УДК 528.412

Г. А. ХАРКЕВИЧ

ТОЧНОСТЬ ЭКСТРАПОЛЯЦИИ ТРЕНДОВ, АППРОКСИМИРУЕМЫХ ПОЛИНОМАМИ ПЕРВЫХ ТРЕХ СТЕПЕНЕЙ

При анализе временных рядов с целью выявления тенденции развития исследуемого явления, например, осадки инженерного сооружения, используются тренды, построенные по результатам наблюдений.

Рассмотрим вопрос о точности экстраполяции таких трендов на один, два шага вперед, представленных сглаженными кривыми, аппроксимированными полиномами 1, 2, 3-й степени соответственно.

Имеем временной ряд

$$S_i = y_i^n + \varepsilon_i, \tag{1}$$

где $i=1, 2, \ldots, k$ — номер цикла наблюдений; S_i — измеренная величина; y_i^n — ее сглаженное значение (неслучайная величина, аппроксимируемая полиномиальной моделью); ε_i — случайная величина — ошибка измерений. Соответствующие дисперсии будут

$$\sigma_S^2 = \sigma_{y^n}^2 + \sigma_\varepsilon^2, \tag{2}$$

где σ_S — общая дисперсия измеренной величины S; σ_{y^n} — дисперсия тренда, представленного полиномам степени n; σ_ε — дисперсия измеренной величины относительно тренда y^n.

После назначения модели y^n и определения ее параметров дисперсии σ_ε, σ_{y^n} оцениваются по формулам:
Если начало счета времени выбрать в средине ряда, то

$$\Sigma t = \Sigma t^3 = \Sigma t^5 = 0.$$

Поэтому

$$T^T T = \begin{vmatrix} k & 0 & \Sigma t^2 & 0 & \Sigma t^4 & \cdots \\ 0 & \Sigma t^2 & 0 & \Sigma t^4 & 0 & \cdots \\ \Sigma t^2 & 0 & \Sigma t^4 & 0 & \Sigma t^6 & \cdots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \end{vmatrix}.$$

Введем обозначение

$$\lambda(k, n, l) = t_L^T (T^T T)^{-1} t_L \tag{6}$$

и найдем распределение λ в зависимости от аргументов k, n, l, где l число шагов экстраполирования после k-го цикла наблюдений.

При $n=1$, 2, 3 соответственно запишем

$$\sigma_{y^1}^2 = \sigma_\varepsilon^2 \left(\frac{1}{k} + \frac{t_l^2}{\Sigma t^2}\right),$$

$$\sigma^2_{y^3} = \sigma^2_\varepsilon \left[\frac{t_l^2}{\Sigma t^2} + \frac{\Sigma t^4 - 2(\Sigma t^2)\, t_l^2 + k t_l^4}{k \Sigma t^4 - (\Sigma t^2)^2} \right], \tag{7}$$

$$\sigma^2_{y^3} = \sigma^2_\varepsilon \left[\frac{\Sigma t^4 - 2(\Sigma t^2)\, t_l^2 + k t_l^4}{k \Sigma t^4 - (\Sigma t^2)^2} + \frac{(\Sigma t^6)\, t_l^2 - 2(\Sigma t^4)\, t_l^4 + (\Sigma t^2)\, t_l^6}{\Sigma t^2\, \Sigma t^6 - (\Sigma t^4)^2} \right],$$

где суммирование выполнено в пределах от $-(k-1)/2$ до $(k-1)/2$.

При k нечетном

$$\Sigma t^2 = (k-1)\, k\, (k+1)/12,$$

$$\Sigma t^4 = \Sigma t^2\, (3k^2 - 7)/20,$$

$$\Sigma t^6 = \Sigma t^2\, (3k^4 - 18k^2 + 31)/112, \tag{8}$$

то, обозначив с целью сокращения записей

$$a^2 = \Sigma t^2, \quad a_4 = \Sigma t^4, \quad a_6 = \Sigma t^6 \tag{9}$$

и учитывая соотношение (6), получаем

$$\lambda(k, 1, l) = \frac{1}{k} + \frac{t_l^2}{a^2},$$

$$\lambda(k, 2, \sigma) = \frac{t_l}{a_2} + \frac{a_4 - 2a_2 t_l^2 + k t_l^4}{k a_4 - a_2^2}, \tag{10}$$

$$\lambda(k, 3, l) = \frac{a_4 - 2a_2 t_l^2 + k t_l^4}{k a_4 - a_2^2} + \frac{a_6 t_l^2 + 2a_4 t_l^4 + a_2 t_l^6}{a_2 a_6 - a_4^2};$$

$$t_l = (k-1)/2 + l. \tag{11}$$

Результаты вычислений приведены в табл. 1.

Коэффициенты $\lambda(k, n, l)$ еще не позволяют указать долю дисперсии $\sigma^2_{y^n}$ тренда y^n в общей дисперсии σ^2_S, так как в соотношении (5) величина σ^2_ε, вычисляемая по формуле (3), зависит от степени n_2 аппроксимирующего полинома.

Заменим в (5) дисперсию σ^2_ε разностью дисперсий σ^2_S, $\sigma^2_{y^n}$ согласно (2).

В результате получим

$$\sigma^2_{y^n} = \frac{\lambda(k, n, l)}{1 + \lambda(k, n, l)}\, \sigma^2_S = \beta(k, n, l)\, \sigma^2_S. \tag{12}$$

В последнем соотношении коэффициент $\beta\ (k, n, l)$ полностью характеризует вклад дисперсии тренда $\sigma^2_{y^n}$ в общую дисперсию σ^2_S в зависимости от аргументов k, n, l (табл. 2).

