Time series analysis

Models and forecasting

Definition

- ***A time series*** Y_t is a sequence of numerical data points $y_{t_1}, y_{t_2}, \dots, y_{t_n}$ taken at successive points in time $t_1, t_2, \dots, t_n$, normally at uniform intervals.

- We consider ***a discrete time series*** with ***a fixed interval*** called by a step (hour, day, week, month, year, ect).

- ***Examples*** of time series are counts of sunspots, and the daily closing value of the Dow Jones Industrial Average.

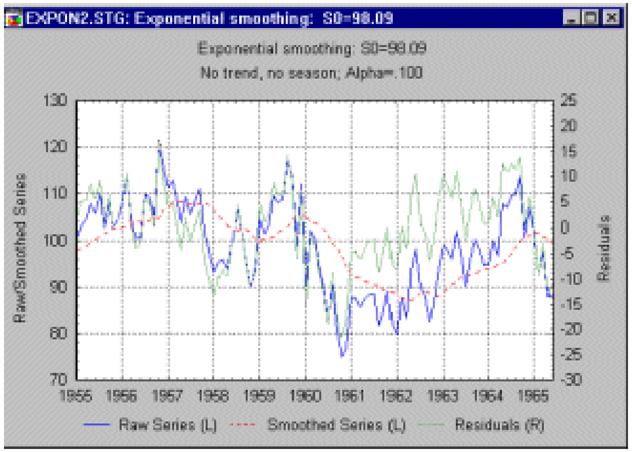
- ***Applications:*** earthquake prediction, electroencephalography, control engineering, astronomy, communications engineering, ect.

❑ A time series can be taken on any variable that changes over time. This can be tracked over the short term, or the long term.

Types of time series

- ***deterministic time series***

- ***stochastic time series***



Time series: basic definitions

- This data in forecasting is usually in the form of a **time series**.
- Suppose that there are T periods of data available, with period T being the most recent.
- We will let the observation on this variable at time period t be denoted by y_t, $t = 1, 2, \ldots, T$. This variable can represent a cumulative quantity, such as the total demand for a product during period t or an instantaneous quantity, such as the daily closing price of a specific stock on the New York Stock Exchange.
- Generally, we will need to distinguish between a **forecast** or **predicted value** of y_t that was made at some previous time period, say, $t - \tau$, and a fitted value of y_t that has resulted from estimating the parameters in a time series model to historical data.
- Note that τ is the forecast lead time. The forecast made at time period $t - \tau$ is denoted by $\hat{y}_t(t - \tau)$. There is a lot of interest in the **lead** - 1 forecast, which is the forecast of the observation in period t, y_t, made one period prior $\hat{y}_t(t - 1)$. We will denote the fitted value of y_t by $\hat{y}_t$.

We will also be interested in analyzing forecast errors. The forecast error that results from a forecast of y_t that was made at time period $t - \tau$ is the lead $-\tau$ forecast error

$$\varepsilon_t(\tau) = y_t - \hat{y}_t(t - \tau). \tag{1}$$

For example, the lead - 1 forecast is

$$\varepsilon_t(1) = y_t - \hat{y}_t(t - 1).$$

The difference between the observation y_t and the value obtained by fitting a time series model to the data, or a fitted value $\hat{y}_t$ defined above, is called a **residual**, and is denoted by

$$\varepsilon_t = y_t - \hat{y}_t$$

The reason for this careful distinction between forecast errors and residuals is that models usually fit historical data better than they forecast. That is, the residuals from a model-fitting process will almost always be smaller than the forecast errors that are experienced when that model is used to forecast future observations.

Time series in R

- The number of international passenger bookings (in thousands) per month on an airline (Pan Am) in the United States were obtained from the Federal Aviation Administration for the period 1949–1960 (Brown, 1963). The company used the data to predict future demand before ordering new aircraft and training aircrew.
- The data are available as a time series in R and illustrate several important concepts that arise in an exploratory time series analysis.

Use functions in R

```
data(AirPassengers)
AP <- AirPassengers
AP
```

"AirPassengers" example

```
> plot(AP)
> data(AirPassengers)
> AP <- AirPassengers
> AP
     Jan Feb Mar Apr May Jun Jul Aug Sep Oct Nov Dec
1949 112 118 132 129 121 135 148 148 136 119 104 118
1950 115 126 141 135 125 149 170 170 158 133 114 140
1951 145 150 178 163 172 178 199 199 184 162 146 166
1952 171 180 193 181 183 218 230 242 209 191 172 194
1953 196 196 236 235 229 243 264 272 237 211 180 201
1954 204 188 235 227 234 264 302 293 259 229 203 229
1955 242 233 267 269 270 315 364 347 312 274 237 278
1956 284 277 317 313 318 374 413 405 355 306 271 306
1957 315 301 356 348 355 422 465 467 404 347 305 336
1958 340 318 362 348 363 435 491 505 404 359 310 337
1959 360 342 406 396 420 472 548 559 463 407 362 405
1960 417 391 419 461 472 535 622 606 508 461 390 432
>
```

"AirPassengers" example

All data in R are stored in objects, which have a range of methods available. The class of an object can be found using the class function:

Use functions in R

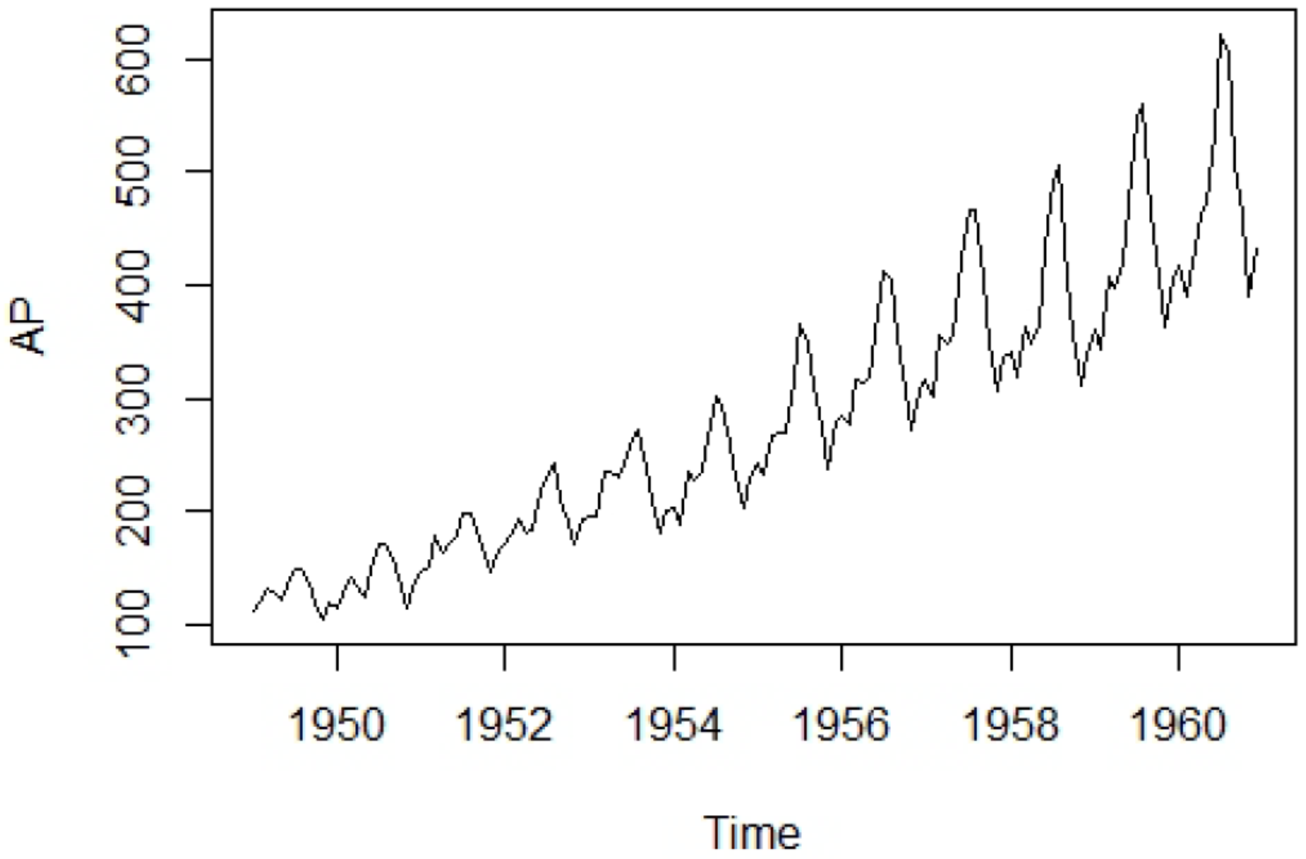
```
class(AP)
start(AP); end(AP); frequency(AP)
```

```
> class(AP)
[1] "ts"
> start(AP); end(AP); frequency(AP)
[1] 1949    1
[1] 1960   12
[1] 12
> summary(AP)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
  104.0   180.0   265.5   280.3   360.5   622.0
> plot(AP)
```

"AirPassengers" example

- In this case, the object is of class `ts`, which is an abbreviation for 'time series'.
- Time series objects have a number of methods available, which include the functions `start` (starting date), `end` (ending date), and `frequency` (period). These methods can be listed using the function methods. In `frequency`: Annual=1, Quarterly=4, Monthly=12, Weekly=52.
- The key thing to bear in mind is that generic functions in R, such as `plot` or `summary`, will attempt to give the most appropriate output to any given input object.
- If your monthly data is already stored as a numerical vector z, then it can be converted to a ts object like this:

"AirPassengers" example



Here, there is a clear and increasing trend. There is also a strong seasonal pattern that increases in size as the level of the series increases.

Stationary time series

Denote a member of a time series at time moment t by y_t.

A stochastic process is called ***strictly stationary*** if a joint probability distribution of m observations $y_{t_1}, y_{t_2}, \ldots, y_{t_m}$ is the same as for the observations $y_{t_1+k}, y_{t_2+k}, \ldots, y_{t_m+k}$ for arbitrary time moments $t_1, t_2, \ldots, t_m$ and $k > 0$.

In other words, time series properties do not change when you change the origin.

A stochastic process is called ***weakly stationary*** if :

- its mathematical expectation and variance are constant over time, so

$$E(y_t) = \mu, \qquad Var(y_t) = E(y_t - \mu)^2 = \sigma_y^2;$$

- a value of the covariance between two observations of a time series depends only on the distance between these observation (the lag).

Autocovariance:

The covariance between the values of y_t and y_{t+k} is called *the autocovariance* with lag k and determined:

$$\rho_k = \gamma_k = \mathrm{cov}(y_t, y_{t+k}) = E[(y_t - \mu)(y_{t+k} - \mu)]$$

$$\gamma_0 = \sigma_y^2 \qquad E(y_t y_{t+k}) - E(y_t) \cdot E y_{t+k}$$

Property of "mean reversion":
the values of a stationary time series fluctuate around mean μ

σ_y^2 - swing of the fluctuation.

