

РАЗДЕЛ II. ЭКОНОМИКО-МАТЕМАТИЧЕСКОЕ МОДЕЛИРОВАНИЕ

УДК 330.43
ББК 65в6

К ВОПРОСУ О МОДЕЛИРОВАНИИ ТРЕНДОВ ВРЕМЕННЫХ РЯДОВ

Ю.Л. Кошкин, канд. техн. наук, доцент кафедры математического моделирования в экономике

Электронный адрес: koshyuri@gmail.com

ФБГОУ ВПО, Вятский государственный университет, 610000, г. Киров, ул. Московская, 36

А.В. Шатров, докт. физ.-мат. наук, зав. кафедрой математического моделирования в экономике

Электронный адрес: avshatrov1@yandex.ru

ФБГОУ ВПО, Вятский государственный университет, 610000, г. Киров, ул. Московская, 36

В экономике и других сферах научной и практической деятельности мы видим интересующие нас объекты, развивающиеся во времени. Для моделирования таких объектов обычно используются эконометрические методы и представления исходной и результирующей информации в виде временных рядов. В настоящее время существует множество методов моделирования временных рядов. Многие методы, разработанные для решения конкретных задач, не являются универсальными. Очень часто исследователи используют динамическую декомпозицию временного ряда на несколько компонентов. Наиболее часто выделяют трендовую составляющую, циклическую составляющую и случайную составляющую. В настоящей работе первый метод (метод взвешенных тангенсов) предполагает разложение на трендовые и циклические компоненты. Второй метод не использует разложения, содержащего циклическую компоненту. Вместо этого метод фазовых трендов использует понятие «фазы», которые могут быть найдены в исходном виде временных рядов. Применение метода фазовых трендов позволяет выполнять кусочную аппроксимацию временных рядов. Современные методы не могут работать с короткими рядами, так как часть статистической информации теряется в предварительном сглаживании. Для коротких временных рядов можно применить метод взвешенных тангенсов при наличии хотя бы одного цикла. Многие методы не учитывают развития временных рядов с течением времени (эволюции). В этом случае мы предлагаем метод фазовых трендов, который, по мнению авторов, во многих случаях дает результаты, не уступающие по качеству при сравнении со сложными современными методами.

Ключевые слова: временной ряд, тренд, циклическая компонента, аддитивная модель, прогнозирование.

Введение

При анализе и построении моделей временных рядов (ВР), отражающих как циклическое, так и трендовое изменение интересующей величины, чаще всего на первом этапе выделяют отдельные компоненты модели в соответствии с имеющимися в наблюдениях исходными данными, а затем эти компоненты (кроме случайной) используют для прогнозирования. Выделение компонент в простейших случаях рассматривают в концепции тренд-сезонного моделирования в виде аддитивной (1) или мультипликативной (2) моделей [1; 3; 7; 8]:

$$y = T + S + E, \quad (1)$$

$$y = T \cdot S \cdot E, \quad (2)$$

где T – тренд, который будем пока считать линейным: $T = a + bt$;

S – циклическая компонента ВР;

E – остаточная компонента.

В [3] предлагается такая последовательность анализа аддитивной модели ВР:

1. Выравнивание данных исходного ВР (методом скользящей средней).
2. Расчёт значений S .
3. Расчёт значений $T + E = y - S$.
4. Расчёт T .
5. Расчёт значений $T + S$.
6. Расчёт и анализ E .

Естественно, что под результатами расчётов понимаются не сами значения компонент и параметров модели (они не наблюдаемы), а их оценки.

К одним из недостатков этого подхода можно отнести то, что в ходе шага 1 приходится терять часть исходной информации на фильтрацию циклической компоненты методом скользящей

средней, а эти потери могут снизить точность расчётов как в шаге 2, так и неизбежно в последующих шагах. Особенно остро этот недостаток отражается при малом объёме выборки (как по числу циклов, так и по числу наблюдений за один цикл).

Другим недостатком следует считать то, что, как правило, метод требует очень большого количества исходной информации (особенно по числу наблюдаемых циклов (периодов)), которой часто не хватает.

