

Баландин Игорь Константинович
Balandin Igor Konstantinovich

Институт экономики и антикризисного управления.
Российская Федерация, г. Москва.
Кафедра антикризисного управления и менеджмента.
e-mail : i_konsta @ mail. ru

The Institute of the economics and anti – crisis management.
The chair of the anti – crisis management.

О некоторых специальных методах сглаживания временных рядов на основе использования скользящих средних.

About some special method smoothing of time series on the base of using of the slipping average.

Аннотация. В данной статье рассматриваются вопросы сглаживания временных рядов и статистического прогнозирования на основе применения скользящих средних. Модификацией использования скользящих средних является метод «кумулятивных сумм».

Abstract. In this article examine some questions of the smoothing of time series and statistical forecasting on the base of using of the slipping average. The modification of the method using of the slipping average is method of cumulative sum.

Ключевые слова: временной ряд, скользящая средняя, метод «кумулятивных сумм», среднесрочное прогнозирование.

Keywords: time series, slipping average, method of cumulative sum, middle forecasting.

Из теории статистики известны методы и приемы сглаживания тенденции во временном ряде под общим названием «расчет скользящей средней» [2] с. 326, [3] с. 300, [4] с. 172. Суть методики – значения временного ряда заменяются некими величинами, определяемыми по формуле простого или взвешенного скользящего среднего:

$$\bar{y}_t = \sum_{i=t-p}^{t+p} y_i / (2p+1) \quad (1),$$

где : \bar{y}_t - значение скользящей средней для момента времени t (1,2... m);
 y_t - фактическое значение уровня временного ряда в момент i ;
 m – активная зона (период) сглаживания (как правило, $m < n$, где n – количество наблюдений в ряду динамики);
 p - величина, определяемая исходя из продолжительности ряда и величины m .

Активная зона сглаживания может определяться четным или нечетным числом.

Предпочтительнее пользоваться нечетными числами при определении длины m , так как в этом случае сглаженное значение оказывается в центре интервала и это позволяет его соотнести с конкретным значением временного ряда. В случае использования четного числа сглаженное значение попадает в промежуток между последовательными членами динамического ряда.

Известно, что при использовании скользящей средней в сглаженном ряде теряется

часть информации. Значение величины p информирует о том, сколько теряет членов выравненный ряд в начале и в конце ряда фактических значений.

Между m и p существует зависимость:

для четного значения m : $m = 2p$; для нечетного $m = 2p + 1$, тогда :

$$p = m / 2, \quad p = m / (2p + 1).$$

В случае применения скользящих средних временный ряд удается избавить от циклических наслоений, более четко обозначить наличие гладкой компоненты в ряде и обоснованно определиться с прогнозом. Механизм вычисления скользящих средних довольно трудоемкий. Количество расчетов увеличивается в зависимости от длины ряда и длины периода сглаживания. Однако имеются соответствующие программы расчетов, например, в оболочке EXCEL.

В формуле (1) знаменатель представляет собой вес каждого значения Y_i в зоне m , причем одинаковый для все значений, поэтому формула (1) характеризует расчеты на основе простой скользящей средней. Использование взвешенных скользящих средних предполагает определенное изменение весовой функции в пределах m .

Известен ряд систем формирования весовых функций, в том числе:

- система Б. Я Ястремского [1] ;
- система Дж. Э. Юла [6];
- система Вульгауса – Шеппарда – Четверикова [5];
- система экспоненциальных весов Брауна – Майера – Д. Эзопо [5] и др.

Выбор системы весов зависит от различных соображений, прежде всего, исходя из того, насколько сглаженный временный ряд близок по своим эволюциям к фактическим значениям ряда., например, для актуарных расчетов. Если же скользящие средние используют только для выявления трендовой составляющей, то вполне достаточно применить функцию с постоянными весами, т. е. простую скользящую среднюю.[2], [3]

В теории статистики известны методы применения скользящих средних для целей прогнозирования , которые позволяют, во – первых, сократить количество усредняемых значений временного ряда, во – вторых, четко отследить, когда начинает изменяться тенденция.

В монографии [4] рассматривается один из методов. К.Д. Льюис называет его «методом кумулятивных сумм» и определяет условия его применения в среднесрочном прогнозировании для тех случаев, когда все используемые данные имеют высокую степень «зашумления» [4] с.115.

Главной задачей в подобной ситуации следует считать выделение эффекта шума и вскрытие истинной картины изменения показателя.