Autocorrelation function - ACF

The covariance between y_t and its value at another time period, say, y_{t+k} is called the **autocovariance** at lag k, defined by

$$\gamma_k = Cov(y_t, y_{t+k}) = E\{(y_t - \mu)(y_{t+k} - \mu)\}.$$

The collection of the values of y_k, $k = 0, 1, 2, \ldots$ is called the **autocovariance function**. Note that the autocovariance at lag $k = 0$ is just the **variance of the time series**, that is, $\gamma_0 = \sigma_y^2$.

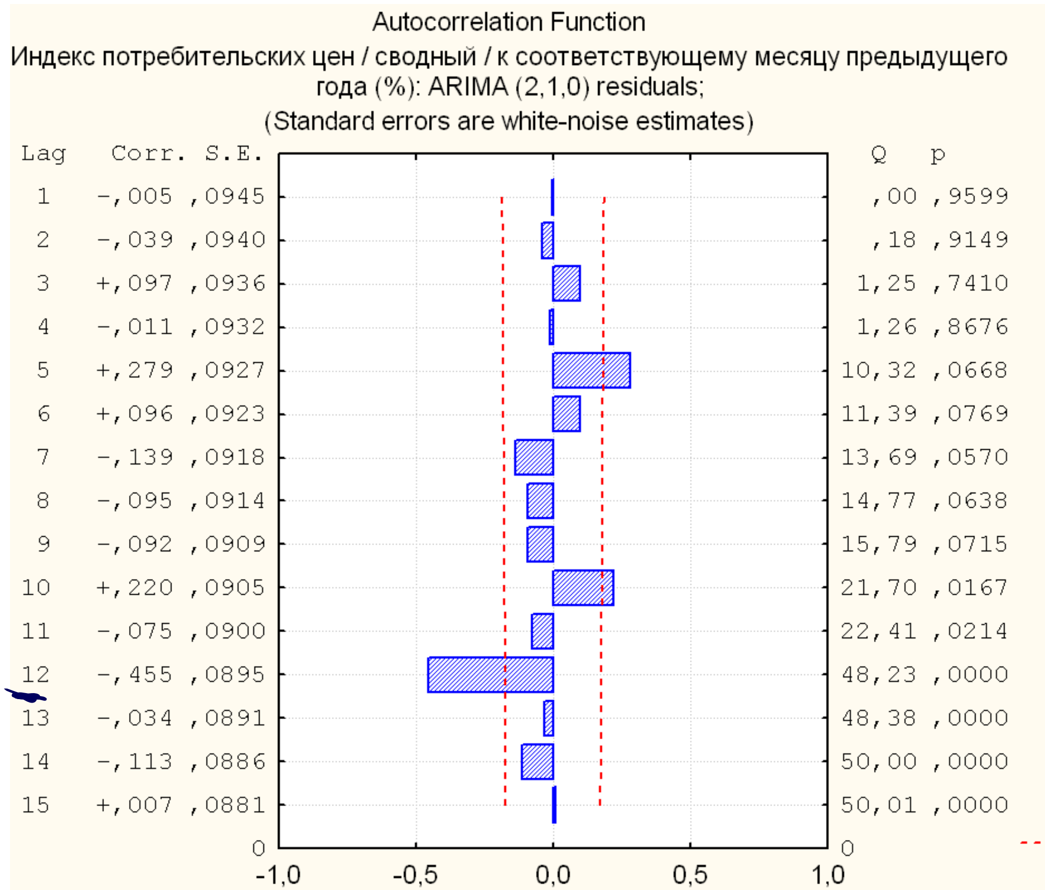
The autocorrelation coefficient at lag k is

$$\rho_k = \frac{E\{(y_t - \mu)(y_{t+k} - \mu)\}}{\sqrt{E(y_t - \mu)^2}\sqrt{E(y_{t+k} - \mu)^2}} = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} = \frac{\gamma_k}{\gamma_0}.$$

Autocorrelation function - ACF

Autocorrelation function $\rho_k = \rho(k)$

Graph of ACF is called ***a correlogram.***



Autocorrelation function with lag k is determined as follows:

$$\rho_k = \frac{E[(y_t - \mu)(y_{t+k} - \mu)]}{\sigma_y^2},$$

$$\rho_k = \frac{\gamma_k}{\gamma_0}.$$

Here $-1 \le \rho_k \le 1$

Autocorrelation function - ACF

The collection of the values of ρ_k, $k = 0, 1, 2, \ldots$ is called the **autocorrelation function (ACF)**. Note that by definition $\rho_0 = 1$. Also, the ACF is independent of the scale of measurement of the time series, so it is a dimensionless quantity. Furthermore,

$$\rho_k = \rho_{-k},$$

that is, the autocorrelation function is **symmetric** around zero. So it is only necessary to compute the positive (or negative) half.

If a time series has a finite mean and autocovariance function, it is said to be second order stationary (or weakly stationary of order 2). If, in addition, the joint probability distribution of the observations at all times is multivariate normal, then that would be sufficient to result in a time series that is strictly stationary.

Autocorrelation function - ACF

It is necessary to estimate the autocovariance and autocorrelation functions from a time series of finite length, say, $y_1, \ldots, y_T$. The usual estimate of the autocovariance function is

$$c_k = \hat{\gamma}_k = \frac{1}{T}\sum_{t=1}^{T}(y_t - \bar{y})(y_{t+k} - \bar{y}), \quad k = 0, 1, \ldots, K,$$

and the autocorrelation function is estimated by the **sample autocorrelation function** (or **sample ACF**)

$$r_k = \hat{\rho}_k = \frac{c_k}{c_0}, \quad k = 0, 1, \ldots, K.$$

A good general rule of thumb is that at least 50 observations are required to give a reliable estimate of the ACF, and the individual sample autocorrelations should be calculated up to lag K, where K is about $T/4$.

ACF in R

These correspond to scatterplots given above. The autocorrelation coefficients are plotted to show the autocorrelation function or ACF. The plot is also known as a **correlogram**.

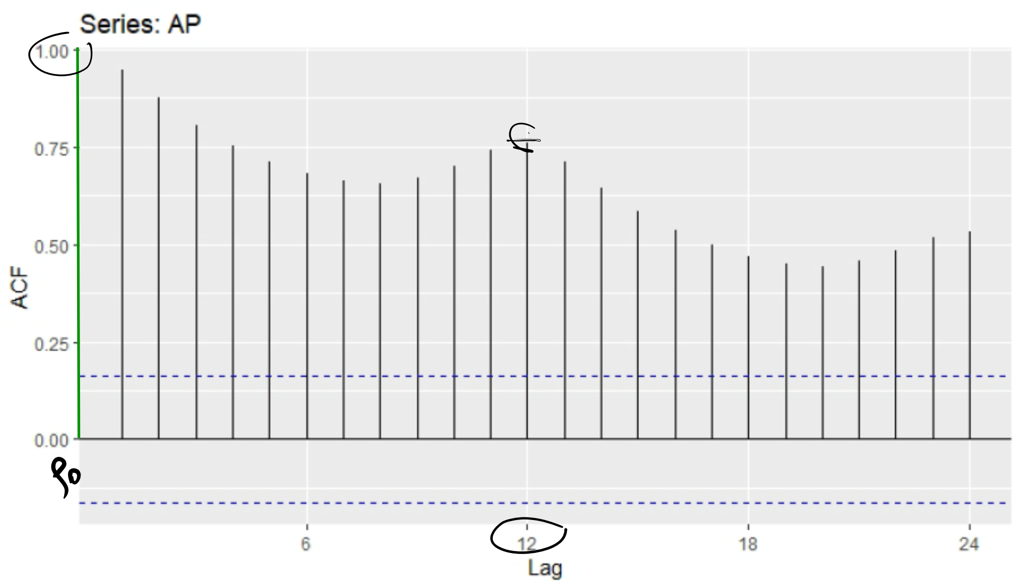
Use function in R

```
acf(AP)
```

or

```
ggAcf(AP)
```

$\rho_0 = r_{y_t y_t} = 1$



- In this graph: $r_0 = 1$, r_{12} shows high autocorrelation. This is due to the seasonal pattern in the data: the peaks tend to be 12 months.
- The dashed blue lines indicate the confidence interval to check if the correlations are significantly different from zero. Outside CI we have significant correlations.
- The slow decrease in the ACF as the lags increase is due to the trend, while the "scalloped" shape is due the seasonality.
- When data have a trend, the autocorrelations for small lags tend to be large and positive because observations nearby in time are also nearby in size. So the ACF of trended time series tend to have positive values that slowly decrease as the lags increase.
- When data are seasonal, the autocorrelations will be larger for the seasonal lags (at multiples of the seasonal frequency) than for other lags.
- When data are both trended and seasonal, you see a combination of these effects.

Test for $H_0 : \rho_1 = \ldots = \rho_m = 0$

We use test by Box and Pierce (1970) for hypothesis:

- $H_0 : \rho_1 = \ldots = \rho_m = 0$.
- H_1 : at least one of $\rho_1, \ldots, \rho_m$ is not equal to zero.

We calculate Q-statistic in the form:

$$Q_1 = \frac{1}{T} \sum_{k=1}^{m} r_k^2$$

where T = sample size, m = maximum lag length.
The correlation coefficients are squared so that the positive and negative coefficients do not cancel each other out. The statistics is distributed by χ^2-statistics with degrees of freedom equal to the number of squares in the sum, that is m.

Use function in R

```
Box.test(AP, lag = 5, type = "Box-Pierce")
```

Test for $H_0 : \rho_1 = \ldots = \rho_m = 0$

We can use another test by Ljung and Box (1978) for the same hypothesis calculating the statistic given by

$$Q_2 = T(T+2)\sum_{k=1}^{m} \frac{r_k^2}{T-k} \sim \chi_m^2.$$

Use function in R

```
Box.test(AP, lag = 5, type = "Ljung-Box")
```

```
> Box.test(AP, lag = 5, type = "Box-Pierce")

        Box-Pierce test

data:  AP
X-squared = 488.46, df = 5, p-value < 2.2e-16
```

$< 0,05$ $\cancel{H_0}$

```
> Box.test(AP, lag = 5, type = "Ljung-Box")

        Box-Ljung test

data:  AP
X-squared = 504.8, df = 5, p-value < 2.2e-16
```

$< 0,05$ $\cancel{H_0}$

Strictly vs. weakly stationary process

These statements are true:

- If a process is strictly stationary, then it is weakly stationary.
- If a process is weakly stationary, then it is not strictly stationary in general case.
- The process is called normal or Gaussian if all its multiple distribution functions are normal.
- For normal processes strictly stationarity follows from weakly stationarity.

