ODHAD ŘÁDU AUTOREGRESNÍHO MODELU

Ladislav Tomášek, IHE, Praha

1. Úvod

Při studiu autoregresních modelů hraje důležitou roli stanovení řádu modelu. Z literatury jsou známy některé metody, které určují řád modelu nepřímo - na základě testů hypotéz o nulových autoregresních koeficientech. Byly však navrženy i přímé metody, které využívají skutečnosti, že odhad rozptylu v modelech s podceněným řádem vede obvykle k hodnotám vyšším, než je skutečná hodnota rozptylu, zatímco v modelech s přeceněným řádem jsou hodnoty obou rozptylů srovnatelné.

Autoregresní posloupností řádu p budeme rozumět stacionární posloupnost $\{ x_t \}$ náhodných veličin, které splňují vztah

$$x_t + a_1 x_{t-1} + \ldots + a_p x_{t-p} = e_t \,, \qquad (1.1)$$

kde $a_p \neq 0$ a kde $\{ e_t \}$ je posloupnost nekorelovaných náhodných veličin s nulovými středními hodnotami a stejnými rozptyly σ^2. Parametry $a_1, \ldots, a_p$ jsou vázány podmínkou, že všechny kořeny polynomu

$$z^p + a_1 z^{p-1} + \ldots + a_p = 0$$

leží uvnitř jednotkoveho kruhu.

Při odhadování autoregresních koeficientů vycházíme z posloupnosti délky N . Neznámé parametry $a' = (a_1, \ldots, a_p)$ seodhadují obvykle metodou nejmenších čtverců. Uvažuje se funkce kvadratických odchylek

$$S(a) = \sum_{t=p+1}^{N} (x_t + a_1 x_{t-1} + \ldots + a_p x_{t-p})^2 \,. \qquad (1.2)$$

Hodnoty vektorového parametru a , které minimalizují součet S(a) se považují za odhad metodou nejmenších čtverců. Řešení â dostaneme ze soustavy normálních rovnic

$$\sum_{t=p+1}^{N} x_{t-j} (x_t + a_1 x_{t-1} + \ldots + a_p x_{t-p}) = 0 \,, \quad j = 1, \ldots, p \,.$$

Označíme-li

$$c_{ij} = \sum_{t=+1}^{N} x_{t-i}x_{t-j} \ , i,j=0,\ldots,p,$$

$$C_0 = (c_{01} \ \ldots \ c_{0p})' \ ,$$

$$C = \begin{pmatrix} c_{11} \ldots c_{1p} \\ \vdots \quad \vdots \\ c_{p1} \ldots c_{pp} \end{pmatrix} ,$$

můžeme za předpokladu regularity matice C vyjádřit řešení ve tvaru $\hat{a} = - C^{-1}C_0$.

Odhad parametru σ^2 se získá na základě veličiny

$$s^2 = S(\hat{a})/(N-p) = (c_{00} - C_0'C^{-1}C_0)/(N-p) \ . (1.3)$$

2. Některé metody odhadu řádu autoregresního modelu

Označme s_p^2 odhad parametru σ^2 za předpokladu, že řád modelu je p. Podstatou tzv. přímých metod odhadu řádu modelu je nalezení minima v konečné posloupnosti vhodně transformovaných hodnot s_p^2. Je zřejmé, že tato transformace musí znevýhodňovat modely s příliš vysokými hodnotami řádu p a dále musí být do transformace zahrnut i parametr N.

Jedním z prvních, kdo se zabývali otázkou odhadu řádu autoregresního modelu, byl H.Akaike. Při konstrukci nejlepší lineární predikce prvku x_{N+1} dospěl k výrazu pro odhad reziduálního rozptylu

$$E\ (\hat{x}_{N+1} - x_{N+1})^2 = \sigma^2(1+ p/N) \ , \qquad (2.1)$$

kde $\hat{x}_{N+1}$ se získá na základě p předcházejících hodnot a odhadů autoregresních koeficientů metodou nejmenších čtverců. Neznámý parametr σ^2 ve výrazu (2.1) H.Akaike nahradil veličinou $S(\hat{a})/(N-2p)$ a tento odhad rozptylu označil jako FPE_p (final prediction error).

