

11.3 Autoregresný proces – AR(p)

Proces AR(p) má tvar

$$y_t = \varepsilon_t + \varphi_1 y_{t-1} + \dots + \varphi_p y_{t-p}.$$

Pomocou operátora spätného posunutia ho môžeme zapísať v tvare

$$\varphi(B)y_t = \varepsilon_t,$$

kde

$$\varphi(B) = 1 - \sum_{j=1}^p \varphi_j B^j$$

je tzv. **autoregresný operátor**.

Proces AR(p) je stacionárny, ak všetky korene rovnice $\varphi(B) = 0$ ležia mimo jednotkového kruhu.

Pre výpočet parametrov $\varphi_1, \varphi_2, \dots, \varphi_p$ procesu AR(p) sa používa aj tzv. Yuleova – Walkerova sústava rovníc

$$\begin{aligned}\rho_1 &= \varphi_1 + \varphi_2 \rho_1 + \dots + \varphi_p \rho_{p-1}, \\ \rho_2 &= \varphi_1 \rho_1 + \varphi_2 + \dots + \varphi_p \rho_{p-2}, \\ &\dots\dots\dots, \\ \rho_p &= \varphi_1 \rho_{p-1} + \varphi_2 \rho_{p-2} + \dots + \varphi_p.\end{aligned}$$

Pre proces AR(1) dostávame

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

s podmienkou stacionarity $|\varphi_1| < 1$. Autokorelačná funkcia procesu AR(1) má tvar ρ_1^k , $k \geq 0$.

Teda $\varphi_1 = \rho_1$.

Proces AR(2) má tvar

$$y_t = \varepsilon_t + \varphi_1 y_{t-1} + \varphi_2 y_{t-2}.$$

Podmienky stacionarity procesu AR(2) sú

$$\varphi_2 + \varphi_1 < 1, \quad \varphi_2 - \varphi_1 < 1, \quad -1 < \varphi_2 < 1.$$

Odpovedajúca sústava rovníc pre φ_1, φ_2 , má tvar

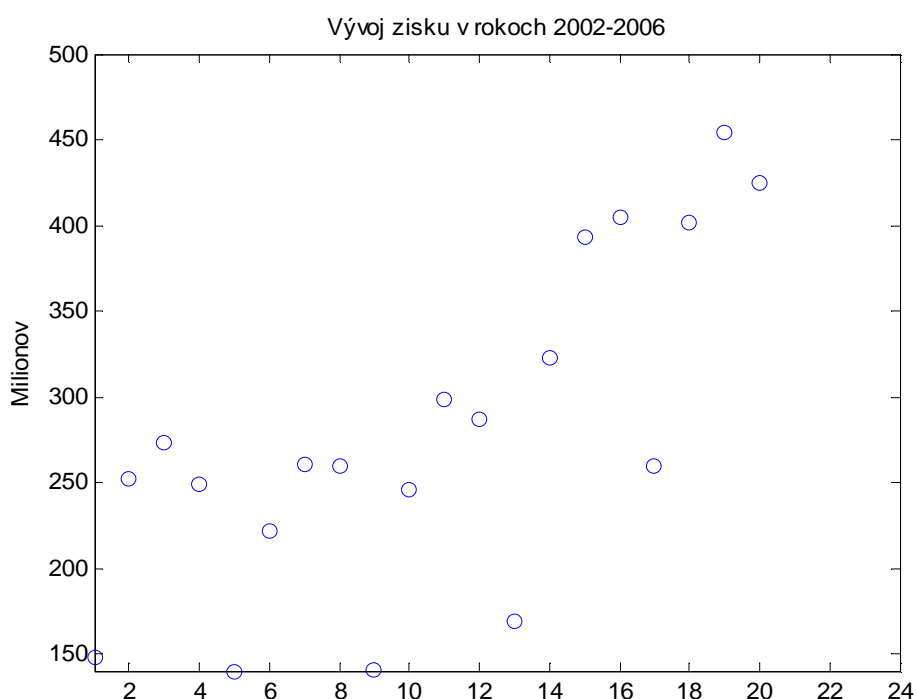
$$\rho_1 = \varphi_1 + \varphi_2 \rho_1, \quad \rho_2 = \varphi_1 \rho_1 + \varphi_2.$$

Príklad 11.1 Pomocou modelu AR(1) urobme predpovede vývoja zisku vo firme XY v jednotlivých štvrtrokoch roku 2007, ak vývoj zisku za minulé obdobie je daný tabuľkou

	1. štvrtrok	2. štvrtrok	3. štvrtrok	4. štvrtrok
2002	147.6	251.8	273.1	249.1
2003	139.3	221.2	260.2	259.5
2004	140.5	245.5	298.8	287.0
2005	168.8	322.6	393.5	404.3
2006	259.7	401.1	452.2	425.0

Dáta boli prevzaté z práce Hanke, J. E. – Reitsch, A.G.: Understanding Business Statistics.

