Time Series

A lecture on time series.

Survey of time series issues

Definition

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

$$x_1, x_2 \ldots x_t \ldots x_T$$

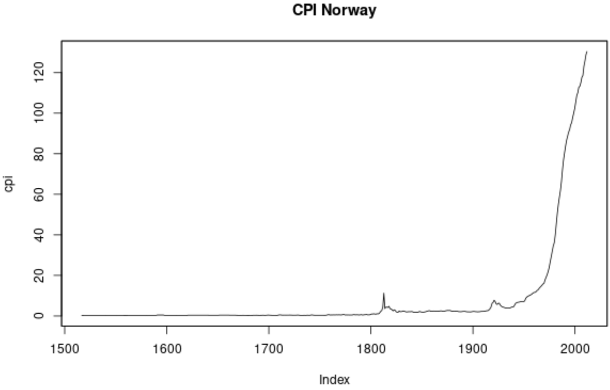
usually sampled at fixed intervals.

Typical intervals: Daily, weekly, monthly, quarterly, annual.

Time imposes ordering.

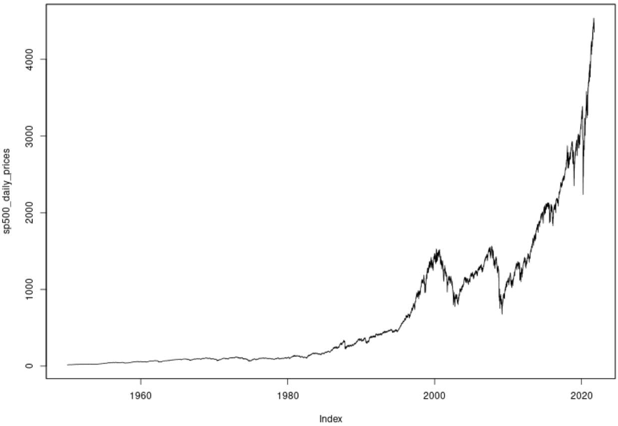
Examples

Example: Consumer Price Index. What is the relative value of one unit of currency.



Another example: Stock market index. What is the total value of an investment in the stock market. Evolution of a stock market index

Evolution of the S&P 500



Stocks or flows?

Common terminology.

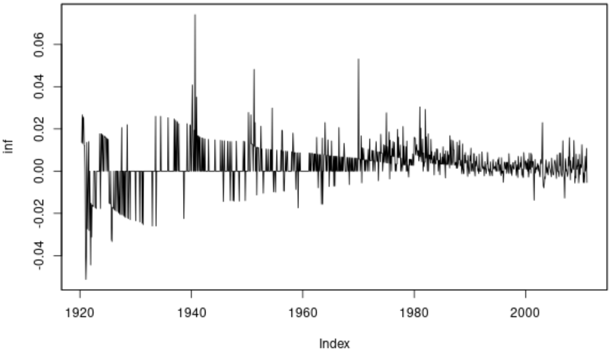
- Stocks measures the value of something at a specific time,
- a flow measures something between two time periods.

The price index is an example of a stock.

Flow example: Inflation

“first differences” of the price indices.

monthly inflation for Norway:

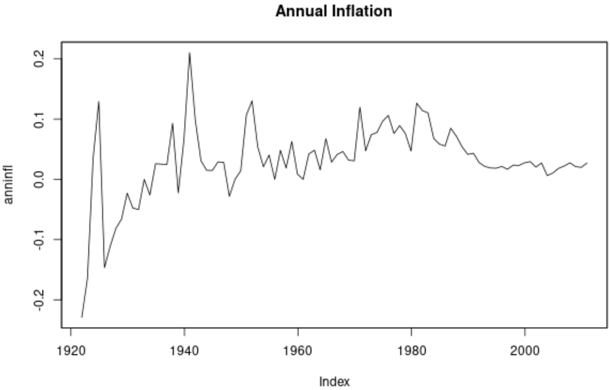


An example: Norwegian Consumer Price Index. First Differences

Convenience: Transformations

Hard to interpret monthly inflation, what is a convenient transformation?

Example:**annual** observations.



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

Typical Problems in Time Series

1. Forecasting.
2. Trend Removal.
3. Seasonal Adjustment.
4. Detection of a Structural Break.
5. Causality–time lags.
6. Distinction between the short and long run
7. Study of agent's expectations.

Trend Removal.

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).