Vzhledem k tomu, že $p \ll N$, můžeme tuto veličinu převést na tvar

$$FPE_p = (N-p)\ s_p^2(1+p/N)/(N-2p) = s_p^2(1+2p/N) \ .$$

V logaritmovaném tvaru bývá označována jako

$$AIC_p = \ln s_p^2 + 2p/N \ . \qquad (2.2)$$

Později zkoumal otázku odhadu řádu modelu G.Schwarz. Na základě bayesovských úvah o řádu regresního modelu jako o veličině s apriorním rozdělením dospěl ke kriteriu pro odhad řádu modelu založeném na veličinách

$$\ln L_p(x_1,\ldots,x_N) - \tfrac{1}{2} p \ln N ,$$

kde L_p je věrohodnostní funkce v modelu řádu p. Aplikací na autoregresní model dostaneme transformační funkci ve tvaru

$$\ln s_p^2 + p(\ln N)/N . \qquad (2.3)$$

Ke stejnému výsledku došel i J.Rissanen na základě úvah založených na co nejúspornějším záznamu posloupnosti $x_1,\ldots,x_N$.

Přibližně ve stejné době zkoumali tento problém E.J.Hannan a B.G.Quinn. Jako funkci penalizující počet parametrů navrhli

$$h(N) = 2c(\ln(\ln N))/N , \quad c > 1 . \qquad (2.4)$$

Odhad řádu modelu založený na takové funkci je konzistentní. Důkaz je založen na odhadu rozptylu pomocí parciálních autokorelací a na zákonu o iterovaném logaritmu.

Později navrhli J.Anděl, M.G.Perez a A.I.Negrao kriterium ve tvaru

$$A_p = s_p^2 (1 + p\, w_N) , \qquad (2.5)$$

kde $w_N \to 0$ a $N w_N \to \infty$ při $N \to \infty$. Za těchto předpokladů je asymptoticky

$$P(A_p < A_k, k \neq p) = 1 .$$

Konkrétně byla navržena penalizační funkce w_N ve tvaru

$$w_N = c\, N^{-\alpha} , \quad c > 0 , \quad 0 < \alpha < \tfrac{1}{2} .$$

Z předcházejících poznámek vyplývá, že kritérium pro odhad řádu autoregresního modelu můžeme obecně založit na veličinách

$$Q(p,N) = (\ln s_p^2) + p\, h_N/N , \qquad (2.6)$$

kde h_N je neklesající funkce N taková, že $h_N/N \to 0$ při $N \to \infty$. Veličinu h_N budeme dále nazývat penalizační faktor.

3. Simulace

K posouzení kvality různých metod odhadu řádu p byly provedny simulace autoregresních posloupností do třetího řádu. Simulované posloupnosti byly získány na základě vztahu (1.1), přičemž pro $t \leq p$ bylo $x_t = 0$ a pro zlepšení stacionarity posloupností bylo prvních 60 členů posloupností vynecháno.

Generující autoregresní koeficienty byly zvoleny následovně:

p	a_1	a_2	a_3
1	0.36		
2	0.24	0.36	
3	0.30	0.36	0.36

Rozptyl σ^2 veličin e_t byl jednotkový. Autoregresní posloupnosti byly simulovány pro hodnoty N = 50, 75, 100, ..., 250 . U každé posloupnosti byly pak vypočteny odhady s_p^2 pro p = 0, 1,..., 7 . Pro každý řád a každou délku posloupnosti bylo uskutečněno 1000 realizací. Kvalitu různých metod ilustrují následující tabulky, které obsahují četnosti modelů pro různé hodnoty p .

N	p = 0	1	2	3	4	5	6	7
50	36	20	89	419	149	96	92	99
75	1	7	53	530	152	98	81	78
100	1	1	8	573	154	101	89	73
125	0	0	4	609	143	91	82	71
150	0	0	2	606	149	96	74	73
175	0	0	0	614	137	105	70	74
200	0	0	0	651	120	95	71	63
225	0	0	0	603	165	93	58	81
250	0	0	0	635	147	72	73	73

Tab. 3.1 Akaikeova metoda pro AR(3)

N	p = 0	1	2	3	4	5	6	7
50	204	42	168	459	78	26	9	14
75	53	39	127	697	53	18	9	4
100	21	9	60	819	71	15	3	2
125	10	4	27	887	48	20	3	1
150	1	0	11	931	42	12	1	2
175	1	1	2	944	39	11	1	1
200	0	0	4	941	41	12	2	0
225	0	0	0	955	36	7	1	1
250	0	0	0	971	27	1	1	0

Tab. 3.2 Schwarzova metoda pro AR(3)

N	p = 0	1	2	3	4	5	6	7
50	226	43	171	447	73	22	7	11
75	55	40	133	691	52	17	9	3
100	21	9	60	817	72	16	3	2
125	10	4	27	885	49	21	3	1
150	1	0	8	926	49	13	1	2
175	0	1	1	936	45	15	1	1
200	0	0	1	937	46	14	2	0
225	0	0	0	942	44	11	2	1
250	0	0	0	962	35	1	2	0