Предположим, что спрос на какой – либо товар на протяжении 60 моментов времени находился на уровне 250 единиц. Однако фактически на 30 – м моменте произошло внезапное изменение спроса, которое визуальное (на специальном графике) заметить практически невозможно из – за того, что вариация временного ряда показателя сопровождается случайными колебаниями, т.е. данные зашумлены.

Для выявления подобных колебаний обратимся к методу «кумулятивных сумм». Для этого воспользуемся данными о спросе в таблице 1, зададимся уровнем прогноза, например, 90 единиц и определим для каждого момента времени отклонение (ошибку) фактического значения показателя относительно прогнозного.

Таблица 1

Расчет кумулятивных сумм #.

Момент времени	Спрос (един.)	Прогноз (един.)	Ошибка (един.)	Сумма предшествующих ошибок						
				S ₁	S ₂	S ₃	S ₄	S ₅	S ₆	S ₇
1	2	3	4	5	6	7	8	9	10	11
1	80	90	- 10	-10						
2	110	90	20	20	10					
3	105	90	15	15	35	25				
4	95	90	5	5	20	40	30			
5	68	90	-22	-22	-17	-2	18	8		
6	74	90	-16	-16	-38	-33	-18	2	-8	
7	71	90	-14	-19	-35	-57	-52	-37	-17	-27
$L_i = \sum_{j=1}^i -a/b * (i + b)$				+ 30 --	+ 40 --	+ 50 --	+ 60 --	+70 --	+ 80 --	+ 90 --

#) Взято из кн. К.Д. Льюиса Методы прогнозирования и экономических показателей.. М.: Финансы и статистика. 1986, с. 119.

Построим по данным таблицы 1 (эти данные выделены) график изменения кумулятивных сумм (CUSUM).

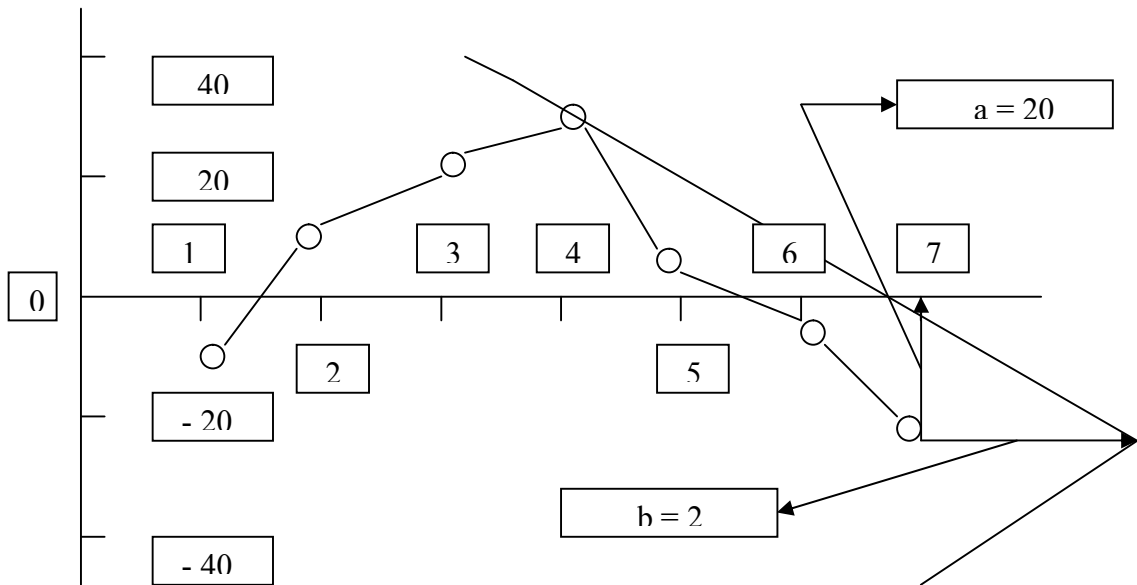


Рис 1. Наложение графика $V / MASK$ на график CUSUM.

Наложим на график кумулятивных сумм другой график, имеющий вид $>$ Назовем его как и в [] – $V / MASK$. Этот график представляет собой два отрезка прямых с угловым коэффициентами $-^+ a / b$, центр пересечения которых находится на расстоянии b вправо от текущей или последней точки временного ряда. Если график CUSUM пересекается нижним концом графика $V / MASK$, то это означает, что в момент пересечения произошел положительный сдвиг в уровне значений показателя, если график CUSUM пересекается с верхним концом $V / MASK$ – отрицательный. Если пересекаются нижний и верхний концы , то предпочтение отдается самому последнему моменту времени пересечения.