Из рассмотрения табл. 1, 2 следует:

1. Полиномиальные модели низших степеней обеспечивают бо́льшую достоверность экстраполированных уровней аппроксимируемых ими трендов

$$\beta(k, 1, l) < \beta(k, 2, l) < \beta(k, 3, l). \tag{13}$$

2. С увеличением циклов наблюдений (уровней ряда) уменьшается вклад дисперсии тренда σ_y^{2n} в общую дисперсию σ_ξ^2 экстраполированного на l шагов значения уровня ряда:

$$\beta(i, n. l) < \beta(j, n, l), \quad i > j. \tag{14}$$

Таблица 1

Параметр λ (*k*, *n*, *l*)

n	1			2			3		
	l			*l*			*l*		
k	0	1	2	0	1	2	0	1	2
3	0.83	2.33	4.83	1.00	19.00	—	—	—	—
5	0.60	1.10	1.80	0.89	4.60	—	1.01	—	—
7	0.46	0.71	1.04	0.76	2.43	6.28	0.93	8.64	—
9	0.38	0.53	0.71	0.66	1.62	3.50	0.86	4.71	—
11	0.32	0.42	0.54	0.58	1.21	2.31	0.79	3.15	10.10
13	0.27	0.35	0.43	0.52	0.96	1.68	0.73	2.33	6.40
15	0.24	0.30	0.36	0.46	0.79	1.30	0.67	1.81	4.50
17	0.22	0.26	0.30	0.42	0.68	1.05	0.62	1.52	3.38
19	0.19	0.23	0.26	0.39	0.59	0.88	0.58	1.28	2.66
21	0.18	0.20	0.24	0.36	0.52	0.75	0.54	1.11	2.16
31									1.05
...							...	...	...
51							0.27	0.36	0.49

Таблица 2

Параметр β (*k*, *n*, *l*)

n	1			2			3		
	l			*l*			1		
k	0	1	2	0	1	2	0	1	2
3	0.45	0.70	0.83	0.50	0.95	—	—	—	—
5	38	52	64	47	82	—	0.50	0.97	—
7	32	42	51	43	71	0.86	48	90	—
9	28	35	42	40	62	78	46	82	—
11	24	30	35	37	55	70	44	75	0.91
13	21	26	30	34	49	63	42	70	86
15	19	23	26	32	44	56	40	64	82
17	18	21	23	30	40	51	38	60	77
19	16	19	21	28	37	47	37	56	73
21	15	17	19	26	34	43	35	53	68
31									61
...									...
51							21	26	33

3. Для рядов, содержащих менее пяти циклов измерений, экстраполяция не имеет смысла для трендов, аппроксимируемых полиномами любой степени, исключая нулевую.

$$\sigma_\varepsilon^2 = \sum_i (S_i - y_i^n)^2/(k-n-1); \tag{3}$$

$$\sigma_S^2 = \sum_i (S_i - \bar{S})^2/(k-1), \tag{4}$$

$$\bar{S} = \sum_i S_i/k.$$

Соотношение (4) есть частный случай из (3) при $n=0$.

Для оценки дисперсии тренда σ_{y^n} воспользуемся известным выражением *

$$\sigma_{y^n}^2 = \sigma_\varepsilon^2\, t_L^T (T^T T)^{-1} t_L, \tag{5}$$

где

$$T = \begin{vmatrix} 1 & t_1 & t_1^2 & \cdots & t_1^n \\ 1 & t_2 & t_2^2 & \cdots & t_2^n \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ 1 & t_k & t_k^2 & \cdots & t_k^n \end{vmatrix} -$$

матрица моментов времени наблюдений;

$$t_L = (1 \quad t_L \quad t_L^2 \cdots t_L^n) -$$

вектор размерностью $n+1$; t_L — интервал экстраполяции.

Произведение матриц $T^T T$ имеет вид

$$T^T T = \begin{vmatrix} k & \Sigma t & \Sigma t^2 & \cdots & \Sigma t^n \\ \Sigma t & \Sigma t^2 & \Sigma t^3 & \cdots & \Sigma t^{n+1} \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ \Sigma t^n & \Sigma t^{n+1} & \Sigma t^{n+2} & \cdots & \Sigma t^{2n} \end{vmatrix},$$

4. С увеличением интервала экстраполяции дисперсия $\sigma_{y^n}^2$ возратает с бо́льшей интенсивностью для трендов, аппроксимируемых полиномами более высокой степени.

Если в (5) принять $\sigma_{y^n}^2 = \sigma_\varepsilon^2$, т. е. дисперсия тренда (неслучайная часть) равна дисперсии ошибок измерений (случайная часть), то экстраполяцию трендов на 1 шаг вперед следует выполнять полиномом первой степени в рядах при $k \leqslant 13$, второй степени при $13 < k \leqslant 21$ и третьей степени при $k > 21$. Если экстраполяция выполняется на два шага вперед, тогда полиномы второй степени следует применять при $17 < k \leqslant 31$, третьей степени при $k > 31$.

---

* *Четыркин Е. М., Калихман И. Л.* Вероятность и статистика. М., 1982.

Найденные по (12) дисперсии σ_y^{2n} рекомендуется использовать при построении доверительных интервалов экстраполированных уровней трендов на один и два шага вперед в равномерных рядах наблюдений.

Статья поступила в редколлегию 31.03.89