Tab. 3.3 Hannan-Quinnova metoda pro AR(3) , c=1.5

N	p = 0	1	2	3	4	5	6	7
50	112	34	139	497	101	55	33	29
75	42	37	123	691	61	24	14	8
100	27	11	63	816	66	13	3	1
125	14	6	37	892	36	14	0	1
150	4	4	18	938	32	4	0	0
175	4	0	10	965	16	5	0	0
200	2	0	6	972	17	3	0	0
225	0	1	4	987	8	0	0	0
250	0	0	1	994	4	1	0	0

Tab. 3.4 Andělova metoda pro AR(3) , c=0.3, α=0.4

Z četností uvedených tabulek je patrné, že Akaikeova metoda obecně přeceňuje řád modelu, což je výrazné zejména u delších posloupností. Tato skutečnost je ostatněznáma z literatury.O něco lépe fungují kritéria Schwarze a Hannana, kde podíl přeceněných modelů je u delších posloupností relativně nižší, u krátkých posloupností je naopak ve srovnání s Akaikeovým kritériem vyšší podíl nedoceněných modelů. Relativně nejlepší výsledky dává metoda Andělova.

4. Empirické odhady penalizačního faktoru

Na základě výsledků simulací byla dále studována otázka, jaké mezní hodnoty penalizačního faktoru lze připustit při předepsanéchybě přecenění, resp. podcenění řádu modelu.

Uvažujme autoregresní posloupnost řádu p . Jestliže má kritérium pro volbu řádu tvar

$$Q(k,N) = \ln s_k^2 + k\, h_N/N , \quad k=0,\ldots,K , \qquad (4.1)$$

je minimální hranice penalizačního faktoru pro odhad, který nepřecení řád p , dána výrazem

$$h_N^- = -N \min_{k>p} (\ln s_k^2 - \ln s_p^2)/(k-p) . \qquad (4.2)$$

Analogicky maximální hranici penalizačního faktoru proti podceněnířádu p lze vyjádřit ve tvaru

$$h_N^+ = -N \max_{k<p} (\ln s_k^2 - \ln s_p^2)/(k-p) . \qquad (4.3)$$

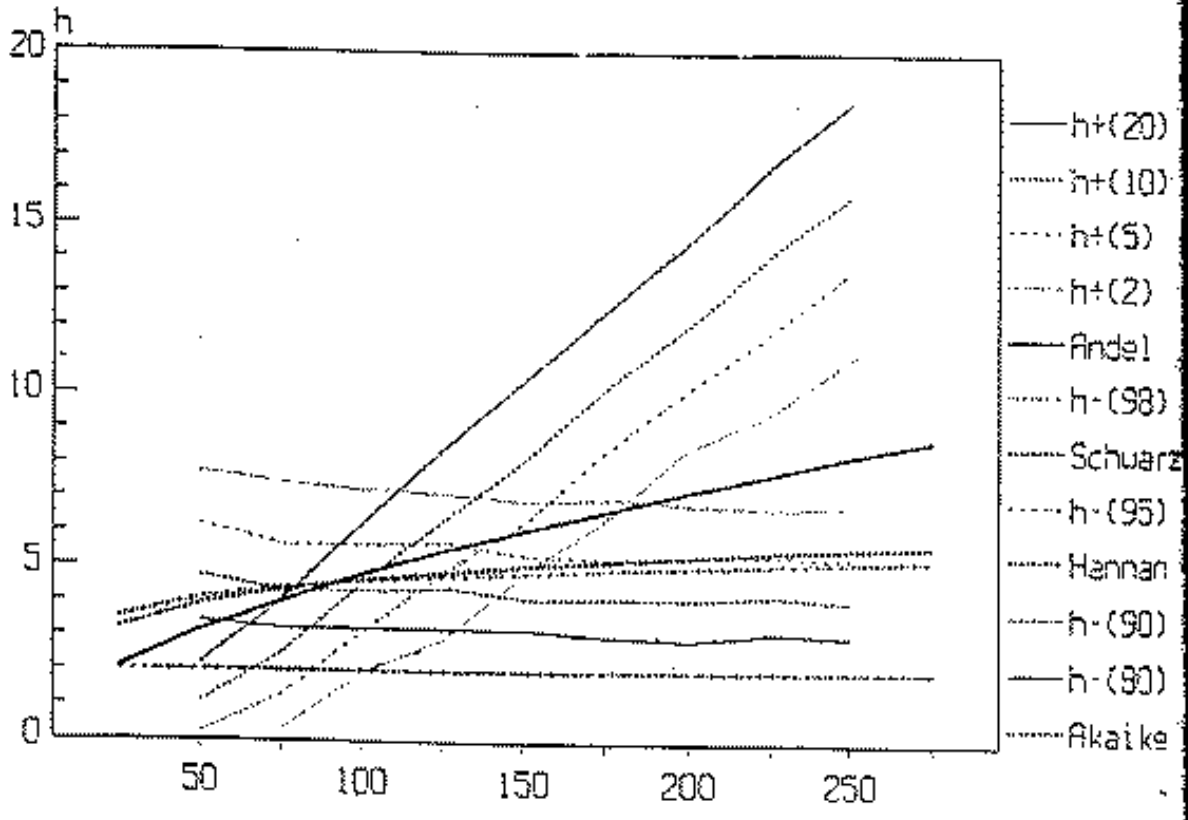
Na základě hodnot h^-a h^+vypočtených v jednotlivych simulovaných posloupnostech lze si utvořit představu o hodnotách penalizačního faktoru pro různé hodnoty N . Nejvhodnější charakteristiky jsou empirické kvantily, které jsou uvedeny v nasledující tabulce. Čísla v závorkách označují procento daného kvantilu.