Partial Autocorrelation function - PACF

In the case of an autoregressive process of order p, there will be direct connections between y_t and y_{t-s} for $s \leqslant p$, but no direct connections for $s > p$.
For example, consider the following $AR(3)$ model

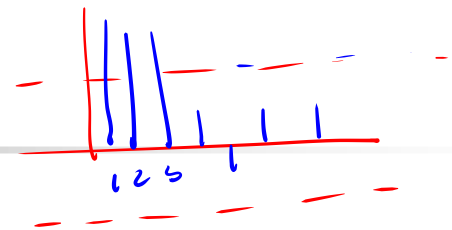
$$y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \phi_3 y_{t-3} + \overset{=\varepsilon_t}{u_t}.$$

There is a direct connection through the model between y_t and y_{t-1}, and between y_t and y_{t-2}, and between y_t and y_{t-3}, but not between y_t and y_{ts}, for $s > 3$.
Hence the **partial autocorrelation function** will usually have non-zero partial autocorrelation coefficients for lags up to the order of the model, but will have zero partial autocorrelation coefficients thereafter. In the case of the $AR(3)$, only the first three partial autocorrelation coefficients will be non-zero.
The partial autocorrelation coefficient at lag k is

$$\rho_k = \frac{E\{(y_t - \mu)(y_{t+k} - \mu)\}}{\sqrt{E(y_t - \mu)^2}\sqrt{E(y_{t+k} - \mu)^2}} = \frac{Cov(y_t, y_{t+k})}{Var(y_t)} = \frac{\gamma_k}{\gamma_0}.$$

Partial Autocorrelation function - PACF

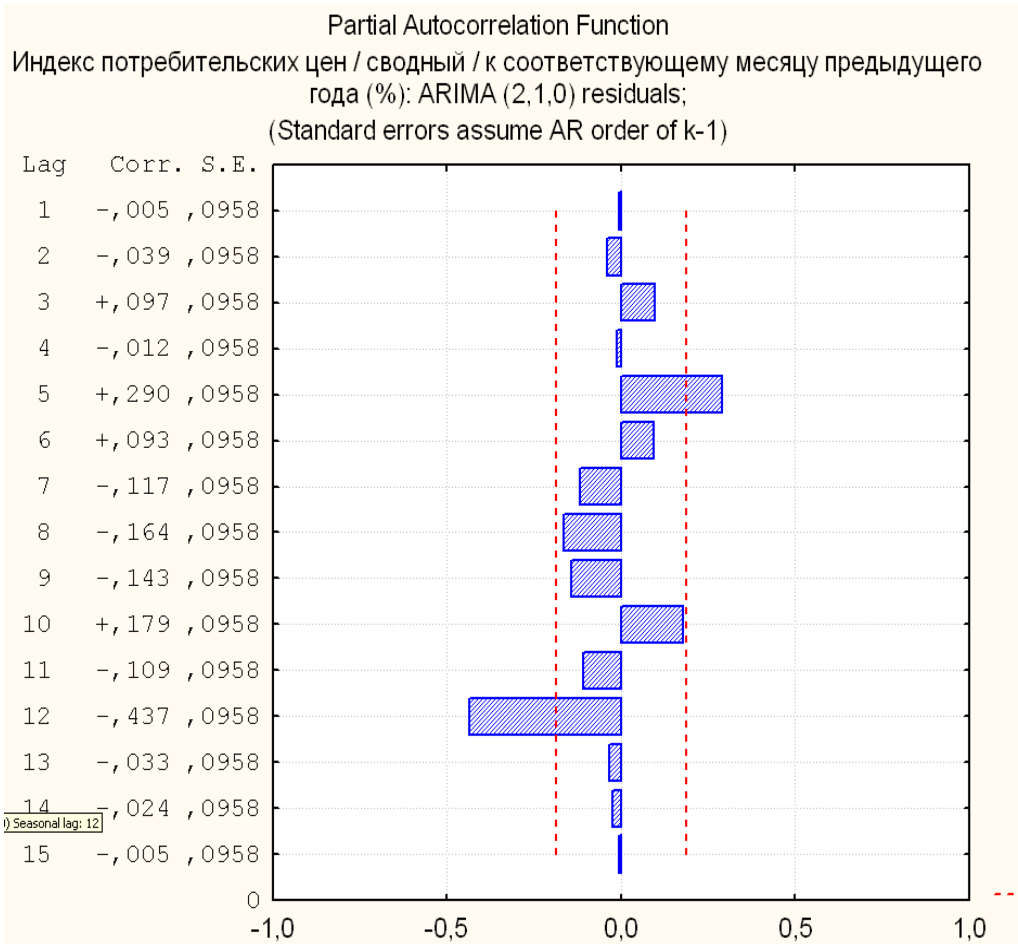
PACF



- Specifically, partial autocorrelations are useful in identifying the order of an autoregressive model.
- The partial autocorrelation of $AR(p)$ process is zero at lag $p + 1$ and greater.
- If the sample autocorrelation plot indicates that an AR model may be appropriate, then the sample partial autocorrelation plot is examined to help identify the order.
- We look for the point on the plot where the partial autocorrelations essentially become zero. Placing a 95% confidence interval for statistical significance is helpful for this purpose.
- The approximate 95% confidence interval for the partial autocorrelations are at $(-2/\sqrt{T}, 2/\sqrt{T})$.

Partial Autocorrelation function - PACF

Partial Autocorrelation function reflects the correlations of time series members separated by k steps that are "purified" from the mediated effect of the intermediate members.



Partial Autocorrelation function with lag k

$$\rho_k^{part} = \rho^{part}(k),$$

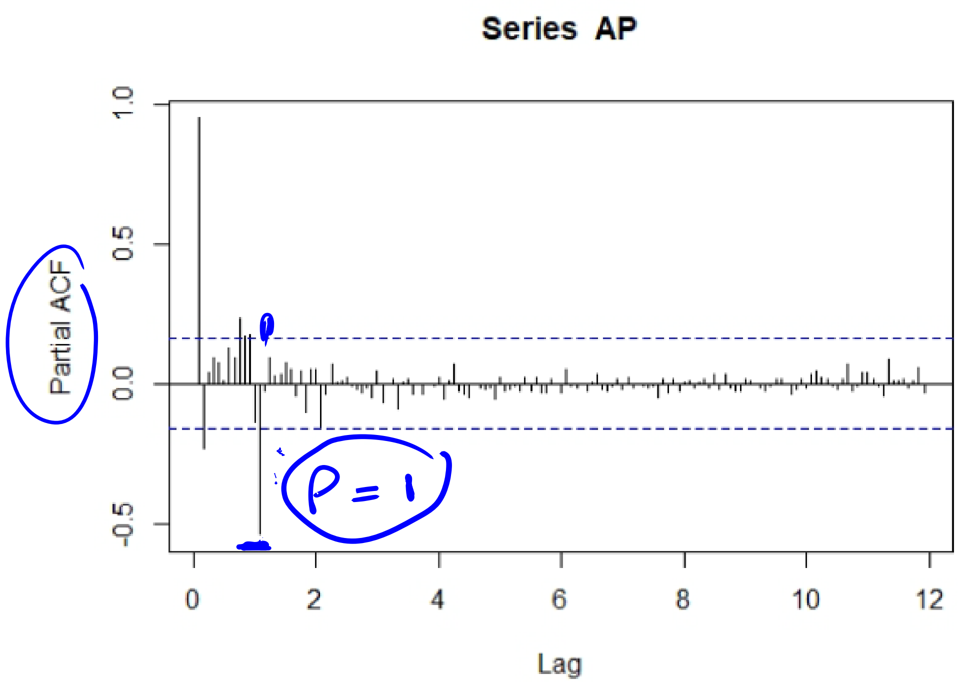
$$-1 \le \rho_k \le 1.$$

ACF and PACF are involved in time series analysis for selecting and identifying a proper model.

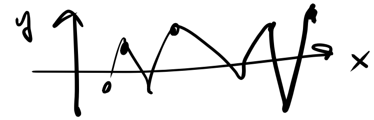
Partial Autocorrelation function - PACF

Use function in R

```
library(tseries)
pacf (AP, lag=length(AP)-1)
```



“White” noise



White noise is a discrete stochastic process whose samples are regarded as a sequence of serially uncorrelated random variables with zero mean and finite (constant) variance.

For white noise $\{\varepsilon_t\}_{t\in T}$ we have $E(\varepsilon_t)=0,\qquad Var(\varepsilon_t)=\sigma^2.$

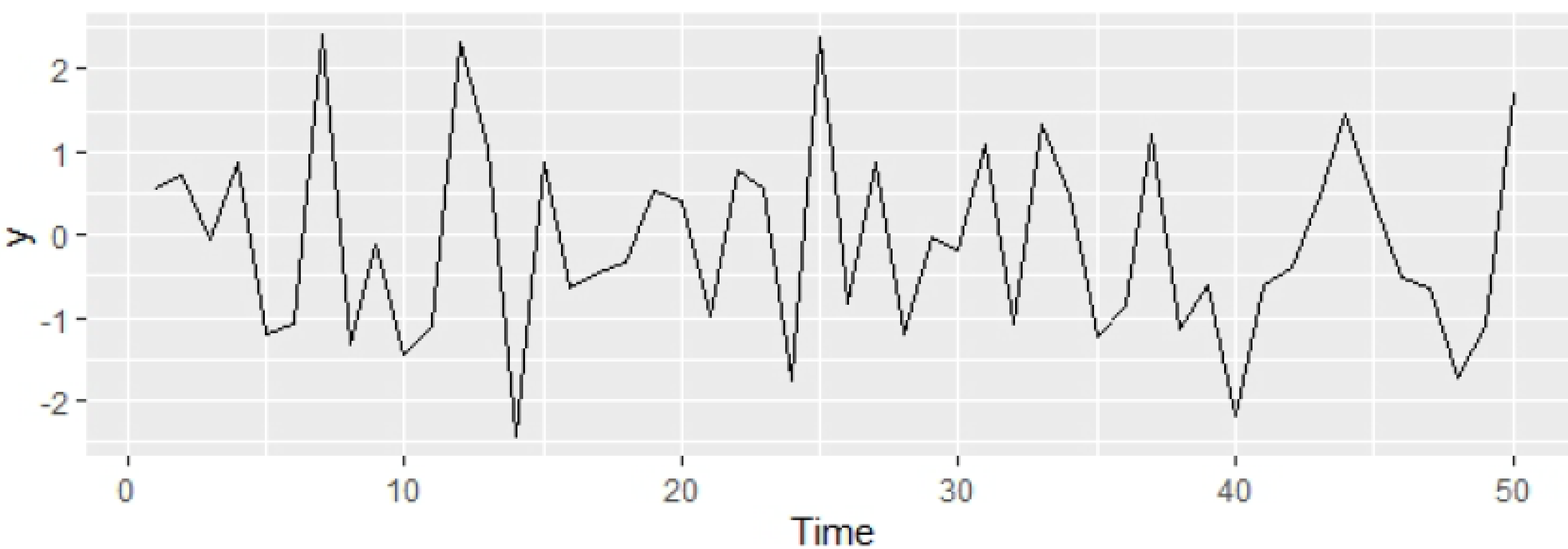
Serially uncorrelated random variables means that two any subsamples (intersecting or disjoint) are independent.

- A single realization of white noise is a random shock.
- Depending on the context, one may also require that the samples be independent and have identical probability distribution. So, i.i.d. is a simplest representative of white noise.
- In particular, if each sample has a normal distribution with zero mean, the stochastic process is said to be Gaussian white noise.