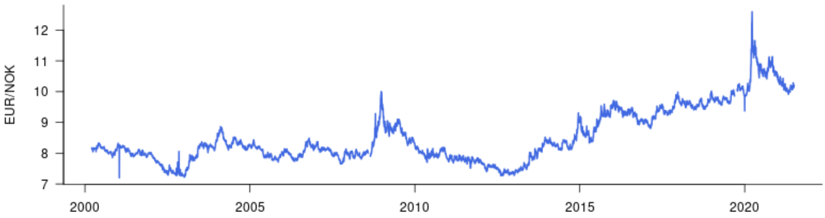
If the series have a clear trend (see the examples of price indices and stock indices), may want to remove the trend and analyze the movements *around* the trend.

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.

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



NOK/EUR exchange rate

Causality–time lags.

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

Distinction between the short and long run

To what degree are relationships *persistent*?

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?

Categorizing time series tools

Distintuish

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

Adjustment models.

Roughly, mechanical models for removing seasonal components and trends.

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.

Example Autopredictive models - AR

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:
autoregressive model of order 1: (an AR(1) process)

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

More general, autoregressive specifications with p terms, AR(p):

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

Example Autopredictive models - MA

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

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

Add terms to produce a MA(q) process:

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

Example Autopredictive models - ARMA

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

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

Issue with autopredictive AR case: Unit root

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*.

Issue with autopredictive AR case: Unit root

Testing for unit root
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 standard normal distribution

Autopredictive toot: Vector Autoregressions (VAR's)

Consider 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.

Autopredictive models - summarizing

"Model-free" in the sense that — not estimating an explicit economic model,
instead: – exploring possible linear relationships between economic variables.
Uses of the autopredictive methods

- forecasting.
- generate "stylized facts" that our economic theories should be able to explain.

Note though: VAR or some other time series model does not *explain* any causality or other regularities in the data.

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.

Reminder – time series in regressions

$$y_t = X_t b + e_t$$

- 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 $t = 1, \ldots T$.

Reminder – time series in regressions

Typical problem:

- the error term e_t will not be independent.

Think of a shock to the economy.
The effects of this shock may very often persist for a time
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.

The Lag operator

Confusing in time series literature: the use of the *lag* operator.
Suppose our time series is y_t.
The lag operator L just shifts the time index one period back

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

Lag operator useful because can do algebra using the lag operator.
For example

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

This will be written

$$L^2 y_t$$

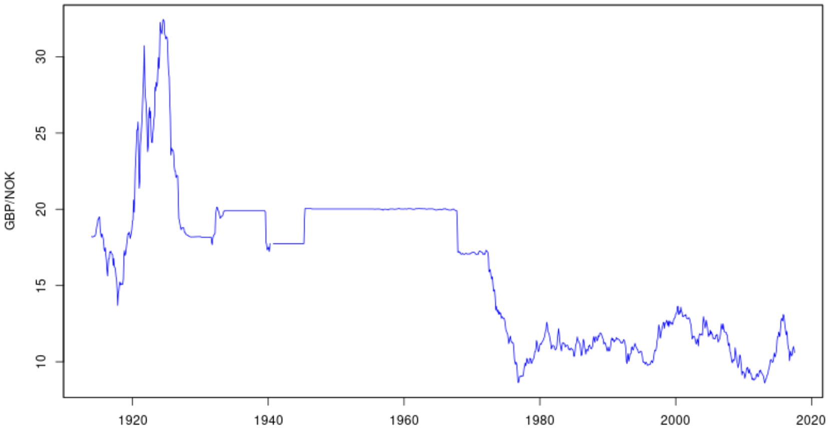
Lag operator – way of transforming time series following the rules of multiplication and addition.
Compactly express complicated time series models.
For example

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

Plotting

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

Example: GBP/NOK



Univariate time series

Go through some of the standard concepts in autorpredictive modelling of univariate time series

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

Can at times 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}.

Stationarity

First question is: Is a series *stationary*?

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

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

is the same as the distribution of

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

for all possible n and k.

Stationarity

weaker concept *weak stationarity*, defined in terms of first and second moments of $\{x_t\}$.
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.

Describing time series dependence - ACF/PACF

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)