Следующим недостатком является то, что, выбор мультипликативной модели предполагает непостоянство амплитуды (возрастание или убывание) циклической компоненты в исходных данных, однако в результате расчётов по модели эта амплитуда оказывается неизбежно постоянной. При этом неизбежно увеличение остаточной суммы квадратов отклонений и несоблюдение условий теоремы Гаусса – Маркова в виде гетероскедастичности остатков, что не позволяет использовать метод наименьших квадратов (МНК) для расчёта параметров моделей.

И, наконец, можно отметить, что при достаточном количестве исходных данных последние используются неэффективно, что снижает показатели надёжности моделей и точность прогнозов.

В [9; 10] рассматриваются и более сложные модели, например смешанные аддитивно-мультипликативные, с использованием гармоник циклической составляющей, с использованием базисов Гребнера [3] и др. Хотя авторы и заявляют, что реализация методов не так сложна, но из-за увеличенного числа идентифицируемых параметров, ужесточаются требования к объёму статистической информации (по мнению авторов [4], на каждый параметр необходимо 6–7 наблюдений), а для построения базисов Гребнера необходимо (для получения достоверных результатов) использовать полиномы высоких степеней и многоразрядные числовые коэффициенты (в примере [8, с. 180–181] рассмотрен полином 20-го порядка с октовигинтиллионной разрядностью коэф-

фициентов). Для реализации таких методов нет разработанных пакетов прикладных программ.

В настоящей работе предлагаются существенно более простые методы.

При малом количестве исходных наблюдений рассматривается аддитивная модель (мультипликативную можно привести к ней логарифмированием) и предлагается модель взвешенных тангенсов, анонсированная в [5]. При достаточно большом числе наблюдений (достаточность должна быть обоснована) предложена анонсированная в [6; 7] модель фазовых трендов без выделения циклической компоненты в виде повторяющейся последовательности чисел. При этом появляется возможность учёта эволюционного развития циклической компоненты и стирается разница подходов в зависимости от амплитуды периодической составляющей, то есть в разделении моделей на аддитивные и мультипликативные.

Метод взвешенных тангенсов [5]

Для того чтобы не терять исходную информацию на фильтрацию перед выделением циклической компоненты (как это необходимо в описанной выше тренд-сезонной модели), в данной работе предлагается изменить последовательность расчётов и на первое место поставить шаг, бывший ранее под номером 4. Остальные шаги остаются, хотя шаг 2 специфичен, а номера шагов 5 и 6 уменьшаются теперь на 1.

При выполнении предложенной последовательности проблематичным становится шаг 1 (расчёт T), так как в прошлой методике после устранения S для расчёта T можно было применить метод наименьших квадратов (МНК). При наличии неслучайной составляющей в виде циклической компоненты МНК непосредственно не применим, поэтому предлагается альтернативный метод.

Так как метод ранее не применялся, то идею его создания изложим подробно и пока на простейших примерах линейной регрессии (без циклической компоненты). Сначала возникла идея метода средних тангенсов (МСТ). Пусть исходные данные отражены в табл.1, где i – номер наблюдения (точки пространства корреляции).

Таблица 1

Исходные данные для иллюстрации МСТ

i	1	2	3	4
x	1	2	3	4
y	1	2	4	5

Для приведённых данных МНК даст такое уравнение регрессии: $y = -0,5 + 1,4x + e$.