Modeling white noise in R

Use function in R

```
library(forecast)
y <- ts(rnorm(50)) # model a random sample of size 50
autoplot(y)
ggAcf(y)
```



$\rho_1 > 0{,}7$.

Autoregressive process of order *p* - *AR(p)*

white noise

AR(1): $y_t = \phi_0 + \phi_1 y_{t-1} + \varepsilon_t,$ $\quad \varepsilon_t \approx iid(0, \sigma^2), \quad t = 1, \ldots, n.$

AR(2): $y_t = \phi_0 + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \varepsilon_t,$ $\quad |\rho_1| > 0{,}7 \quad |\rho_2| > 0{,}7$

...

AR(p): $y_t = \phi_0 + \phi_1 y_{t-1} + \ldots + \phi_p y_{t-p} + \varepsilon_t,$

It can be shown that the process of *AR(1)* is stationary if $|\phi_1| < 1,$
i.e. you must define the following characteristics

$$E(y_t) - ? \quad Var(y_t) - ? \quad Cov(y_t, y_{t-k}) - ?$$

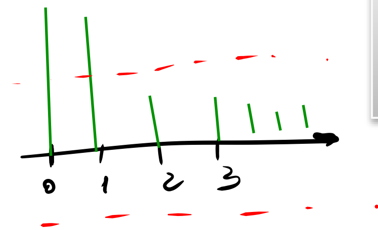
Then its autocorrelation function is $\rho_1 = \phi_1, \quad \rho_2 = \phi_1^2, \ \ldots, \ \rho_k = \phi_1^k.$

Autoregressive processes of orders 1 and 2
AR(1) and AR(2)

For ***AR(1):*** a time series is stationary if $|\phi_1| < 1$.

From Yule-Walker's system of equations

ACF: $\rho_1 = \phi_1, \quad \rho_2 = \phi_1^2, \; \ldots, \rho_k = \phi_1^k$



ACF ***decreases sharply*** after the first few values.

PACF(1)=ACF(1) – by definition. *PACF* values ***equal 0 for k>1,*** i.e.

$$\rho_1^{part} = \rho_1, \quad \rho_k^{part} = 0, \;\; k > 1.$$

For ***AR(2):*** a time series is stationary if $|\phi_1| < 1, \;\; \phi_1 + \phi_2 < 1, \;\; \phi_2 - \phi_1 < 1.$

Let $\phi_0 = 0.$

ACF: $\rho_1 = \dfrac{\phi_1}{1-\phi_2}, \;\; \rho_2 = \dfrac{\phi_1^2}{1-\phi_2} + \phi_2, \;\; \rho_k = \phi_1 \rho_{k-1} + \phi_1 \rho_{k-2}, \;\; k = 1,2,\ldots$

ACF ***exponentially decreases*** in the case the roots of characteristic polynomial are real, or ***varies in a sine wave with an exponentially decreasing amplitude***, if the roots are complex. *PACF(k)* values ***equal 0 for k>2.***

An **autoregressive model** is one where the current value of a variable, y_t, depends upon only the values that the variable took in previous periods plus an error term. An autoregressive model of order p, denoted as $AR(p)$, can be expressed as

$$y_t = \mu + u_t + \phi_1 t_{t-1} + \phi_2 y_{t-2} + \ldots + \phi_p y_{t-p},$$

where u_t is a white noise disturbance term.
For an $AR(1)$ model:

1. When $\phi_1 = 0$, then y_t is equivalent to white noise.
2. When $\phi_1 < 0$, then y_t tends to oscillate around the mean.

We normally restrict autoregressive models to stationary data, in which case some constraints on the values of the parameters are required.

1. For an $AR(1)$ model: $-1 < \phi_1 < 1$.
2. For an $AR(2)$ model: $-1 < \phi_2 < 1$, $\phi_1 + \phi_2 < 1$, $\phi_2 - \phi_1 < 1$.

When $p \geqslant 3$, the restrictions are much more complicated. R takes care of these restrictions when estimating a model.

Shift operator and unit roots

$y_t = \varphi_0 + \varphi_1 y_{t-1} + \varepsilon_t$

$y_t - \varphi_0 - \varphi_1 y_{t-1} = \varepsilon_t$

We introduce the shift operator L such that $Ly_t = y_{t-1}$.

$\varphi(L) y_t = (1 - \varphi_1 L) y_t$

Then, *AR(1)*-model can be represented as

$$\Phi(L)y_t = (1-\phi_1 L)y_t = \varepsilon_t, \quad \varepsilon_t \approx iid\,(0,\sigma^2).$$

AR(2)-model is represented in the form

$$\Phi(L)y_t = (1-\phi_1 L - \phi_2 L^2)y_t = \varepsilon_t, \quad \varepsilon_t \approx iid\,(0,\sigma^2).$$

Polynomial $\Phi(L)$ can be factorized in this way:

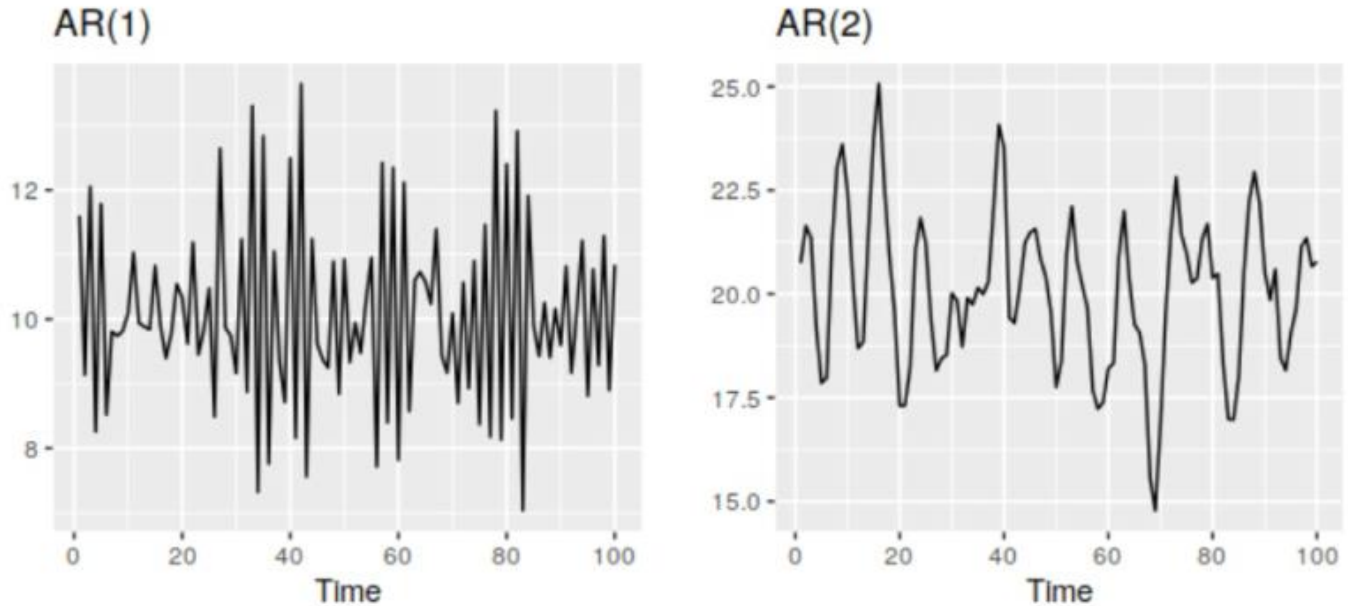
$$\Phi(L) = (1-\phi_1 L - \phi_2 L^2) = (1-\lambda_1 L)(1-\lambda_2 L).$$

In order to the inverse operator would exist it is required that all $|\lambda_i| < 1$.

Or all roots $w_i = \dfrac{1}{\lambda_i}$ of the polynomial $\Phi(L)$ must lie outside the unit disk.

AR(p) processes: examples

AR(p)



$AR(1)$: $y_t = 8 - 0.8y_{t-1} + u_t$,
$AR(2)$: $y_t = 18 + 1.3y_{t-1} - 0.7y_{t-2} + u_t$.

Moving average processes

Rather than using past values of the forecast variable in a regression, a moving average model uses past forecast errors in a regression-like model.

$$y_t = \mu + u_t + \theta_1 u_{t-1} + \theta_2 u_{t-2} + \ldots + \theta_q u_{t-q},$$

where u_t is a white noise ($Eu_t = 0$, $var(u_t) = \sigma^2$ for any t). We refer to this as an $MA(q)$ model, a **moving average model of order** q. Of course, we do not observe the values of u_t, so it is not really a regression in the usual sense.

Remark

Notice that each value of y_t can be thought of as a weighted moving average of the past few forecast errors. However, moving average models should not be confused with the moving average smoothing we discussed above. A moving average model is used for forecasting future values, while moving average smoothing is used for estimating the trend-cycle of past values.

A moving average model is simply a linear combination of white noise processes, so that y_t depends on the current and previous values of a white noise disturbance term.

The process of moving average of order *q* - *MA(q)*

white noise

MA(1): $y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1},$ $\varepsilon_t \approx iid(0, \sigma^2),$ $t = 1, \ldots, n.$

$y_t = \delta + u_t - \theta_1 u_{t-1}$

...

MA(q): $y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \theta_2 \varepsilon_{t-2} - \ldots - \theta_q \varepsilon_{t-q}.$

It can be shown that the process of *MA(q)* is stationary with $\forall q$ and $\forall \theta_i$.

Let us formulate the reversibility condition for MA-process - presenting it in the form of AR-process.

