

Масштабирование при применении метода взвешенной регрессии

к.т.н. Ивашнев Л.И., Горбачёва С.Л.

Аннотация. В статье представлен вариант метода взвешенной регрессии, названный методом масштабирования, причем этот метод позволяет получать уравнения регрессии для функций, которые на области существования имеют точки разрыва. Этот метод также позволяет получать уравнения регрессии, описывающие переходные процессы возле точек разрыва и расширяет инструментарий математической статистики, расширяют диапазон их применения, повышают точность и эффективность получаемых результатов исследований.

Ключевые слова: математическая статистика, варианты, метод взвешенной регрессии, метод масштабирования, инструментарий, точность, эффективность.

В статье [4] показано, что применение метода взвешенной регрессии может быть весьма результативным. В ней показано, что, используя метод взвешенной регрессии, не меняя исходной выборки, можно получить уравнение регрессии методом наименьших квадратов по части выборки, исключив из нее значения, выходящие за допустимые пределы или потерявшие достоверность. Для этого надо обнулить значения весовых коэффициентов строк выборки, искажающих статистическую закономерность, а остальным весовым коэффициентам присвоить значение 1.

Более того, устанавливая различные значения весовым коэффициентам, можно задавать важности или достоверности строкам выборки и получать взвешенные уравнения регрессии. Используя метод взвешенной регрессии, можно также получать и нелинейные уравнения регрессии.

Однако в указанной статье не нашел отражения еще один вариант применения метода взвешенной регрессии – масштабирование исходной выборки для обработки статистической информации, т.е. взвешивание строк исходной выборки только по значениям показателя Y . Это может потребоваться для линеаризации числового ряда значений показателя Y , если эта последовательность содержит точки разрыва или не является линейной.

Действительно, вспомним денежную реформу 90-х годов, когда из-за галопирующей инфляции масштаб денежной единицы был увеличен в 1000 раз, т.е. из цен убрали 3 нуля. Такое же изменение денежного масштаба было и при Хрущеве, в 60-х годах прошлого столетия, но только в 10 раз. В итоге таких деноминаций числовые ряды экономических показателей становились совершенно несопоставимыми, хотя экономические процессы в стране оставались прежними, просто менялся масштаб цен и масштабы экономических показателей.

Для того, чтобы сделать масштабы экономических показателей сопоставимыми можно использовать метод взвешенной регрессии. Для этого к исходной выборке:

$$\left\{ \begin{array}{l} y_1, x_{11}, x_{21}, \dots, x_{m1} \\ \dots\dots\dots \\ r y_j, x_{1j}, x_{2j}, \dots, x_{mj} \cdot \\ \dots\dots\dots \\ r y_n, x_{1n}, x_{2n}, \dots, x_{mn} \end{array} \right. \quad (3)$$

Далее полученная выборка (1) обрабатывается либо методом наименьших квадратов, либо любым другим методом множественной регрессии. Если используется метод наименьших квадратов, система нормальных уравнений имеет вид:

$$\left\{ \begin{aligned} na_0 + a_1 \sum_{i=1}^n x_{1i} + a_2 \sum_{i=1}^n x_{2i} + \dots + a_m \sum_{i=1}^n x_{mi} &= \sum_{i=1}^n r_i y \\ a_0 \sum_{i=1}^n x_{1i} + a_1 \sum_{l=1}^n x_{1l} x_{li} + a_2 \sum_{i=1}^n x_{1i} x_{2i} + \dots + a_m \sum_{i=1}^n x_{1i} x_{mi} &= \sum_{i=1}^n r_i x_{1i} y_i \\ a_0 \sum_{i=1}^n x_{2i} + a_1 \sum_{l=1}^n x_{2l} x_{li} + a_2 \sum_{i=1}^n x_{2i} x_{2i} + \dots + a_m \sum_{i=1}^n x_{2i} x_{mi} &= \sum_{i=1}^n r_i x_{2i} y_i \\ &\dots \\ a_0 \sum_{i=1}^n x_{mi} + a_1 \sum_{l=1}^n x_{ml} x_{li} + a_2 \sum_{i=1}^n x_{mi} x_{2i} + \dots + a_m \sum_{i=1}^n x_{mi} x_{mi} &= \sum_{i=1}^n r_i x_{mi} y_i \end{aligned} \right. \quad (4)$$

где: $a_0, a_1, a_2, \dots, a_m$ – коэффициенты искомого линейного уравнения регрессии, r_1, r_2, \dots, r_{j-1} равны 1, а r_j, r_{j+1}, \dots, r_n равны соответствующему коэффициенту масштабирования, например 10 или 1000.