Чувствительность критерия , в том числе величина и искомого сдвига, определяется управляющими параметрами a и b . На практике значения a и b идентифицируются с с приближительной величиной сдвига и их количеством. Выбором разных значений этих параметров можно добиться разных положений $V / MASK$.

Из графика на рис 1 следует, что снижение значения временного ряда по отношению к его средней величине (отрицательный сдвиг) при $a = 20$ и $b = 2$ обнаруживается после четвертого момента времени. После того как момент сдвига определен, можно вычислить новый средний уровень ряда в виде среднего значения фактических уровней, соответствующих моментам времени после сдвига. По данным таблицы 1 – это среднее из трех последних значений временного ряда :

$$(68 + 74 + 71) / 3 = 71 \text{ един.}$$

Это означает .что послесдвиговый прогноз спроса составляет 71 единицу.

Метод кумулятивных сумм является критерием сравнения значений S_i (где i – число моментов времени назад от текущей или последней точки) с линейными границами L_i :

$$L_i = -^+ a / b * (i + b) \quad (2)$$

Для примера, приведенного в таблице 1, формула 2 приобретает следующий вид:

$$L_i = -^+ 20 / 2 * (i + 2) \quad (3)$$

Вернемся к анализу данных таблицы 1. В ней значение S_3 , соответствующее последнему моменту времени пересекает нижнюю границу L_3 , что означает падение спроса и его переход на новый уровень - 71един.

Можно привести пример кумулятивных сумм, когда изменение значений временного ряда по отношению к выбранному первоначально значению первичного прогноза будет положительным (т. е. превысит зафиксированный уровень). Эта ситуация следующим образом отразится на расчетах S_i и L_i (см. таблицу 2).

Таблица 2. Вычисление кумулятивных сумм.

Момент времени	Спрос (един.)	Прогноз (един.)	Ошибка (един.)	Сумма предшествующих ошибок						
				S_1	S_2	S_3	S_4	S_5	S_6	S_7
1	2	3	4	5	6	7	8	9	10	11
1	107	100	7	7						
2	94	100	-6	-6	1					
3	110	100	10	10	4	11				
4	92	100	-8	-8	2	-4	3			
5	116	100	16	16	8	18	12	19		
6	121	100	21	21	37	29	39	33	40	
7	123	100	23	23	44	60	54	62	56	63
$L_i = -^+ a/b * (i + b)$				± 36	± 45	± 54	± 63	± 72	± 81	± 90

Для заполнения последней строки таблицы 2 были использованы следующие данные: $a = 27$, $b = 3$. Эти параметры выбраны из следующих соображений: величина b , равная 3, обусловлена тем, что визуальное обследование временного ряда указывает на наличие в нем трех сдвигов при $t = 2$, $t = 4$, $t = 5$, а величина $a = 27$, выбрана, исходя из средней величины отклонений от прогноза, равного 100 единиц за три последние момента времени (т.е. 20 единиц) и соображения максимальной кратности числу 3. Числа 21 и 24 не подходят из-за меньшей точности отображения характеристики динамики ряда. Для выяснения этого были проведены соответствующие расчеты L_i для: $a = 21$ и $a = 24$.

Анализ данных таблицы 2 показывает, что величина S_3 для момента времени 7 пересекает верхнюю границу графика $V / MASK$ ($60 > 54$). Это означает, что спрос по крайней за последние три периода вышел на новый средний уровень выше величины 100 единиц. Подобный сдвиг начался с момента времени 5. Прогноз на следующий промежуток времени можно составить:

$$(116 + 121 + 123) / 3 = 120 \text{ единиц.}$$

Метод поиска изменений в динамике показателей, основанный на подсчете кумулятивных сумм, трудоемкий для реализации автоматической системы контроля особенно с применением, например, даже с помощью ПК, так как приходится запоминать предыдущие вычисления

Харрисон и Дэвис [7] предложили модификацию метода CUSUM, которая сокращает количество расчетов и легко встраивается в систему автоматических расчетов для контроля среднесрочного прогнозирования.

Модификация метода заключается в использовании двух уравнений следующего типа:

$$D_t = \min(D_{t-1}; a) + (a/b + e_t) \quad (4),$$

$$R_t = \min (R_{t-1} ; a) + (a / b - e_t) \quad (5),$$

где :
 а и b – параметры V / MASK;
 e_t - текущая или последняя прогностическая ошибка прогноза;
 $(D_{t-1} ; a)$ – минимальное значение из предыдущего значения D_t и а;
 $(R_{t-1} ; a)$ - минимальное значение из предыдущего значения R_t и а.