The process of moving average of order *q* - *MA(q)*

AR(p)	***MA(q)***
(PACF, ACF) ⟺	*(ACF, PACF)*

Consider ***MA(1)*:** $y_t = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} = \delta + \Theta(L)\varepsilon_t,\qquad \varepsilon_t \approx iid(0,\sigma^2)$

It can be presented as ***AR(∞)*:** $\Theta(L)^{-1} y_t = \Theta(L)^{-1}\delta + \varepsilon_t$

In case ***the reversibility condition*** holds: $|\theta_1| < 1.$

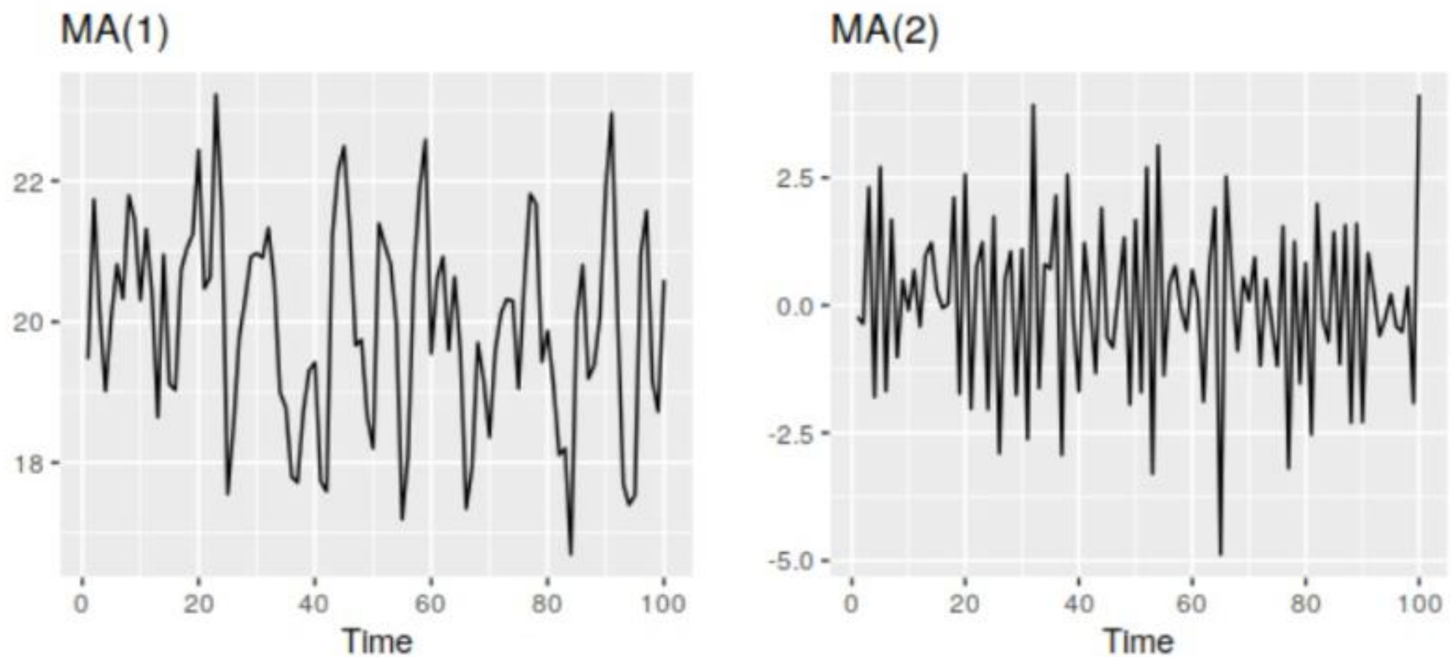
Свойства ***MA(1)*** процесса: $E(y_t) = \delta,\ Var(y_t) = \sigma_\varepsilon^2(1+\theta_1^2).$

ACF: $$\rho_1 = \frac{-\theta_1}{1+\phi_1^2},\qquad \rho_k = 0,\quad k > 1.$$

For *MA(q)* it holds that $\rho_k = 0,\quad k > q,$

i.e. ***ACF values equal 0 для k>q.***

MA(q) processes: examples



$MA(1)$: $y_t = 20 + u_t + 0.8u_{t-1}$,
$MA(2)$: $y_t = u_t - u_{t-1} + 0.8u_{t-2}$.

STATIONARITY OF STOCHASTIC PROCESS

- It is easier to work with stationary time series.
- $MA(q)$ is always stationary for any q.
- $AR(p)$ may be non-stationary.
- We need to examine conditions for stationarity of the time series.

Theorem 1 (Wold decomposition, 1938)

The weakly stationary stochastic process without constant can be represented in the following form:

$$y_t - \mu = \sum_{\tau=0}^{\infty} \psi_\tau u_{t-\tau},$$

where ψ_τ are constants, $u_{t-\tau}$ is a white noise.

So, to examine any weakly stationary stochastic process or time series we need to examine the properties of linear combination of white noises.

θ ψ_0

Some comments on Wold's theorem:

1. The series in the right-hand side should converge.
2. It is assumed that $\psi_0 = 1$.
3. The larger ψ_τ, the more influence of $u_{t-\tau}$ on y_t.
4. The infinite number of elements in decomposition implies technical problems, but fortunately in many known stochastic processes the number of elements is finite.
5. In $MA(q)$ there are q elements in Wold's decomposition. Therefore, $MA(q)$ is weakly stationary.
6. We need to find conditions under which $AR(p)$ can be represented by Wold decomposition to prove stationarity of $AR(p)$.

Augmented Dickey-Fuller Test (adf test)

H_0: there is a unit root of characteristic polynomial.
H_1: all roots of the characteristic polynomial are greater than 1 (the process is stationary).
If p-value of the test is less than 0.05, we can't reject the stationarity of the process.