Если критической точкой является некоторое событие, изменившее ход экономических процессов, начиная, например, с некоторого года j , то изменения экономического процесса могут изображаться изменениями коэффициента масштабирования r , начиная с года j . При этом для лет, раньше критической точки (до года j), следует использовать уравнение регрессии с коэффициентами $a_0, a_1, a_2, \dots, a_m$, а для последующих лет – использовать коэффициенты, в r раз меньшие, т.е. $a_0 / r, a_1 / r, a_2 / r, \dots, a_m / r$.

Таблица 1.

Получение уравнения регрессии выручки от года и числа покупателей, если в 5-й год денежная единица деноминирована в 10 раз

	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Масшт	Год	Покупат..	Выручка	Ур		12	*a0+	78	*a1+	5511	*a2=	12122970	
2	1	1	513	1152470	1077664		78	*a0+	650	*a1+	35348	*a2=	78821761	
3	1	2	492	1025094	1043835		5511	*a0+	35348	*a1+	2573103	*a2=	5,65E+09	
4	1	3	542	1076225	1146419			a0+	6,5	*a1+	459,25	*a2=	1010248	
5	1	4	460	994952	995391,5			a0+	8,333333	*a1+	453,1795	*a2=	1010535	
6	10	5	470	100469,5	102112,3	*		a0+	6,414081	*a1+	466,9031	*a2=	1024391	
7	10	6	401	93665,1	89507,3				1,833333	*a1+	-6,07051	*a2=	287,8782	
8	10	7	386	87811,5	87277,24				-1,91925	*a1+	13,72362	*a2=	13855,87	
9	10	8	356	79095,02	82165,25					a1+	-3,31119	*a2=	157,0245	
10	10	9	419	91586,84	94921,25					a1+	-7,1505	*a2=	-7219,41	
11	10	10	442	102515,1	99992,1						-3,83931	*a2=	-7376,43	
12	10	11	472	107507,6	106407,8	1-4 год:	a0=	85522,98	a1=	6518,779	a2=	1921,29		
13	10	12	558	124772,2	123582,8	5-12 год:	a0=	8552,298	a1=	651,8779	a2=	192,129		
14														

Примером получения такой зависимости может служить распечатка экрана компьютера (таблица 1), содержащая расчет уравнения регрессии, характеризующего такой процесс деноминации, когда в 5-й год произошла деноминация денежной единицы, при которой масштаб цен понизился в 10 раз.

В соответствии с полученными результатами для 1 – 4-го года получено уравнение регрессии $Y = 85623 + 6518,8 \cdot \Gamma + 1921,29 \cdot \mathcal{U}$, а для последующих лет уравнение регрессии имеет вид: $Y = 8562,3 + 651,88 \cdot \Gamma + 192,129 \cdot \mathcal{U}$. График такого уравнения регрессии показан на рисунке 1.

С помощью масштабного коэффициента можно, например, определить переходные процессы, которые возникают при резких скачках в экономических системах, как и в электрических цепях с большими индуктивностями. В таких системах, как и в электрических цепях, при резких изменениях показателя возникают переходные процессы, иногда даже колебания, которые кроме как масштабированием отобразить очень сложно.

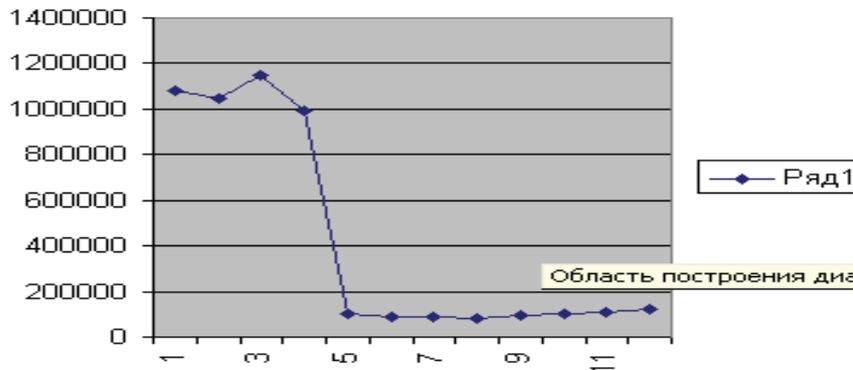


Рисунок 1. График уравнения регрессии

$$Y = \begin{cases} 85623 + 6518,8 \Gamma + 1921,29 \text{ Ч, если } 0 < t < 5 \\ 8562,3 + 651,88 \Gamma + 192,129 \text{ Ч, если } t \geq 5, \end{cases}$$