p	N	$h^-(98)$	$h^-(95)$	$h^-(90)$	$h^-(80)$	$h^+(20)$	$h^+(10)$	$h^+(5)$	$h^+(2)$
	50	7.6	5.8	4.5	3.4	2.5	1.1	0.1	-0.7
	75	7.2	5.5	4.1	3.0	4.9	2.9	1.3	0.2
	100	7.1	5.1	4.2	3.1	7.6	5.5	3.5	2.2
	125	7.3	5.6	4.1	3.0	10.7	7.1	4.5	2.6
1	150	7.4	5.8	4.4	3.3	13.2	10.3	7.7	5.1
	175	7.0	5.1	4.0	2.9	15.8	12.6	9.5	6.6
	200	6.3	4.9	3.8	2.8	18.4	14.5	11.8	8.8
	225	6.5	5.2	4.2	3.1	22.1	18.4	15.6	11.4
	250	7.2	5.4	4.1	3.0	23.6	19.7	16.3	12.9
	50	7.8	6.1	4.6	3.3	2.3	1.3	0.6	-0.1
	75	7.8	5.6	4.4	3.2	3.7	2.5	1.7	1.0
	100	7.1	5.4	4.3	3.1	5.5	3.8	2.8	1.6
	125	7.4	5.8	4.6	3.3	7.6	5.7	4.8	2.9
2	150	6.6	4.6	3.8	2.9	8.8	7.0	5.5	4.2
	175	7.4	5.4	4.1	3.0	10.9	9.0	7.2	5.8
	200	6.7	5.4	4.1	2.9	12.9	10.9	9.4	7.4
	225	6.6	5.2	4.2	2.9	14.8	12.2	10.7	8.6
	250	7.0	5.4	4.1	3.0	16.3	13.8	11.4	10.0
	50	8.0	6.6	5.0	3.6	1.9	0.8	0.1	-0.4
	75	7.4	5.6	4.4	3.2	3.8	2.4	1.1	0.1
	100	7.5	6.2	4.6	3.3	6.1	4.6	3.1	1.9
	125	6.6	5.6	4.3	3.2	8.2	6.6	5.2	3.1
3	150	7.2	5.3	4.1	3.0	10.4	8.4	7.3	5.7
	175	6.9	5.2	4.1	3.1	12.4	10.4	8.9	7.0
	200	7.1	5.5	4.2	3.1	14.3	12.5	10.5	8.9
	225	7.0	5.2	4.4	3.1	16.9	14.4	12.6	10.4
	250	6.1	4.7	3.9	2.9	18.9	16.5	14.8	12.4

Tab. 4.1 Empirické kvantily veličin h_N^- a h_N^+ v AR modelech

Z uvedené tabulky jsou patrné relativně malé rozdíly empirických kvantilů pro různé řády p. Má tedy smysl spojit soubory hodnot h^- a h^+pro modely o různých řádech a určit empirické kvantily.

Následující grafické vyjádření ilustruje vztah těchto empirických kvantilů a penalizačních faktorů metod z části 2.



Obr. 4.1 Souhrnné empirické kvantily veličin h^- a h^+ v závislosti na N

Literatura

1 Akaike, H.: Fitting autoregressive models for prediction. Ann.Inst.Stat.Math. 21, 1969

2 Anděl, J., Perez, M.G., Negrao, A.J.: Estimating the dimension of a linear model. Kybernetika, Vol.17, 1981

3 Hannan, E.J., Quinn, B.C.: The determination of the order of an autoregression. J.Roy.Stat.Soc., Ser.B 41, 1979

4 Schwarz, G.: Estimating the dimension of a model. Ann.Stat. 6, 1978