Use function in R

```
library(tseries)
adf.test(AP)
```

```
> adf.test(AP)

        Augmented Dickey-Fuller Test

data:  AP
Dickey-Fuller = -7.3186, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

$< 0{,}05$

In the test the lag order is used. It is calculated as $(T-1)^{1/3}$.

The process of autoregression and moving average of orders *p* and *q* - *ARMA(p,q)*

Mixed processes:

$$y_t - \phi_1 y_{t-1} - \dots - \phi_p y_{t-p} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1} - \dots - \theta_q \varepsilon_{t-q}, \qquad \varepsilon_t \approx iid\,(0, \sigma^2),$$

$y_t = \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p} + \delta + \underbrace{\qquad\qquad}$

or $$\Phi(L) y_t = \delta + \Theta(L) \varepsilon_t, \qquad \varepsilon_t \approx iid\,(0, \sigma^2),$$

where $\Phi(L) = 1 - \phi_1 L - \dots - \phi_p L^p$

$\Theta(L) = 1 - \theta_1 L - \dots - \theta_q L^q$ are polynomials of shift operators.

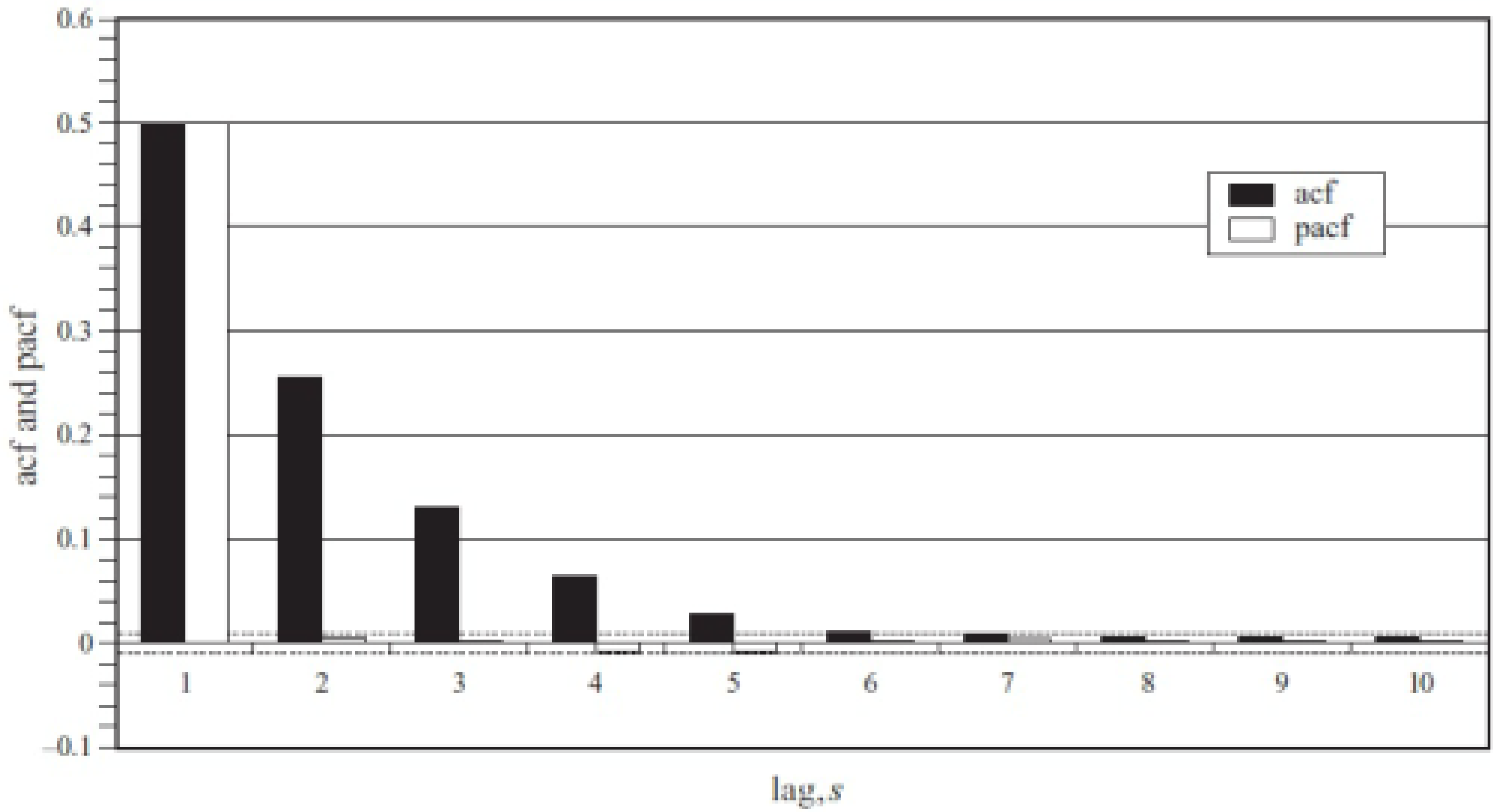
Autoregressive moving average or ***ARMA(p,q)***

Consider *ARMA(1,1):* $$y_t - \phi_1 y_{t-1} = \delta + \varepsilon_t - \theta_1 \varepsilon_{t-1},$$

Let $|\phi_1| < 1, \;\; |\theta_1| < 1.$

Prove that under this condition the process is stationary.

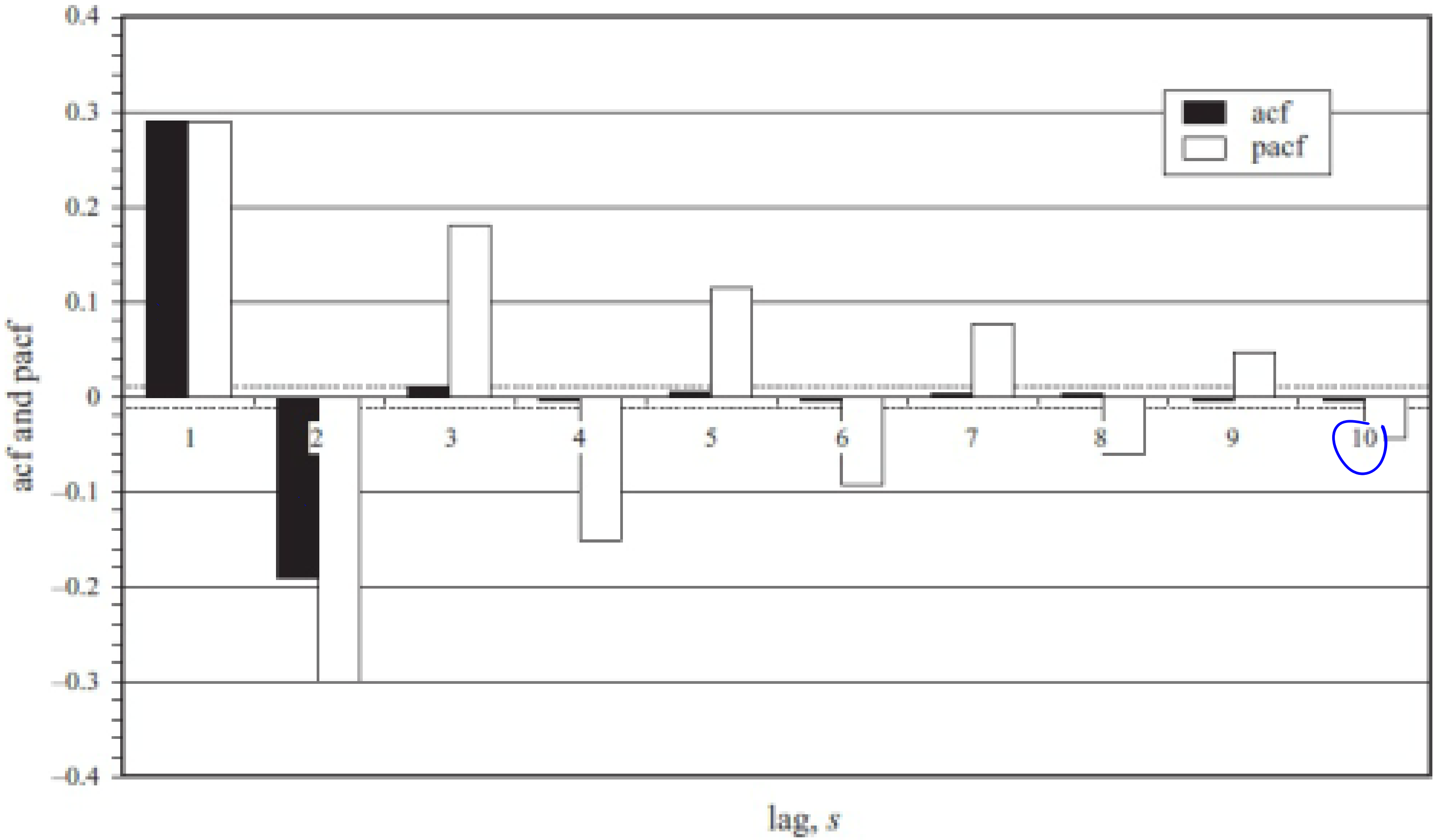
Plot: ACF and PACF in AR model



Sample autocorrelation and partial autocorrelation functions for a decaying $AR(1)$ model: $y_t = 0.5y_{t-1} + u_t$.

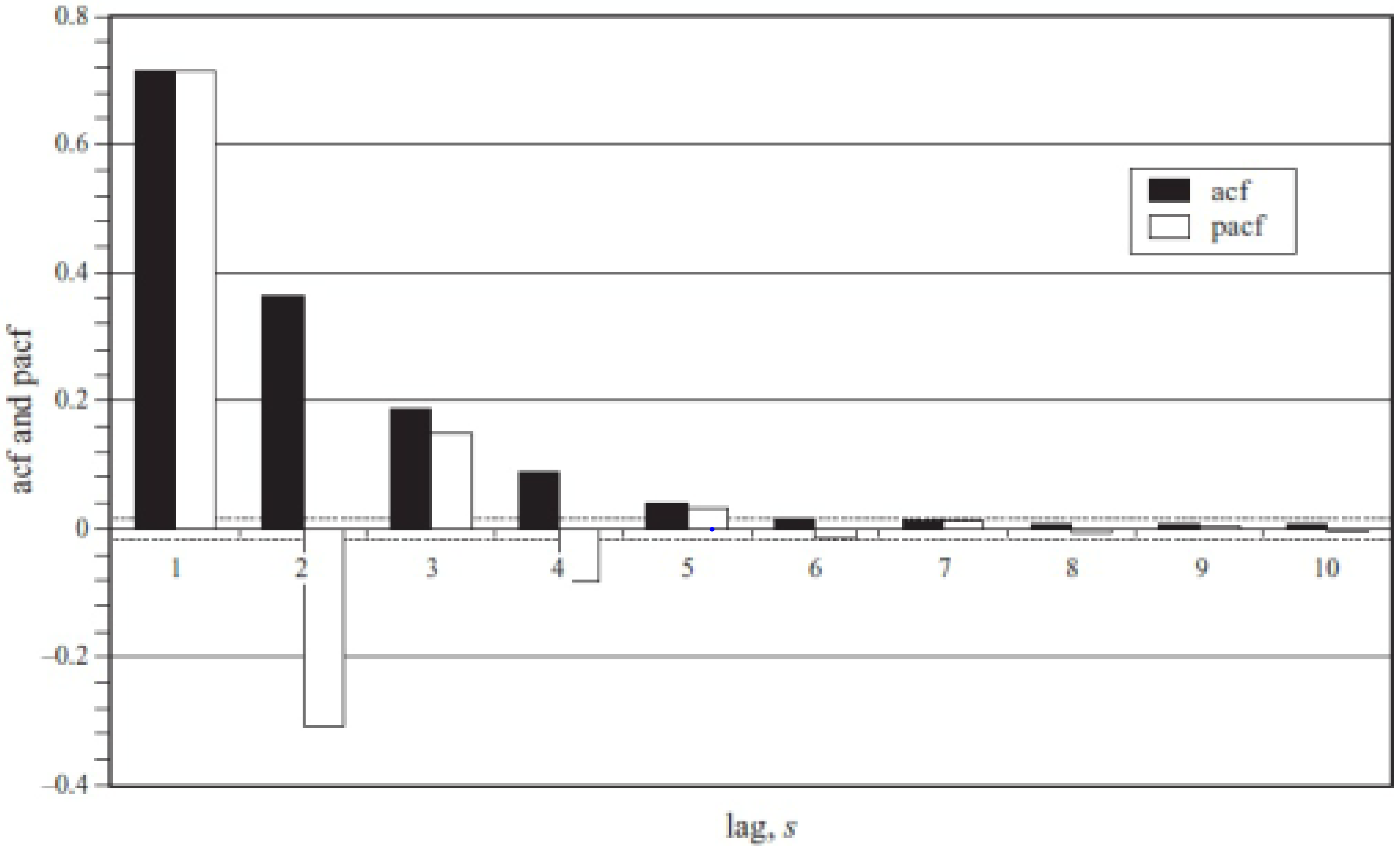
Plot: ACF and PACF in MA model

MA(2)



Sample autocorrelation and partial autocorrelation functions for an $MA(2)$ model: $y_t = 0.5u_{t-1} - 0.25u_{t-2} + u_t$.

Plot: ACF and PACF in ARMA model



Sample autocorrelation and partial autocorrelation functions for an $ARMA(1,1)$ model: $y_t = 0.5y_{t-1} + 0.5u_{t-1} + u_t$

The analysis of a time series stationarity

An analysis of a time series:

- essence of the process;
- the causes;
- consequences;
- methods of detection;
- methods of correction.

Example of a n/s time series:
Random walk:

$$y_t = y_{t-1} + \varepsilon_t, \quad \varepsilon_t \approx iid(0, \sigma^2)$$

Prove that
1. Random walk is non-stationary;
2. After taking the first difference becomes stationary.

Methods of detection:

- Graphical analysis *(trend, seasonality, "spread" over time, and others);*
- ACF is weakly damped *(for stationary series ACF(k) decreases rapidly with increasing k);*
- Applying specified tests:
  - Dickey-Fuller test *(if the ADF-statistic is in the critical region, it might be a case of non-stationarity).*
  - Ljung-Box test *(based on statistic* Q_{LB} *).*

Methods of correction:

- Taking d first differences and/or seasonal differences;
- Discharging of a linear or a nonlinear trend;
- Removing seasonal components.

Typical examples of non-stationary time series

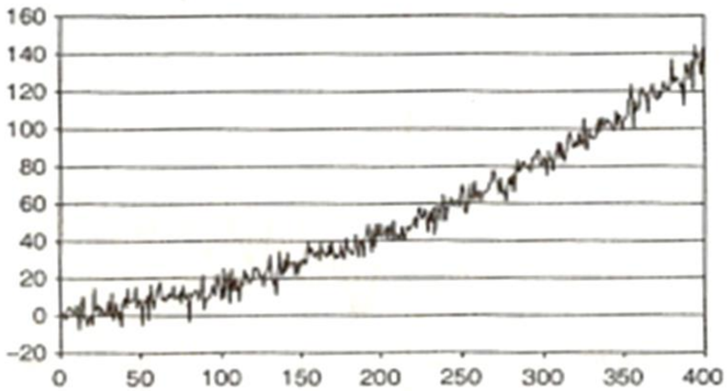


Рис. 12.1. Тренд (модельный пример)

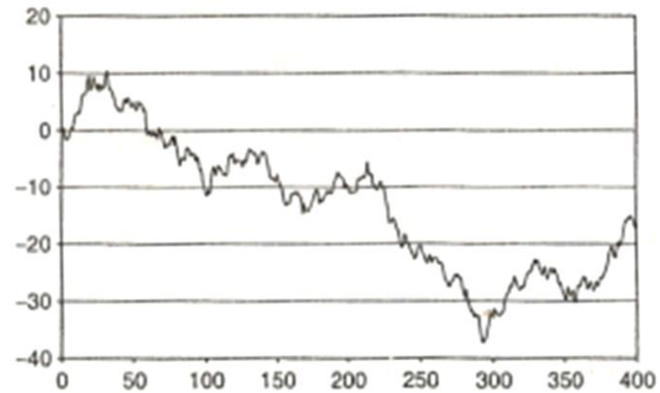


Рис. 12.2. Случайное блуждание (модельный пример)

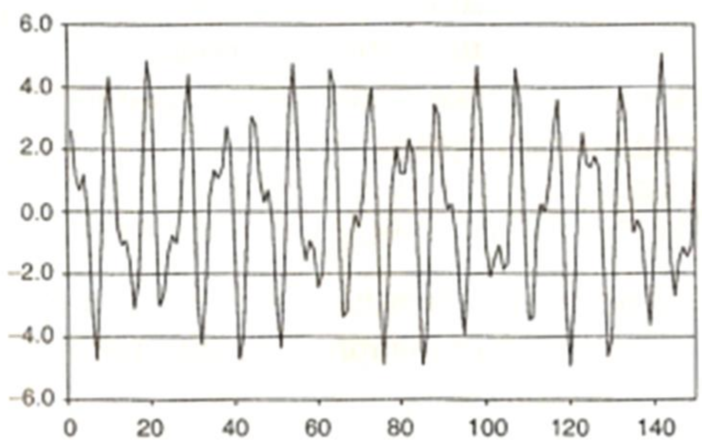


Рис. 12.3. Сезонность (солнечные пятна, числа Вольф

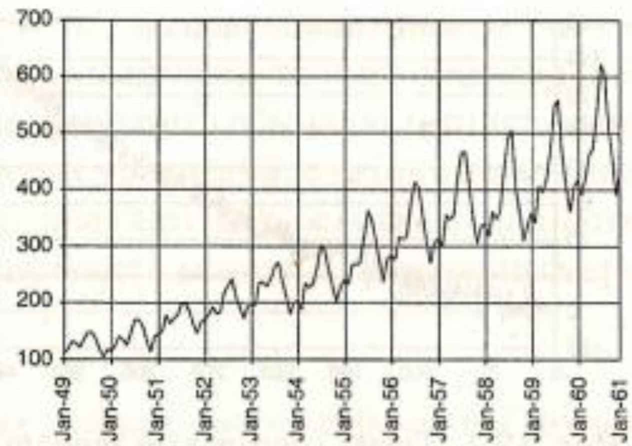
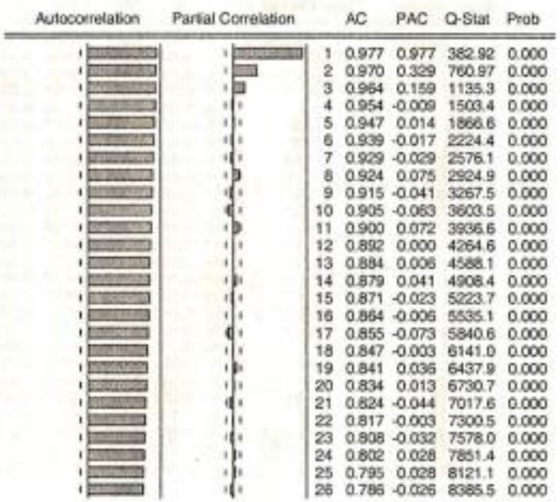


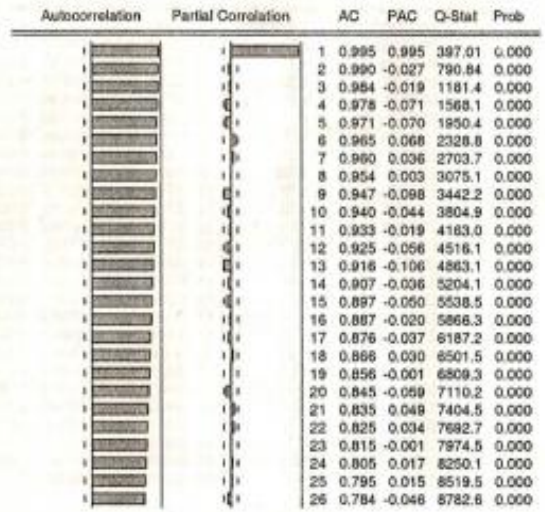
Рис. 12.4. Тренд и сезонность (объем авиаперевозок)

ACF и PACF for typical examples of non-stationary time series



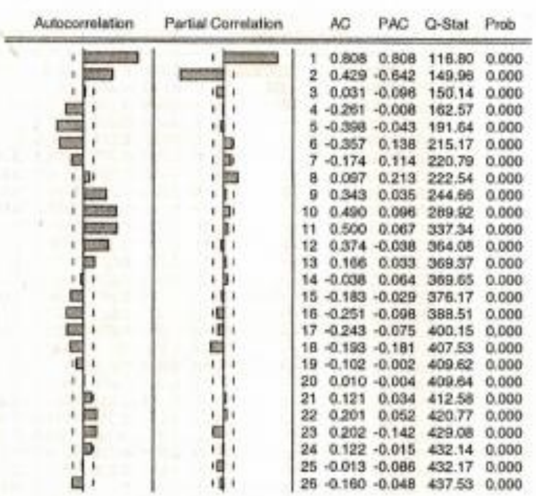
Lag	AC	PAC	Q-Stat	Prob
1	0.977	0.977	382.92	0.000
2	0.970	0.329	760.97	0.000
3	0.964	0.159	1135.3	0.000
4	0.954	-0.009	1503.4	0.000
5	0.947	0.014	1866.6	0.000
6	0.939	-0.017	2224.4	0.000
7	0.929	-0.029	2576.1	0.000
8	0.924	0.075	2924.9	0.000
9	0.915	-0.041	3267.5	0.000
10	0.905	-0.063	3603.5	0.000
11	0.900	0.072	3936.6	0.000
12	0.892	0.000	4264.6	0.000
13	0.884	0.006	4588.1	0.000
14	0.879	0.041	4908.4	0.000
15	0.871	-0.023	5223.7	0.000
16	0.864	-0.006	5535.1	0.000
17	0.855	-0.073	5840.6	0.000
18	0.847	-0.003	6141.0	0.000
19	0.841	0.036	6437.9	0.000
20	0.834	0.013	6730.7	0.000
21	0.824	-0.044	7017.6	0.000
22	0.817	-0.003	7300.5	0.000
23	0.808	-0.032	7578.0	0.000
24	0.802	0.028	7851.4	0.000
25	0.795	0.028	8121.1	0.000
26	0.786	-0.026	8385.5	0.000

Рис. 12.5. Тренд (модельный пример)



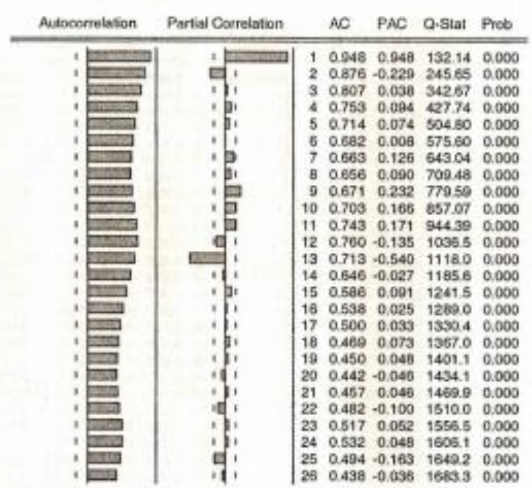
Lag	AC	PAC	Q-Stat	Prob
1	0.995	0.995	397.01	0.000
2	0.990	-0.027	790.84	0.000
3	0.984	-0.019	1181.4	0.000
4	0.978	-0.071	1568.1	0.000
5	0.971	-0.070	1950.4	0.000
6	0.965	0.068	2328.8	0.000
7	0.960	0.036	2703.7	0.000
8	0.954	0.003	3075.1	0.000
9	0.947	-0.098	3442.2	0.000
10	0.940	-0.044	3804.9	0.000
11	0.933	-0.019	4163.0	0.000
12	0.925	-0.056	4516.1	0.000
13	0.916	-0.106	4863.1	0.000
14	0.907	-0.036	5204.1	0.000
15	0.897	-0.050	5538.5	0.000
16	0.887	-0.020	5866.3	0.000
17	0.876	-0.037	6187.2	0.000
18	0.866	0.030	6501.5	0.000
19	0.856	-0.001	6809.3	0.000
20	0.845	-0.059	7110.2	0.000
21	0.835	0.049	7404.5	0.000
22	0.825	0.034	7692.7	0.000
23	0.815	-0.001	7974.5	0.000
24	0.805	0.017	8250.1	0.000
25	0.795	0.015	8519.5	0.000
