu) + \varepsilon_t$$

Thus, the best estimate of tomorrows price is today's price plus the one-period expected return. If $\mu > 0$ this is what is called a supermartingale

$$E[P_t] > P_{t-1}$$

5 Readings

Specialized textbook for Finance and R: Tsay (2013).
The Bible: Hamilton (1994).
Much of the overview is taken from Gourieroux and Monfort (1997).
A good discussion of lag operators is in Hamilton (1994), which is also the preferred advanced source for time series analysis.
Most textbooks in econometrics, like Greene (1997) and Davidson and MacKinnon (1993) will have sections on time series analysis. There are also a huge number of pure time series texts.
For unit roots and integration: Campbell and Perron (1991) is a short overview, alternatively (Davidson and MacKinnon, 1993, 20.1), (Hamilton, 1994, Ch 17).
For VAR's:(Davidson and MacKinnon, 1993, 19.5),(Hamilton, 1994, Ch 11: 11.1, 11.2).

References

Chris Brooks. *Introductory econometrics for finance.* Cambridge, 2002.

John Y Campbell and Pierre Perron. Pitfalls and opportunities: What Macroeconmists should know about Unit Roots. In Olivier Jean Blanchard and Stanley Fisher, editors, *NBER Macroeconomics annual 1991.* MIT Press, 1991.

Russel Davidson and James G MacKinnon. *Estimation and Interference in Econometrics.* Oxford University Press, 1993.

Christian Gourieroux and Alain Monfort. *Time Series and Dynamic Models.* Cambridge University Press, 1997.

William H Greene. *Econometric Analysis.* Prentice–Hall, third edition, 1997.

James D Hamilton. *Time Series Analysis.* Princeton University Press, 1994.

Ruey S Tsay. *An Introduction to Analysis of Financial Data with R.* Wiley, Hoboken, New Jersey, 2013.