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

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

Example

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

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

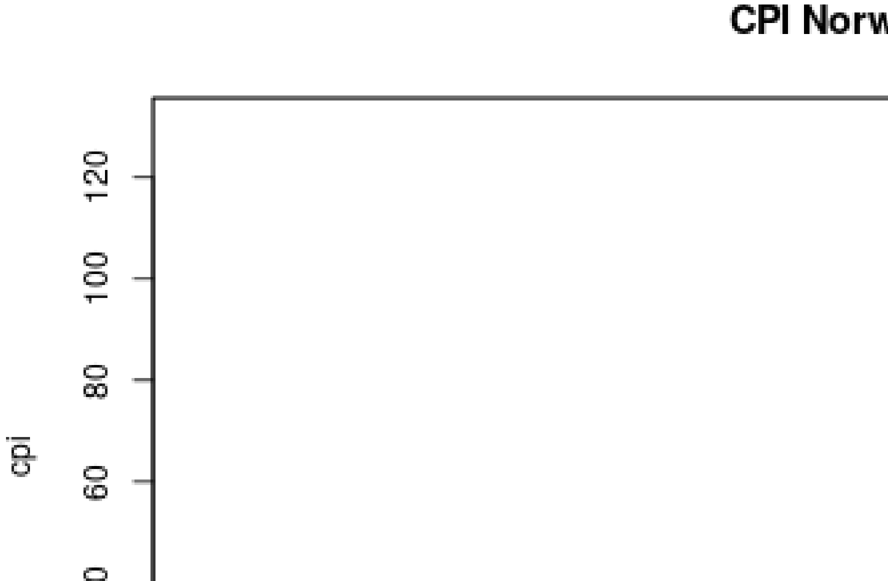
based on the CPI series. Let $q = 5$.

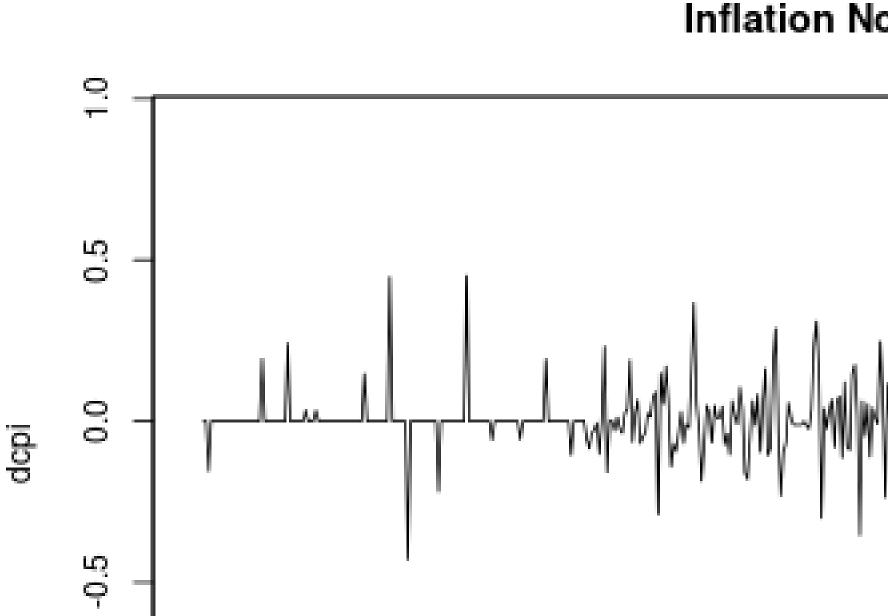
1. Plot the two series.
2. Estimate the autocorrelations (acf) of orders 1 through q for the CPI series.
3. Estimate the partial autocorrelations (pacf) of orders 1 through q for the CPI series.
4. Estimate the average inflation for the period.
5. Estimate the autocorrelations (acf) of orders 1 through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