26	0.784	-0.046	8782.6	0.000

Рис. 12.6. Случайное блуждание (модельный пример)



Lag	AC	PAC	Q-Stat	Prob
1	0.808	0.808	116.80	0.000
2	0.429	-0.642	149.96	0.000
3	0.031	-0.096	150.14	0.000
4	-0.261	-0.008	162.57	0.000
5	-0.398	-0.043	191.64	0.000
6	-0.357	0.138	215.17	0.000
7	-0.174	0.114	220.79	0.000
8	0.097	0.213	222.54	0.000
9	0.343	0.035	244.66	0.000
10	0.490	0.096	289.92	0.000
11	0.500	0.067	337.34	0.000
12	0.374	-0.038	364.08	0.000
13	0.166	0.033	369.37	0.000
14	-0.038	0.064	369.65	0.000
15	-0.183	-0.029	376.17	0.000
16	-0.251	-0.098	388.51	0.000
17	-0.243	-0.075	400.15	0.000
18	-0.193	-0.181	407.53	0.000
19	-0.102	-0.002	409.62	0.000
20	0.010	-0.004	409.64	0.000
21	0.121	0.034	412.58	0.000
22	0.201	0.052	420.77	0.000
23	0.202	-0.142	429.08	0.000
24	0.122	-0.015	432.14	0.000
25	-0.013	-0.086	432.17	0.000
26	-0.160	-0.048	437.53	0.000

Рис. 12.7. Сезонность (солнечные пятна, числа Вольфа)

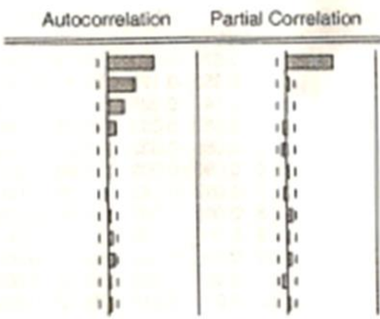


Lag	AC	PAC	Q-Stat	Prob
1	0.948	0.948	132.14	0.000
2	0.876	-0.229	245.65	0.000
3	0.807	0.038	342.67	0.000
4	0.753	0.094	427.74	0.000
5	0.714	0.074	504.80	0.000
6	0.682	0.008	575.60	0.000
7	0.663	0.126	643.04	0.000
8	0.656	0.090	709.48	0.000
9	0.671	0.232	779.59	0.000
10	0.703	0.166	857.07	0.000
11	0.743	0.171	944.39	0.000
12	0.760	-0.135	1036.5	0.000
13	0.713	-0.540	1118.0	0.000
14	0.646	-0.027	1185.6	0.000
15	0.586	0.091	1241.5	0.000
16	0.538	0.025	1289.0	0.000
17	0.500	0.033	1330.4	0.000
18	0.469	0.073	1367.0	0.000
19	0.450	0.048	1401.1	0.000
20	0.442	-0.046	1434.1	0.000
21	0.457	0.046	1469.9	0.000
22	0.482	-0.100	1510.0	0.000
23	0.517	0.052	1556.5	0.000
24	0.532	0.048	1606.1	0.000
25	0.494	-0.163	1649.2	0.000
26	0.438	-0.036	1683.3	0.000

Рис. 12.8. Тренд и сезонность (объем авиаперевозок)

ACF and PACF for standard stationary time series

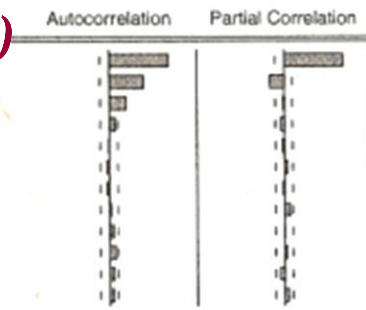
AR(1)



	AC	PAC	Q-Stat	Prob
1	0.539	0.539	116.40	0.000
2	0.319	0.041	157.37	0.000
3	0.190	0.004	171.91	0.000
4	0.092	-0.029	175.35	0.000
5	0.014	-0.044	175.43	0.000
6	0.012	0.033	175.50	0.000
7	-0.013	-0.026	175.56	0.000
8	0.025	0.059	175.81	0.000
9	0.042	0.018	176.52	0.000
10	0.069	0.042	178.47	0.000
11	0.027	-0.051	178.78	0.000
12	0.036	0.028	179.32	0.000

Рис. 12.9. AR(1). $Y_t = 0.5Y_{t-1} + \varepsilon_t$. Корень $\mu = 2$

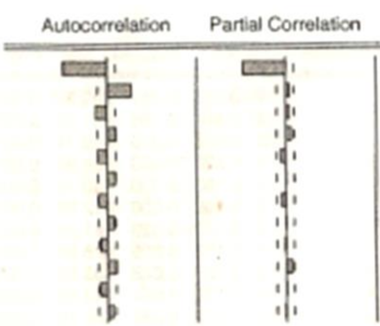
AR(2)



	AC	PAC	Q-Stat	Prob
1	0.700	0.700	196.54	0.000
2	0.403	-0.171	261.80	0.000
3	0.203	-0.016	278.34	0.000
4	0.072	-0.037	280.46	0.000
5	-0.006	-0.023	280.47	0.000
6	-0.021	0.035	280.64	0.000
7	-0.022	-0.016	280.84	0.000
8	0.017	0.071	280.95	0.000
9	0.049	0.008	281.93	0.000
10	0.071	0.025	283.99	0.000
11	0.051	-0.043	285.05	0.000
12	0.048	0.045	286.00	0.000

Рис. 12.11. AR(2). $Y_t = 0.8Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$.
Корни $\mu_1 = 2 + i, \mu_2 = 2 - i$

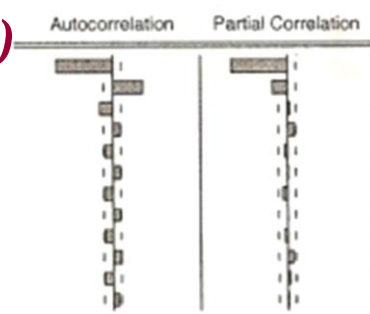
AR(1)



	AC	PAC	Q-Stat	Prob
1	-0.500	-0.500	100.19	0.000
2	0.281	0.041	131.88	0.000
3	-0.125	0.041	138.15	0.000
4	0.104	0.063	142.49	0.000
5	-0.106	-0.049	147.01	0.000
6	0.090	0.009	150.33	0.000
7	-0.096	-0.043	154.11	0.000
8	0.080	0.011	156.70	0.000
9	-0.068	-0.010	158.57	0.000
10	0.103	0.074	162.91	0.000
11	-0.081	0.009	165.60	0.000
12	0.063	-0.002	167.23	0.000

Рис. 12.10. AR(1). $Y_t = -0.5Y_{t-1} + \varepsilon_t$. Корень $\mu = -2$

AR(2)



	AC	PAC	Q-Stat	Prob
1	-0.670	-0.670	179.75	0.000
2	0.353	-0.173	229.82	0.000
3	-0.147	0.028	238.48	0.000
4	0.087	0.083	241.55	0.000
5	-0.088	-0.032	244.67	0.000
6	0.090	0.009	247.99	0.000
7	-0.097	-0.042	251.78	0.000
8	0.088	0.007	254.96	0.000
9	-0.086	-0.030	257.98	0.000
10	0.106	0.062	262.57	0.000
11	-0.092	0.029	266.04	0.000
12	0.071	0.010	268.12	0.000

Рис. 12.12. AR(2). $Y_t = -0.8Y_{t-1} - 0.2Y_{t-2} + \varepsilon_t$.
Корни $\mu_1 = -2 + i, \mu_2 = -2 - i$

ACF and PACF for standard stationary time series

MA(2)

	AC	PAC	Q-Stat	Prob
1	-0.593	-0.593	140.88	0.000
2	0.124	-0.351	147.01	0.000
3	0.004	-0.185	147.02	0.000
4	0.026	-0.034	147.29	0.000
5	-0.069	-0.068	149.21	0.000
6	0.076	0.003	151.55	0.000
7	-0.074	-0.050	153.79	0.000
8	0.056	-0.014	155.06	0.000
9	-0.055	-0.058	156.32	0.000
10	0.088	0.050	159.47	0.000
11	-0.077	0.024	161.89	0.000
12	0.035	0.010	162.40	0.000

Рис. 12.16. MA(2). $Y_t = \varepsilon_t - 0.9\varepsilon_{t-1} + 0.2\varepsilon_{t-2}$.
Корни $\mu_1 = 2.5$, $\mu_2 = 2$

ARMA(1,1)

	AC	PAC	Q-Stat	Prob
1	0.687	0.687	189.53	0.000
2	0.336	-0.259	235.01	0.000
3	0.177	0.129	247.59	0.000
4	0.071	-0.107	249.63	0.000
5	0.003	0.018	249.63	0.000
6	-0.013	0.004	249.70	0.000
7	-0.015	-0.009	249.79	0.000
8	0.016	0.067	249.90	0.000
9	0.051	0.009	250.97	0.000
10	0.066	0.023	252.76	0.000
11	0.042	-0.044	253.50	0.000
12	0.039	0.057	254.13	0.000

Рис. 12.20. ARMA(1,1). $Y_t = 0.4Y_{t-1} + \varepsilon_t + 0.5\varepsilon_{t-1}$.
Корни $\mu_{AR} = 2, \mu_{MA} = -2$

MA(2)

	AC	PAC	Q-Stat	Prob
1	-0.074	-0.074	2.1884	0.139
2	-0.151	-0.158	11.407	0.003
3	0.048	0.024	12.338	0.006
4	0.008	-0.010	12.365	0.015