Riešenie:



Vypočítame 95% interval spoľahlivosti pre autokorelačný koeficient

$$0 \pm y \left(\frac{1}{\sqrt{n}} \right) = 0 \pm 1.96 \left(\frac{1}{\sqrt{20}} \right) = 0 \pm 0.44 .$$

Výpočtom autokorelačných koeficientov dostávame

k	1	2	3	4	5	6	7	8	9	10
r_k	0.46	0.15	0.25	0.60	0.12	-0.17	-0.12	0.18	-0.16	-0.33

Autokorelačný koeficient pre dĺžku 4 je rôzny od nuly, pretože $0.60 > 0.44$. Pre predpoveď teda bude vhodný model AR(1) tvaru

$$y_t = a_0 + a_1 t + a_2 y_{t-4},$$

kde je zahrnutá aj trendová zložka $a_0 + a_1 t$ (modelu AR(1) odpovedá časť $a_2 y_{t-4}$).

Výpočet realizujeme pomocou MAPLu, pričom $yt = y_t, yt4 = y_{t-4}$. Použitím MNŠ dostávame:

```
> with(stats):
fit[leastsquare[[t,yt4,yt]]](
[[5,6,7,8,9,10,11,12,13,14,15,16,17,18,19,20],
[147.6,251.8,273.1,249.1,139.3,221.2,260.2,259.5,140.5,245.5,298.8,287.0,168.8,322.6,393.5,404.3],
[139.3,221.2,260.2,259.5,140.5,245.5,298.8,287.0,168.8,322.6,393.5,404.3,259.7,401.1,454.2,425.0]]);
```

$$yt = -26.89210400 + 8.322962152t + 0.8484004218yt4$$

Jednotlivé štvrťroky roku 2007 odpovedajú hodnotám $t = 21, 22, 23, 24$.

```
> y21=-26.89210400+8.322962152*21+.8484004218*259.7;
y21 = 368.2196907
```

```
> y22=-26.89210400+8.322962152*22+.8484004218*401.1;
y22 = 496.5064725
```

```
> y23=-26.89210400+8.322962152*23+.8484004218*454.2;
y23 = 549.8794971
```

```
> y24=-26.89210400+8.322962152*24+.8484004218*425.0;
y24 = 533.4291669
```

Predpovede na rok 2007 teda sú

	1.štvrťrok	2.štvrťrok	3.štvrťrok	4.štvrťrok
pred. zisk	368.2196907	496.5064725	549.8794971	533.4291669

Správnosť daného modelu sa pokúsime overiť na hodnotách z roku 2006 (teda urobíme pomocou získaného modelu predikciu na rok 2006). Dostávame

	1.štvrťrok	2.štvrťrok	3.štvrťrok	4.štvrťrok	Suma
skutočnosť y_t	259.7	401.1	452.2	425.0	
pred. zisk \hat{P}_t	257.8	396.6	465.1	482.58	
$(y_t - \hat{P}_t)^2$	3.61	20.25	166.41	3315.46	3505.73
y_t^2	67444.09	160881.21	204484.84	180625.00	613435.14