Solution

```
cpi <- read.zoo("../../../../data/norway/nb_historical_statistics/cpi_norway_1
                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

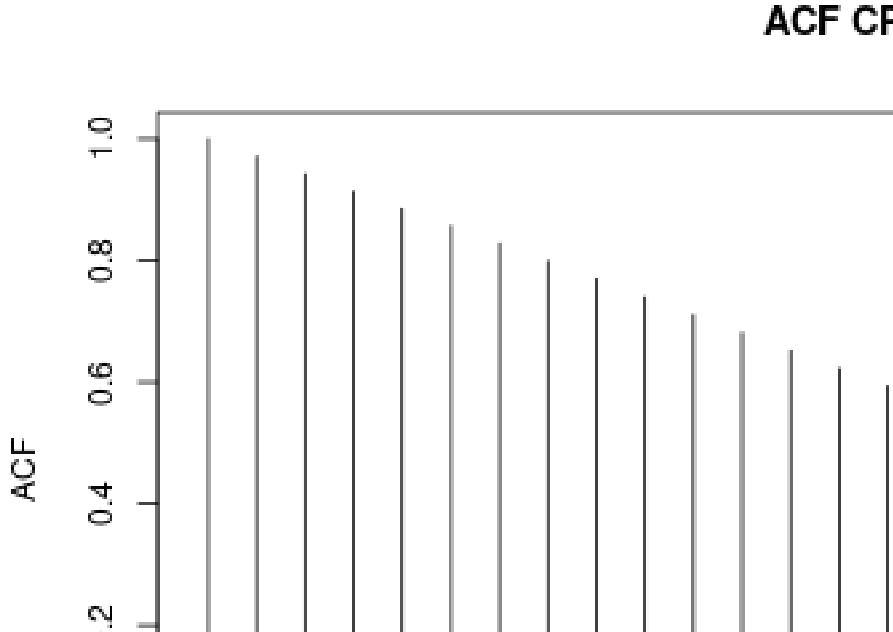


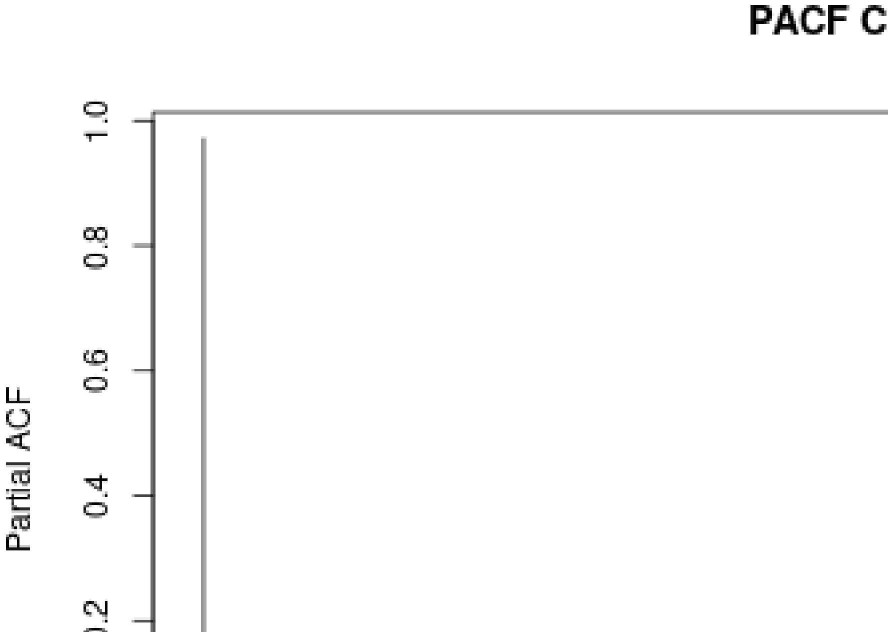


Inflation No
1.0
0.5
0.0
-0.5
dcpi

Describing the cpi series

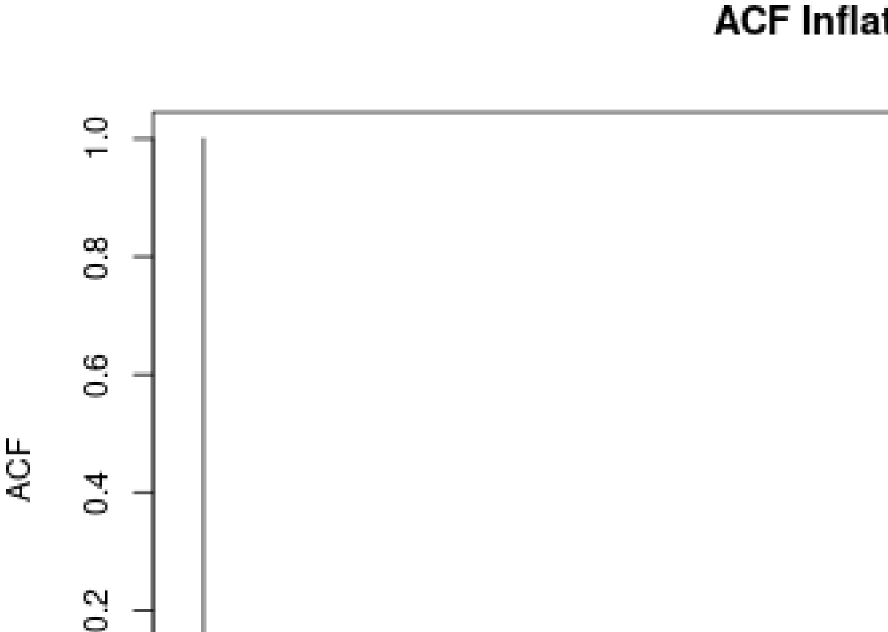
```
> acf(cpi,plot=FALSE,lag.max=5)