Критерии работают следующим образом: либо R_t , либо D_t меньше 0, то изменение в уровне следует считать обнаруженным (это эквивалентно пересечению графиков CUSUM и V / MASK). Если $D_t < 0$, то следует ожидать снижение среднего уровня спроса, если $R_t < 0$, то - повышение спроса. Реализуем вычислительную схему, определяемую формулами (4) и (5) на примере данных таблицы 2, используя следующую информацию: $a = 27$, $b = 3$, первоначальный прогноз 100 единиц, начальные значения $\min (R_t ; a) = 27$, $\min (D_t ; a) = 27$ единиц (см. таблицу 3).

Таблица 3 Реализация метода кумулятивных сумм по Харрисону и Дэвису.

Момент времени	Фактическое наблюдение	Прогноз	Ошибка	$a/b-t$	$\min (R_{t-1})$	R_t	$a/b+t$	$\min (D_{t-1})$	D_t
1	2	3	4	5	6	7	8	9	10
1	107	100	7	2	27*	29	16	27*	43
2	94	100	-6	15	27	42	3	27	30
3	110	100	10	-1	27	26	19	27	46
4	92	100	-8	17	26	43	1	27	28
5	116	100	16	-7	27	20	25	27	52
6	121	100	21	-12	20	8	30	27	57
7	123	100	23	-14	8	-6	32	27	59

*Начальные значения

$R_t < 27$

Данные таблицы 3 показывают, что предложенная схема достаточно чутко улавливает изменения в значениях ряда. Если соблюдается соотношение :

$$R_t - D_t < 0 \quad (6),$$

то наблюдается сокращение значения ряда (и по отношению к прогнозу). При соотношении:

$$R_t - D_t > 0 \quad (7)$$

фиксируется увеличение значения ряда. Начиная с пятого момента времени, значения R_t меньше 27, кроме того, величина R_t для седьмого момента времени отрицательна, что дополнительно подтверждает тезис о росте значений в конце периода наблюдений.

Результаты использования метода кумулятивных сумм полностью зависят от выбора параметров а и b V / MASK. Для точного определения этих параметров не существует простого и стандартного правила. Параметры а и b находят эмпирически. Для имеющегося временного ряда при выбранном размере сдвига, т.е. отклонения от среднего уровня ряда строится график кумулятивных сумм, затем при заданном значении b изменением половины угла V / MASK необходимо добиться того, чтобы график V / MASK пересек график CUSUM хотя бы в одной точке. Это позволит получить значение параметра а.

В заключении данной заметки хотелось бы продемонстрировать достаточно высокую чувствительность расчетной схемы Харрисона и Дэвиса в зависимости от корректного выбора начальных параметров и величин а и b V / MASC. Рассмотрим временный ряд спроса на товар в тоннах за 10 десять промежутков времени, который изобилует резкими

«up and down» (см. таблицу 4).

Таблица 4. Изменение величины спроса

Спрос в тоннах	Промежутки времени Y_t									
	1	2	3	4	5	6	7	8	9	10
1	2	3	4	5	6	7	8	9	10	11
Y_t	40	21,8	46,4	25,7	59,6	41	30	68,2	28,2	24,6

В данном ряде 2.4 и 10 –е наблюдения явно ниже первоначального прогноза , а 5 и 8 – е – выше. В качестве первоначального прогноза для анализируемого ряда использована величина 33,27 тонн – средняя величина ряда из восьми первых значений ряда. Расчеты прогноза по схеме Харрисона и Дэвиса для двух вариантов значений параметров $V / MASC$:

I вариант $a = 24,4$, $b = 2$,

II вариант $a = 16$, $b = 2$.

К сожалению, первый вариант оказался практически нечувствительным к фактическим изменениям временного ряда, так как при расчетах R_t , D_t не было зафиксировано отрицательных величин (что свидетельствует о необходимости тщательного подбора параметров $V / MASC$). Второй вариант показал определенную чувствительность к изменению данных временного ряда при использовании метода Харрисона и Дэвиса (см. таблицу 5)

Таблица 5 Прогнозирование спроса по методу Харрисона и Дэвиса

Время	Y_t	R_t	D_t	\hat{Y}_t
1	2	3	4	5
1	40	17,25	30,75	33,25
2	21,8	34,45	12,55	33,25
3	46,4	10,85	33,7	33,25
4	25,7	26,4	16,45	33,25
5	59,6	-2,35	50,35	42,65
6	41,0	7,3	22,35	42,65
7	30,0	27,95	11,35	42,65
8	68,2	-1,55	44,9	49,1
9	28,2	27,35	3,1	49,1
10	24,0	49,1	-14,1	26,4

Графа 5 в таблице показывает сглаженный временной ряд с использованием метода Харрисона и Дэвиса. Наиболее вероятным прогнозным значением величины спроса для $t = 11$ будет 26,4 тонны. Эта величина представляет собой среднее из двух последних членов временного ряда. В целом использование данной вычислительной схемы позволило обнаружить 3 значительных сдвига. Это может быть проиллюстрировано рисунком 2.