Идея МСТ состоит в том, что каждая линия, соединяющая любые пары точек поля корреляции (x_i, y_i) и (x_j, y_j) , является в некотором смысле линией регрессии, а тангенс угла α наклона этой линии по отношению к оси абсцисс отражает коэффициент регрессии для этой линии. Так как при объёме выборки, равном n ,

получается $\sum_{i=1}^{n-1} i = \frac{(n-1)n}{2}$ линий, то в

соответствии с МСТ на эту величину и нужно разделить сумму тангенсов по всем отрезкам, после чего получим средний коэффициент регрессии для всей совокупности исходных данных. В нашем случае $tg\alpha_{1-2} = 1$; $tg\alpha_{1-3} = 3/2$;

$$tg\alpha_{1-4} = 4/3; tg\alpha_{2-3} = 2;$$

$$tg\alpha_{2-4} = 3/2; tg\alpha_{3-4} = 1; \frac{(n-1)n}{2} = \frac{(4-1)4}{2} = 6.$$

$$\hat{b}_{MCT} = \frac{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n tg\alpha_{i-j}}{(n-1)n} = \frac{1+3/2+4/3+2+3/2+1}{6} = \frac{25}{18} \approx 1,39, \quad (3)$$

где первая сумма отражает перебор первых (левых) точек отрезка, вторая – вторых.

Свободный член можно вычислить по

$$\text{формуле } \hat{a}_{MCT} = \frac{\sum_{i=1}^n y_i - \hat{b}_{MCT} \sum_{i=1}^n x_i}{n} = \frac{12-13,9}{4} \approx -0,48,$$

где суммирование проводится по всем значениям соответствующих переменных табл. 1.

Таким образом, МСТ даёт уравнение регрессии $y = -0,48 + 1,39x + e$.

Как видим, МНК и МСТ дают близкие, но всё же различающиеся результаты.

Метод взвешенных тангенсов отличается от МСТ тем, что каждой линии «промежуточной регрессии» и её тангенсу (коэффициенту этой регрессии) присваивается вес, пропорциональный разности абсцисс образующих линию точек. При этом формула (3) преобразуется к виду

$$\hat{b}_{MBT} = \frac{2 \sum_{i=1}^{n-1} \sum_{j=i+1}^n \lambda_{i-j} tg\alpha_{i-j}}{(n-1)n}, \quad (3^*)$$

где вес λ_{i-j} частичного отрезка определяется равенством

$$\lambda_{i-j} = \frac{x_j - x_i}{\bar{L}} = \frac{(x_j - x_i)n}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_j - x_i)}, \quad (4)$$

а средняя длина $\bar{L}$ абсциссы отрезков равна

$$\bar{L} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_j - x_i)}{n}.$$

С учётом (4), а также того, что $tg\alpha_{i-j} = \frac{y_j - y_i}{x_j - x_i}$,

формула (3) примет вид

$$\hat{b}_{MBT} = \frac{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (y_j - y_i)}{\sum_{i=1}^{n-1} \sum_{j=i+1}^n (x_j - x_i)}. \quad (5)$$

Применим (5) к данным табл. 1:

$$\hat{b}_{MBT} = \frac{(2-1) + (4-1) + (5-1) + (4-2) + (5-2) + (5-4)}{(2-1) + (3-1) + (4-1) + (3-2) + (4-2) + (4-3)} = \frac{14}{10} = 1,4.$$

Для расчёта свободного члена применима формула (2), в соответствии с которой получим $\hat{a}_{MBT} = (12-14)/4 = -0,5$.

Полученные по МВТ результаты совпали с результатами по МНК. Случайно ли это? В общем случае да.

Если, например, имеем данные в соответствии с табл. 2, то получим одинаковые уравнения регрессии: $y = -3 + 3x + e$.

Таблица 2

Данные, для которых результаты по МВТ и МНК совпадают

i	1	2	3	4	5
x	1	2	2	3	4
y	1	2	3	5	10

Если же теперь незначимо поменять значения y при неизменных (повторяющихся) значениях x во втором и третьем столбцах (табл. 3), то уравнение регрессии по МНК не изменится, а

МВТ даст другое уравнение: $y = -2,66 + 2,86x + e$.

Таблица 3

Перестановка второго и третьего столбцов в табл. 2

i	1	3	2	4	5
x	1	2	2	3	4
y	1	3	2	5	10

В следующем разделе покажем, что если наблюдения по факторной переменной равноотстоящие (что характерно для ВР), то методы МНК и МВТ эквивалентны.