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
```


ACF CP
ACF
1.0
0.8
0.6
0.4

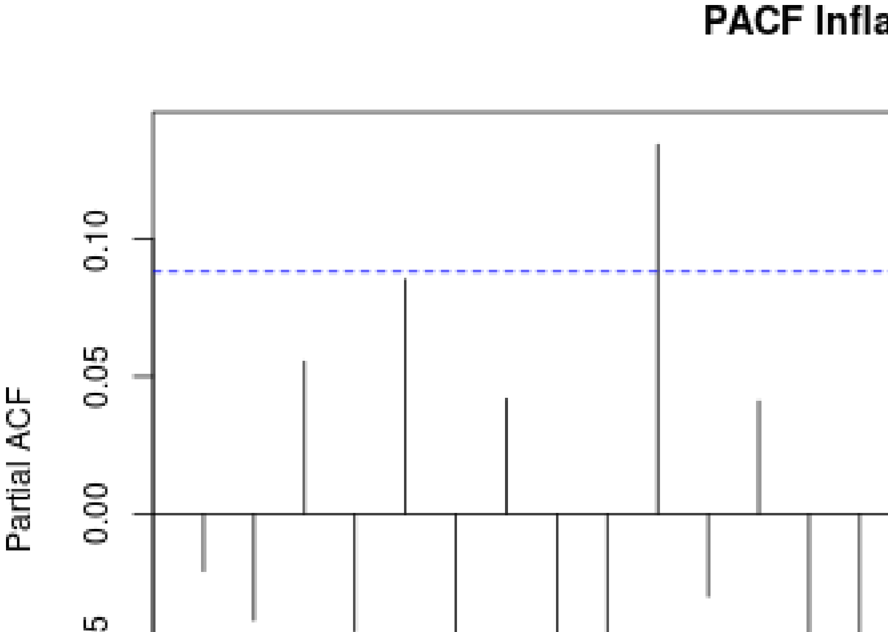

PACF C
Partial ACF
1.0
0.8
0.6
0.4
0.2

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
```



ACF Inflat
ACF
1.0
0.8
0.6
0.4
0.2


PACF Infla
Partial ACF
0.10
0.05
0.00

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.

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.

Testing for autocorrelation

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.

Example

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

```
cpi <- read.zoo("../../../../data/norway/nb_historical_statistics/cpi_norway_1
                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
 0.0079  -0.0538   0.0788  -0.0109   0.0544  -0.0270   0.0
      9        10        11        12        13        14
-0.0353   0.1307  -0.0198   0.0247  -0.0857  -0.1147   0.0
     17
 0.1365
Order selected 17  sigma^2 estimated as  0.01315
```

So the optimal autoregressive model involves 17 autoregressive terms.

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`.

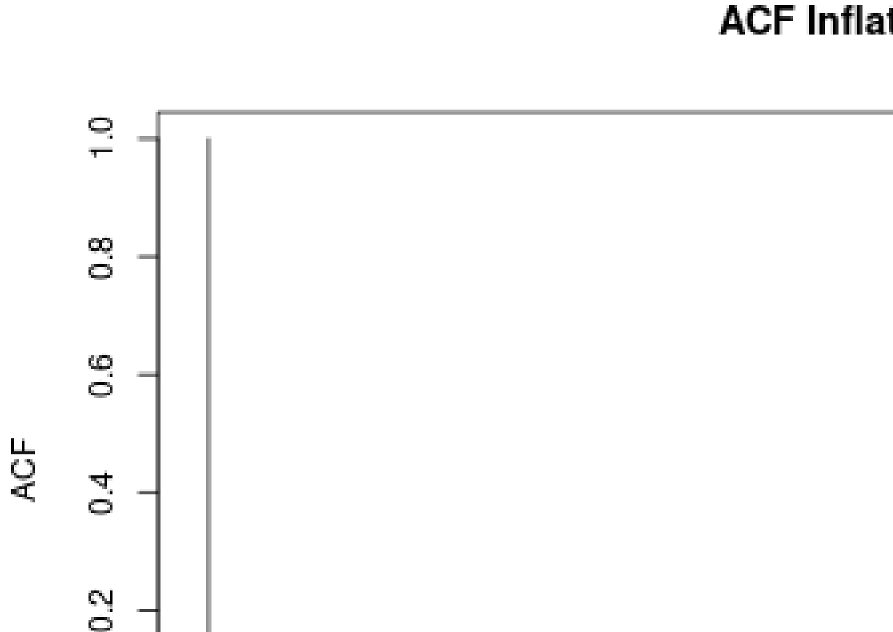
```
> ar(dcpi,AIC=true,order.max=5)
Call:
ar(x = dcpi, order.max = 5, AIC = true)

Order selected 0  sigma^2 estimated as  0.01382
```