5-й год которого является точкой разрыва

Пример такого процесса представлен в таблице 2, где столбец В – порядковый номер месяца, столбец С – численность трудящихся на предприятии, столбец D – выручка предприятия за произведенную продукцию. Кроме того, таблица 2 содержит столбец масштабных множителей А, которые используются для линеаризации экономических процессов, произошедших при реструктуризации указанного предприятия, которая состояла в изменении технологии производства на предприятии, сопровождавшейся заменой технического обеспечения и сокращением штатов.

Таблица 2.

Результаты реструктуризации на предприятии, проведенной в 8 - 9 месяце

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q
1	Масшт.	Мес.	Числ.	Выручка						12	*a0+	78	*a1+	1317	*a2=	578410	576480
2	R	f	Ч, чел	S, т.руб	Y _{1р}	Y _{2р}	S-Y _{1р}	S-Y _{2р}		78	*a0+	650	*a1+	8263	*a2=	4211620	4142300
3	1	1	105	31500	27214,23	30602,02	4285,77	897,984		1317	*a0+	8263	*a1+	146617	*a2=	62864580	62648030
4	1	2	112	33120	32395,67	34620,97	724,3303	-1500,97			a0+	6,5	*a1+	109,75	*a2=	48200,83	48040
5	1	3	117	36250	37131,99	38363,41	-881,994	-2113,41			a0+	8,333333	*a1+	105,9359	*a2=	53995,13	53106,41
6	1	4	123	41070	42090,88	42244,11	-1020,88	-1174,11			a0+	6,274108	*a1+	111,3265	*a2=	47733,17	47568,74
7	1	5	124	46070	45936,97	45433,51	133,0294	636,4947				1,833333	*a1+	-3,8141	*a2=	5794,295	5066,41
8	1	6	128	51200	50450,74	49037,68	749,2624	2162,318				-2,05923	*a1+	5,390602	*a2=	-6261,96	-5537,67
9	1	7	128	52500	54074,27	52088,82	-1574,27	411,1816					a1+	-2,08042	*a2=	3160,524	2763,497
10	1,3	8	95	41000	50353,42	38905,67	-9353,42	2094,335					a1+	-2,61778	*a2=	3040,931	2689,201
11	1,1	9	95	51700	53976,95	48753,18	-2276,95	2946,818						-0,53736	*a2=	-119,594	-74,2958
12	0,9	10	96	60500	57823,05	63131	2676,951	-2631			a0=	13033,55	a1=	3051,136	a2=	138,2603	
13	0,9	11	96	64500	61446,59	66521,15	3053,413	-2021,15			a0=	222,1641	a1=	3623,537	a2=	222,5574	
14	0,9	12	98	69000	65515,24	70218,54	3484,761	-1218,54			ско1=	290,3968	ско2=	101,5453			
15																	

Эта реструктуризация привела в 8 месяце к значительному снижению выручки, которая восстановилась в 10 – 12 месяцах, причем рост выручки не только восстановился, но и темп ее роста существенно увеличился.

Столбец E этой таблицы, т.е. значения Y_{1р} изображают расчетные значения выручки, полученные методом наименьших квадратов, используя столбцы B, C и D исходной выборки. Столбец F, т.е. значения Y_{2р} получены по тем же данным, но используя масштабные ко-

эффиценты из столбца А. Столбцы G и H содержат абсолютные отклонения фактической выручки от ее расчетных значений Y_{1p} и Y_{2p} . В правой части таблицы 2 показан процесс получения уравнений регрессии, т.е. составлены и решены системы нормальных уравнений, причем диапазон ячеек J1:P12 показывает как получено уравнение регрессии методом наименьших квадратов, причем полученные коэффициенты уравнения регрессии представлены в ячейках L12, N12 и P12.

Левая часть системы нормальных уравнений для получения масштабированного уравнения регрессии (до знаков равенства) совпадает с левой частью системы нормальных уравнений для метода наименьших квадратов, а правая часть помещена в столбце Q1:Q11. Коэффициенты полученного масштабированного уравнения регрессии пред L13, N13 и P13.