МВТ и МНК при равноотстоящих факторах

Как известно, оценка коэффициента регрессии с использованием МНК вычисляется по формуле

$$\hat{b}_{MНК} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}, \quad (6)$$

где $\bar{x}, \bar{y}$ – средние значения в наблюдениях фактора и результата.

Обозначим $\Delta x = x_{i+1} - x_i$ расстояние (шаг) между соседними наблюдениями фактора и будем считать его постоянным. Тогда

$$\bar{x} = \frac{x_1 + x_n}{2} = \frac{x_1 + x_1 + \Delta x(n-1)}{2} = x_1 + \frac{\Delta x(n-1)}{2}, \quad (7)$$

$$x_i - \bar{x} = x_i + \Delta x(i-1) - (x_1 + \Delta x(\frac{n-1}{2})) = \frac{\Delta x}{2}(2i - n - 1). \quad (8)$$

Подставим (8) в (6):

$$\hat{b}_{МНК} = \frac{\frac{\Delta x}{2} \sum_{i=1}^n (2i - n - 1)(y_i - \bar{y})}{\frac{\Delta^2 x}{4} \sum_{i=1}^n (2i - n - 1)^2} = \frac{2 \left[\sum_{i=1}^n (2i - n - 1)y_i - \bar{y} \sum_{i=1}^n (2i - n - 1) \right]}{\Delta x \sum_{i=1}^n (2i - n - 1)^2} \quad (9)$$

$$= \frac{2 \sum_{i=1}^n (2i - n - 1)y_i}{\Delta x \sum_{i=1}^n (2i - n - 1)^2},$$

$$\sum_{i=1}^n (2i - n - 1) = 2 \sum_{i=1}^n i - n^2 - n = 0. \quad (10)$$

Здесь учтено, что

$$\sum_{i=1}^n i = \frac{n^2 + n}{2}.$$

Рассмотрим сейчас коэффициент регрессии для МВТ. Чтобы учесть разницу смысла индексов в (5) и (9), перепишем (5) с другими обозначениями:

$$\hat{b}_{МВТ} = \frac{\sum_{p=1}^{n-1} \sum_{q=i+1}^n (y_q - y_{pi})}{\sum_{p=1}^{n-1} \sum_{q=i+1}^n (x_q - x_{pi})}, \quad (11)$$

где p – номер первой точки частичного отрезка, q – номер второй точки этого отрезка.

Для удобства сравнения (9) и (11) нужно преобразовать двойные суммы в (11) к одинарным. Перепишем пока числитель $Ч$ дроби в правой части (11):

$$\begin{aligned} Ч &= (y_2 - y_1) + (y_3 - y_1) + (y_4 - y_1) + \dots \\ &\quad + (y_{n-2} - y_1) + (y_{n-1} - y_1) + (y_n - y_1) + \\ &\quad + (y_3 - y_2) + (y_4 - y_2) + \dots + (y_{n-2} - y_2) + \\ &\quad \quad (y_{n-1} - y_2) + (y_n - y_2) + \\ &\quad + (y_4 - y_3) + \dots + (y_{n-2} - y_3) + (y_{n-1} - y_3) + \\ &\quad \quad (y_n - y_3) + \\ &\quad \quad \dots \\ &\quad + (y_{n-2} - y_{n-3}) + (y_{n-1} - y_{n-3}) + (y_n - y_{n-3}) + \\ &\quad \quad + (y_{n-1} - y_{n-2}) + (y_n - y_{n-2}) + \\ &\quad \quad + (y_n - y_{n-1}). \end{aligned}$$

Для удобства группировки все элементы со знаком «минус» просуммируем по строкам, а со знаком «плюс» – по столбцам, тогда

$$\begin{aligned} Ч &= (n-1)(-y_1) + (n-2)(-y_2) + (n-3)(-y_3) + \dots \\ &\quad + 3(-y_{n-3}) + 2(-y_{n-2}) + 1(-y_{n-1}) + \end{aligned} \quad (12)$$

$$\begin{aligned} &+ y_2 + 2y_3 + \dots + (n-4)y_{n-3} + (n-3)y_{n-2} + \\ &+ (n-2)y_{n-1} + (n-1)y_n = \\ &+ (1-n)y_1 + (3-n)y_2 + (5-n)y_3 + \dots \\ &+ (n-5)y_{n-2} + (n-3)y_{n-1} + (n-1)y_n = \sum_{i=1}^n (2i - n - 1)y_i. \end{aligned}$$