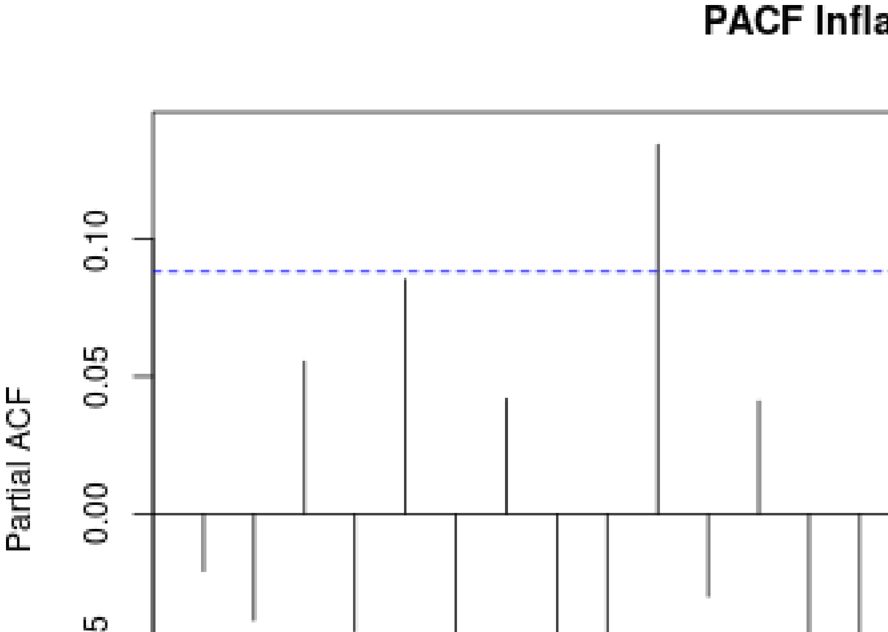
If the maximal lag length is five, the model selected has zero autoregressive terms.

```
> ar(dcpi,AIC=true,order.max=10)

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

Coefficients:
      1        2        3        4        5        6
-0.0016  -0.0429   0.0595  -0.0382   0.0758  -0.0399   0.0
      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
-0.0022  -0.0482   0.0692  -0.0252   0.0644  -0.0371   0.02
      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
```

Once the max is above 10, the order increases. One can figure out why by looking at the ACF and PACF plots.


ACF Inflat
ACF
1.0
0.8
0.6
0.4
0.2


PACF Infla
Partial ACF
0.10
0.05
0.00

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

The Lag operator

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

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

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

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.

Wold decomposition theorem

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

ARMA processes

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

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

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

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

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

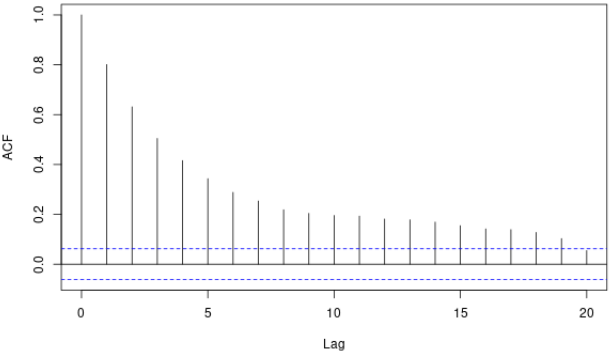
where u_t is Gaussian White Noise with $\sigma_u^2 = 1$. Setting $\rho = 0.8$ and $y_0 = 0$ simulate $T = 1000$ realizations of this process, and plot the ACF and PACF of the resulting data series for lags 1-20.

Doing the simulation

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

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

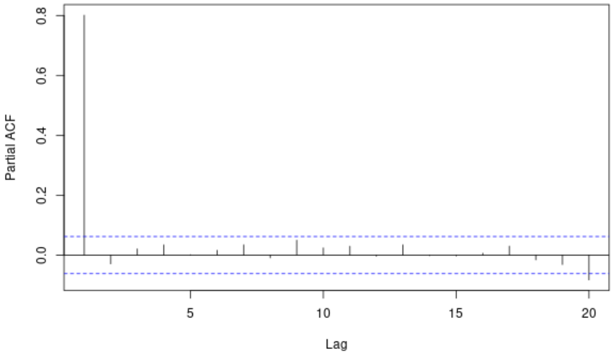
```
> acf(series,plot=FALSE,max.lag=20)
Autocorrelations of series ‘series’, by lag
     0      1      2      3      4      5      6      7      8
 1.000  0.791  0.654  0.505  0.398  0.305  0.223  0.168  0.124
    11     12     13     14     15     16     17     18     19
 0.048  0.057  0.068  0.088  0.108  0.115  0.104  0.079  0.060
    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
```