t	y_t	Predpoveď \hat{P}_t	$y_t - \hat{P}_t$	$(y_t - \hat{P}_t)^2$	$\frac{ y_t - \hat{P}_t }{y_t}$	$e_t - e_{t-1}$
1,00	147,60					
2,00	251,80					
3,00	273,10					
4,00	249,10					
5,00	139,30	139,946609	-0,65	0,42	0,004642	
6,00	221,20	236,6728951	-15,47	239,41	0,06995	-14,83
7,00	260,20	263,0667863	-2,87	8,22	0,011018	12,61
8,00	259,50	251,0281383	8,47	71,77	0,032647	11,34
9,00	140,50	166,1967341	-25,70	660,32	0,182895	-34,17
10,00	245,50	244,0036908	1,50	2,24	0,006095	27,19
11,00	298,80	285,4142694	13,39	179,18	0,044798	11,89
12,00	287,00	293,1433513	-6,14	37,74	0,021405	-19,53
13,00	168,80	200,5066632	-31,71	1005,31	0,187836	-25,56
14,00	322,60	297,9116697	24,69	609,51	0,076529	56,39
15,00	393,50	351,4543743	42,05	1767,83	0,10685	17,36
16,00	404,30	349,7662115	54,53	2973,93	0,134884	12,49
17,00	259,70	257,8082438	1,89	3,58	0,007284	-52,64
18,00	401,10	396,6151908	4,48	20,11	0,011181	2,59
19,00	452,20	465,0897429	-12,89	166,15	0,028505	-17,37
20,00	425,00	482,5754296	-57,58	3314,93	0,135472	-44,69
Suma		4681,2		11060,66	1,061991	-56,93
21,00		368,2196907				
22,00		496,5064725				
23,00		548,1826962				
24,00		533,4291669				
Výpočet koeficientov testovania kvality modelu pre t=5,...,20						
M.A.P.E.	v percent.				4,56 %	12304,70
M.S.E.				876,4325		
RootM.S.E.				29,6046		
D-W						1,112474
Theil					7,56 %	

Ako mieru prognostickej kvality získaného modelu používame **koeficient nesúladu** (nesúlad medzi simulovanou predpoveďou a v tom čase už známou skutočnosťou) pomocou Theilovho koeficienta

$$T_H^2 = \frac{\sum_{j=1}^D (y_{N+j} - \hat{P}_j)^2}{\sum_{j=1}^D (y_{N+j})^2},$$

kde n je dĺžka časového radu pre odhad modelu,

$D = n - N$, je skrátenie časového radu,

\hat{P}_j je extrapolácia (predpoveď) na j období dopredu modelom odhadnutým na základe prvých N pozorovaní.

Napríklad pre rok 2006 je

$$T_H^2 = \frac{3505,73}{613435,14} = 0,0057149, T_H \approx 7,56\% .$$

V percentuálnom vyjadrení ($T_H \times 100\%$) môžeme túto veličinu interpretovať ako chybu predpovede vyjadrenú v percentách. Ak sa pohybuje v hraniciach 3-5%, potom je chyba predpovede považovaná za „malú“ a uvažovaný model môže byť vhodným nástrojom pre prognózovanie. Ak je väčšia ako 5%, ale menšia ako 8%, nie je ďalšie použitie daného modelu vylúčené. Je zrejmé, že pre vytvorenie lepšieho modelu je potrebná väčšia vzorka.

Absolútne miery presnosti iba pre rok 2006:

$$M.S.E. = \frac{1}{N} \sum_{t=1}^N (y_t - \hat{P}_t)^2 = \frac{3505,73}{4} = 876,4325 - \text{priemerná kvadratická chyba rozptylu rezíduí},$$

$$Root\ M.S.E. = \sqrt{M.S.E} = 29,6046 - \text{štandardná odchýlka rezíduí}.$$

Relatívna miera presnosti iba pre rok 2006:

$$M.A.P.E. = \frac{1}{N} \sum_{t=1}^N \left| \frac{y_t - \hat{P}_t}{y_t} \right| \times 100\% = \frac{1}{4} 0,1824 \times 100\% = 0,0456 \times 100\% = 4,56\% - \text{relatívna}$$

absolútna chyba rezíduí.

Pomocou Durbin – Watsonovej charakteristiky pre roky 2003 - 2006 dostávame

$$D - W = \frac{\sum_{t=6}^{20} (e_t - e_{t-1})^2}{\sum_{t=5}^{20} e_t^2} = \frac{12304,7}{11060,66} = 1,112474$$

kde $e_t = y_t - \hat{y}_t$, pre $t = 1, 2, \dots, N$.

Je zrejmé, že je potrebné zvážiť použiteľnosť daného modelu pre predikcie budúcich hodnôt. Durbin – Watsonova charakteristika pre obe hladiny významnosti (0,01 a 0,05) hovorí, že výsledok testu je nepresvedčivý (nejasný). Pridaním premenných v modeli môžeme vylepšiť predikciu. Potrebovali by sme tiež asi viac hodnôt z minulosti.