5	-0.052	-0.042	13.451	0.020
6	0.016	0.008	13.559	0.035
7	-0.043	-0.057	14.316	0.046
8	0.009	0.008	14.352	0.073
9	0.015	0.000	14.438	0.108
10	0.067	0.075	16.307	0.091
11	-0.040	-0.027	16.974	0.109
12	-0.002	0.009	16.975	0.151

Рис. 12.17. MA(2). $Y_t = \varepsilon_t - 0.1\varepsilon_{t-1} - 0.2\varepsilon_{t-2}$.
Корни $\mu_1 = -2.5, \mu_2 = 2$

ARMA(1,1)

	AC	PAC	Q-Stat	Prob
1	-0.662	-0.662	175.75	0.000
2	0.281	-0.279	207.58	0.000
3	-0.114	-0.117	212.83	0.000
4	0.082	0.021	215.55	0.000
5	-0.095	-0.047	219.20	0.000
6	0.095	0.008	222.85	0.000
7	-0.095	-0.047	226.53	0.000
8	0.082	-0.006	229.26	0.000
9	-0.079	-0.046	231.83	0.000
10	0.102	0.054	236.10	0.000
11	-0.091	0.029	239.50	0.000
12	0.063	0.014	241.12	0.000

Рис. 12.21. ARMA(1,1). $Y_t = -0.4Y_{t-1} + \varepsilon_t - 0.5\varepsilon_{t-1}$.
Корни $\mu_{AR} = -2, \mu_{MA} = 2$

Model of autoregression and integrated moving average *ARIMA(p,d,q)*

$y_1 \; y_2 \; y_3 \; . \; . \; . \; y_t$

$\Delta y_t = y_2 - y_1 \quad y_3 - y_2 \quad y_4 - y_3 \; . \; . \; .$

$\Delta^2 y_t$

Let y_t be an integrated **non-stationary** time series.

Applying the first difference operator finite times, we obtain a stationary time series $\Delta^{d} y_t$

Box and Jenkins methodology ***(Box, Jenkins, 1976)***

1. Model specification

1.1. A time series is checked for stationarity by means of :

- graphical analysis *(trend, seasonality, "spread" over time, and others);*
- applying specified tests:
  - Advanced Dickey-Fuller test *(if the ADF-statistic is in the critical region, it might be a case of non-stationarity).*
  - Ljung-Box test *(based on statistic* Q_{LB} *).*

If necessary, the first differences should be taken finite times. For stationary series *ACF(k)* decreases rapidly with increasing *k*. This means that the necessary degree is achieved.
As a result we need to get a stationary time series.

1.2. For the considered data set *ACF* and *PACF* are constructed and hypotheses on the orders of autoregression (*p*) and moving average (*q*) are formalized. Analysis of *ACF* and *PACF* is needed to be done. You should consider $\underbrace{p+q}\le 4.$!

You should select several appropriate models.

Model of autoregression and integrated moving average *ARIMA(p,d,q)*

2. Model identification and verification its adequacy

2.1. For each model the parameters are estimated and the residuals (errors) are checked for being “white noise”.

2.2. Each model is tested on the adequacy:

- the model coefficients are checked for statistical significance using *t*-test (Student’s criteria).

The model should not contain redundant parameters.

2.3. The «best» model is chosen on the basis of two requirements:

- the minimal number of model parameters;
- sufficiently good quality of a model fitting.

Among "good" models the simplest one usually should be chosen (with the least number of parameters). But it depends on…

3. Forecasting

At this step using the «best» model you can make forecasts for several steps forward.

The adequacy of ARIMA models

- *Check statistical significance* of the coefficients estimates for the selected level of significance.

- Analysis of residuals: ε_t - *white noise.*

  If corresponding tests show the presence of autocorrelation in the residuals, ***the model badly fit the data set: it should be modified.***

- *Choose the "best" model* based on the Akaike information criterion and the Schwarz-Bayes information criterion (AIC и SBC).

MAPE

$$AIC = 2\frac{p+q}{n} + \ln\left(\frac{\sum_{t=1}^{n}\varepsilon_t^2}{n}\right) \qquad SBC = \frac{(p+q)\ln n}{n} + \ln\left(\frac{\sum_{t=1}^{n}\varepsilon_t^2}{n}\right)$$

AIC and SBC in its ideology are close to the adjusted coefficient of determination: They both consider two requirements:

1) The minimal number of model parameters;
2) Sufficiently good quality of a model fitting.

SARIMA

Seasonal model of autoregression and integrated moving average *ARIMA(p,d,q)(P,D,Q)s*

is $s=12$

ACF
PACF

$y_t - y_{t-1}$

$y_{13} - y_1$

Consider seasonal model *ARIMA(p, d, q)(P, D, Q)s*

Here *P* is a seasonal autoregressive parameter,

D is an order of seasonal differences,

Q is a seasonal moving average parameter.

s is a seasonal lag.

1. Continuation of seasonal model specification

1.1. In a case of a pronounced seasonal component, it is advisable to include seasonal differentiation into the model. For example, $\Delta^D(\Delta^d y_t)$.

However, when solving practical economic problems, it is not recommended to use seasonal differentiation of more than the first order for each seasonal lag.

1.2. *ACF* and *PACF* are considered and seasonal parameters of autoregression *SAR(P)*

and moving average *SMA*(*Q*) are defined. It is necessary to take into account this condition for parameters: $n - (p + P + D + D \cdot s) > 1$.

Several appropriate models can be considered. But the decision of which model to choose is made by a researcher .