ACF
1.0
0.8
0.6
0.4
0.2
0.0
0
5
10
15
20
Lag

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

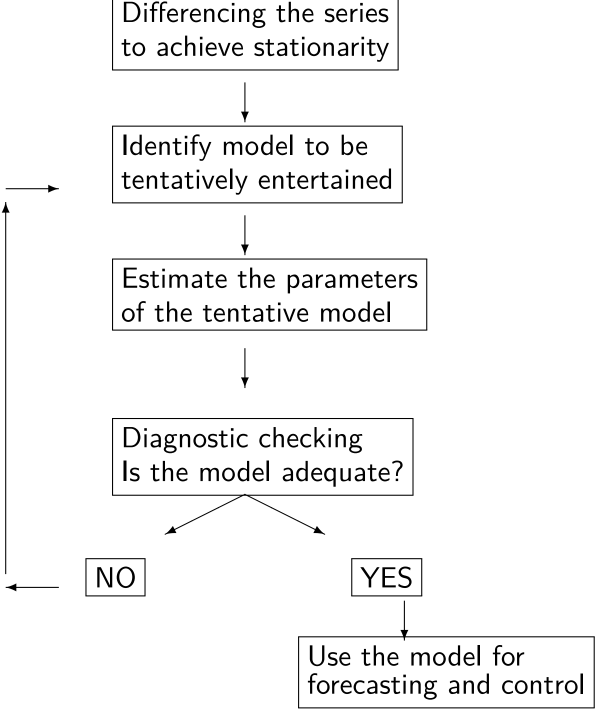
```
> pacf(series,plot=FALSE,max.lag=20)

Partial autocorrelations of series ‘series’, by lag

     1      2      3      4      5      6      7      8
 0.791  0.077 -0.091  0.012 -0.013 -0.033  0.015  0.001 -0.
    12     13     14     15     16     17     18     19
 0.043  0.028  0.040  0.026 -0.003 -0.034 -0.040  0.003  0.
    23     24     25     26     27     28     29     30
 0.011  0.000  0.007 -0.042  0.003 -0.055 -0.023 -0.007
```



Partial ACF
0.0
0.2
0.4
0.6
0.8
5
10
15
20
Lag

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.

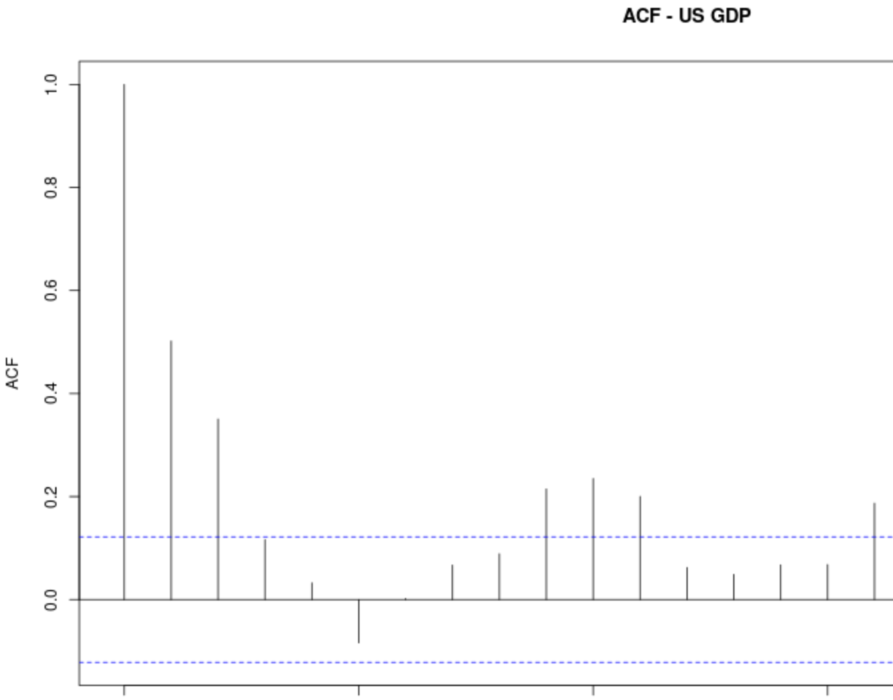

Differencing the series
to achieve stationarity
Identify model to be
tentatively entertained
Estimate the parameters
of the tentative model
Diagnostic checking
Is the model adequate?
NO
YES
Use the model for
forecasting and control