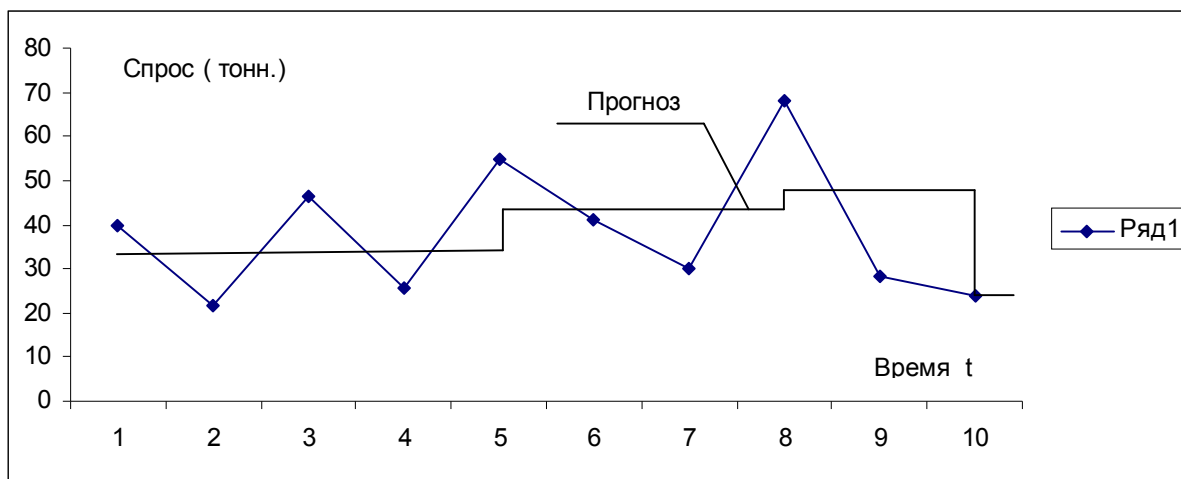


Рис 2. Прогноз спроса по методу Харрисона и Дэвиса.

Ступенчатая линия, изображенная на рисунке 5, названа в источнике [4] «Диаграммой Манхэттена»

Рассмотренная схема весьма легко может быть реализована на персональном компьютере с добавлением расчетов верхней и нижней границы прогноза, согласно современным представлениям о верификации подобных вычислений.

Таким образом, исследование возможностей совершенствования среднесрочного прогнозирования на базе метода кумулятивных сумм позволяет сделать некоторые выводы и рекомендации:

1. Предлагаемый К. Д. Льюисом метод прогнозирования представляет собой модернизацию скользящей средней и может быть использован для временных рядов, включающих выраженную «шумовую» составляющую.
2. Его реализация в виде схемы Харрисона и Дэвиса имеет ряд преимуществ перед методом скользящей средней. Во-первых, сокращается количество вычислений самой средней; во-вторых данная схема более эффективно реагирует на начало изменений во временном ряду, предупреждая об этом исследователей (достаточно посмотреть на величину разницы между R_t и D_t).
3. Схема расчетов может быть модернизирована за счет изменения весовой функции при определении скользящих средних, а также построения доверительных интервалов прогноза.

Литература

1. Венецкий И.Г. , Венецкая В. И. Основные математико – статистические понятия и формулы в экономическом анализе. М.: Статистика. 1974, 275 с.
2. Елисеева И. И., Юзбашев М. М. Общая теория статистики. Москва: Финансы и статистика, 1998, 480 с.
3. Ефимова М. Р., Петрова Е. В., Румянцев В. Н., Общая теория статистики. М: Инфра – М, 1996. 416 с.
4. Льюис К. Д., Методы прогнозирования экономических показателей. М: Финансы и статистика. 1986. 133 с.
5. Сборник Статистический анализ экономических временных рядов и прогнозирование. М.: Наука , 1973 295 с.
6. Юл Дж. Э., Кенделл М. Теория статистика. М.: Госстатиздат. 1960, 376 с.
7. Harrison P. J., Davies O. L. The use cumulative sum (CUSUM) techniques for the control of routine forecasts of product demand – Operation Research, 12, 325, 1964.