Полученные уравнения регрессии имеют вид:

- для всех строк выборки по методу наименьших квадратов

$$Y = 13033,55 + 3051,136 \cdot t + 138,2603 \cdot Ч ,$$

- по методу масштабирования

$$Y / r = 222,1641 + 3623,537 \cdot t + 222,5574 \cdot Ч ,$$

где значения масштабного коэффициента заданы в столбце А.

Таким образом, для метода масштабирования определено 4 уравнения регрессии:

- для строк 1-7 получено уравнение

$$Y = 222,1641 + 3623,537 \cdot t + 222,5574 \cdot Ч ,$$

- для строки 8 - уравнение

$$Y / 1,3 = 222,1641 + 3623,537 \cdot t + 222,5574 \cdot Ч ,$$

- для строки 9 –

$$Y / 1,1 = 222,1641 + 3623,537 \cdot t + 222,5574 \cdot Ч ,$$

- для строк 10 - 12 –

$$Y / 0,9 = 222,1641 + 3623,537 \cdot t + 222,5574 \cdot Ч .$$

Графики полученных зависимостей представлены на рисунке 2.

Ряд 1 на этом рисунке показывает ежемесячный рост фактической выручки предприятия, ряд 2 – линия, изображающая линейное уравнение регрессии, полученное методом наименьших квадратов как функция $Y=f(t, Ч)$. Ряд 3 – это график функции, полученной методом взвешенной регрессии с использованием масштабирования выручки. Легко видеть, что масштабированная числовая последовательность (ряд 3) практически совпадает с исходной выручкой предприятия (ряд 1). Таким образом, метод масштабирования позволяет более точно описывать исследуемую зависимость, чем метод наименьших квадратов.

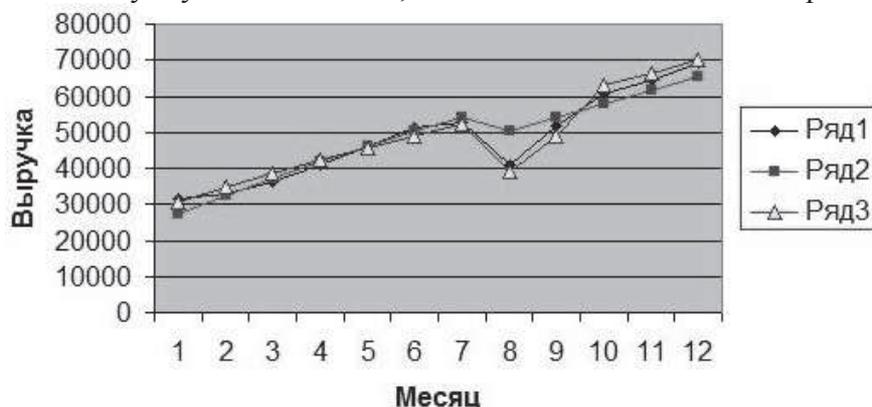


Рисунок 2. Графики фактической и расчетной выручки предприятия

Это подтверждает и расчет среднеквадратичного отклонения. В ячейках L14 и M14 даны оценки, полученные по формуле $\sigma = \sqrt{Q/n}$, где Q – сумма квадратов отклонений, а n

– число строк выборки. При этом получено, что среднеквадратичное отклонение уравнения, полученное методом наименьших квадратов 290,4, а методом масштабирования – всего 101,55, то есть почти в 3 раза меньше. Это показывает, что применение метода масштабирования может существенно повысить точность получаемых уравнений регрессии, и главное, обрабатывать статистическими методами числовые ряды, содержащие точки разрыва и значительные скачки исследуемых показателей.

Литература

1. Ивашнев Л.И. Методы и модели в экономике: Учеб. пособие. – М.: Издательский дом «Лидер-М», 2011. – 328 с.
2. Ивашнев Л.И. Методы регрессии в экономической математике: Монография. – М.: Изд-во МГОУ, 2005. – 200 с.
3. Ивашнев Л.И. Методы и модели в экономике: Учеб. пособие. – М.: Изд-во МГОУ, 2004. – 110 с.
4. Ивашнев Л.И., Горбачёва С.Л. Варианты метода взвешенной регрессии. – М.: Вестник МГОУ, Серия «Экономика и право», 2013, № 2.