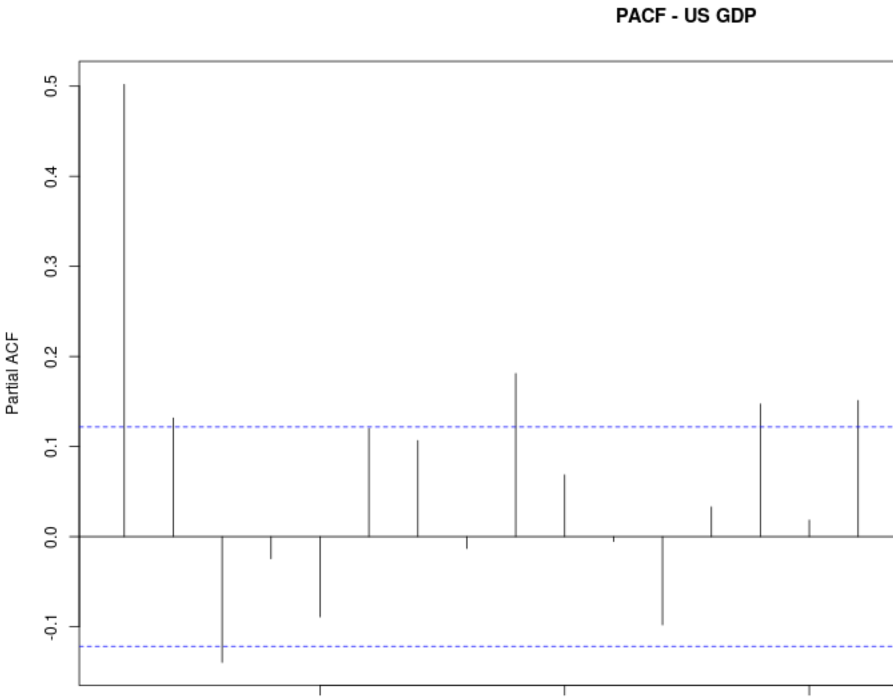
Collect quarterly data from the US on the aggregate GDP (Gross Domestic Product) for the period 1947:I to 2007:II. Calculate the log difference of the GDP series $(\ln(Y_t) - \ln(Y_{t-1}))$. You want to model this series using time series, and apply the Box Jenkins methodology to select a reasonable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

```
> GDP <- read.table("../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
 1.000  0.502  0.350  0.116  0.033 -0.084  0.002  0.067  0.
    11     12     13     14     15     16     17     18
 0.200  0.062  0.049  0.067  0.068  0.187  0.195  0.224  0.
    22     23     24
 0.038  0.008  0.062
```

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

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

     1      2      3      4      5      6      7      8
 0.502  0.132 -0.140 -0.025 -0.089  0.120  0.107 -0.013  0.
    12     13     14     15     16     17     18     19
-0.098  0.033  0.147  0.018  0.151  0.017  0.044 -0.004  0.
    23     24
 0.000  0.047
```


ACF - US GDP
ACF
1.0
0.8
0.6
0.4
0.2
0.0


PACF - US GDP
Partial ACF
0.5
0.4
0.3
0.2
0.1
0.0
-0.1

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.

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.

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

```
> GDP <- read.table("../../../../data/usa/us_macro_data/us_
> 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
 0.4307   0.1888  -0.1029   0.0190  -0.1704   0.0926   0.03
      9       10       11       12       13       14
 0.1490   0.0776   0.0716  -0.1384  -0.0195   0.1076  -0.04

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

Forecasting

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.

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 $ARMA(p,q)$.

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

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?

One step ahead forecasts

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

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

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

Two step ahead forecasts

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

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

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

Three step ahead forecasts

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

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

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

Four step ahead forecasts

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

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

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

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

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

Stock prices

When we in finance (and economics) talk about “time series analysis” we usually have in mind the relationship between *past* realizations of a variable, and the *next* realization, i.e. *prediction*. 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:
The theory of efficient markets states (roughly):
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 a well informed market can not do better than the market.

Random Walk Model

This statement needs to be formalized in some way to make it testable.

The simplest possible such formalization is the

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

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: the path of a drunk left in the middle of a field.

Expanding this

$$\begin{aligned} P_t &= P_{t-1} + \varepsilon_t \\ &= P_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= P_{t-3} + \varepsilon_{t-2} + \varepsilon_{t-1} + \varepsilon_t \\ &= P_0 + \sum_{j=0}^{t-1} \varepsilon_{t-j} \end{aligned}$$

Observe that in the Random Walk model, the effect of a shock

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

$$E[R_t] = \mu$$

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

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

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

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

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