Аналогичные действия со знаменателем дроби в правой части (11) дают

$$З = \sum_{i=1}^n (2i - n - 1)x_i. \quad \text{С учётом (10) можно}$$

записать

$$З = \sum_{i=1}^n (2i - n - 1)(x_i - \bar{x}) = \sum_{i=1}^n (2i - n - 1) \left(\frac{\Delta x}{2} (2i - n - 1) \right) = \frac{\Delta x}{2} \sum_{i=1}^n (2i - n - 1)^2. \quad (13)$$

Деление $Ч$ на $З$ позволяет сделать вывод, что $b_{МНК} = b_{МВТ}$.

Так как свободный член для МНК и МВТ можно вычислить по одной и той же формуле (2), то при равноотстоящих значениях факторы МНК и МВТ эквивалентны.

Применение МВТ для анализа ВР

Данные ВР обычно наблюдаются через равные дискреты времени, что должно обеспечить применимость МВТ и его эффективность (минимум остаточной суммы квадратов по «наследству» от МНК в силу эквивалентности) при выделении тренда. Обычно заранее известна периодичность циклической составляющей в виде числа циклов (периодов) и числа наблюдений на один цикл (в экономических исследованиях это чаще всего 4 квартала или 12 месяцев).

Пусть наблюдения каждого периода циклической компоненты пронумерованы от $t = 1$ до $t = k_s$ (k_s – количество наблюдений за 1 цикл). С учетом этого можно определить время для всего ВР:

$$p = \overline{1, s_p}; \quad c = \overline{1, s_c}; \quad t = p + c(s_p - 1).$$

При наличии S-компоненты (циклической) для непосредственного выделения тренда можно использовать соответствующие точки t разных циклов. Тогда линии, соединяющие эти точки, будут частными линиями регрессии (трендов), а совокупность частных коэффициентов регрессии позволит найти общий коэффициент регрессии тренда ВР, например по МСТ, но лучше по МВТ. В последнем случае формулу (11) можно преобразовать к виду

$$\hat{b}_{МВТ} = \frac{\sum_{t=1}^{k_s} \sum_{p=1}^{k_c-1} \sum_{q=i+1}^{k_c} (y_q - y_{pi})}{k_s \sum_{p=1}^{k_s} \sum_{q=i+1}^{k_c} (q - p)}, \quad (14)$$

где k_c – количество циклов временного ряда;

p – номер цикла (периода ВР), в котором находится первая (левая по траектории) из точек с номером t отрезка частного тренда;

q – номер цикла для второй точки отрезка;

первый индекс в массиве y уровней ВР означает принадлежность к циклу, второй – номер точки внутри цикла.

Свободный член можно вычислить по формуле

$$\hat{a}_{MBT} = \frac{\sum_{n_c=1}^{k_c} \sum_{t=1}^{k_s} (y_{n_c,t} - \hat{b}_{MBT}(t + k_s(n_c - 1)))}{k_c \cdot k_s}, \quad (15)$$

где $n_c = \overline{1, k_c}$ – номер цикла.

Зная уравнение тренда $T = a + bt$, можно выделить сумму циклической компоненты с остатками $S + E = y - T$, но так как значения E

случайны и не повторяются из цикла в цикл, то для получения повторяющихся последовательностей значений S нужно каждую её точку t усреднить по всем циклам:

$$S(t) = \frac{\sum_{n_c=1}^{k_c} (S_{n_c,t} + E_{n_c,t})}{k_c}, t = \overline{1, k_s}. \quad (16)$$

Для сравнения приведём основные результаты расчётов для известного [4, с. 311-317] примера.

Исходные данные для примера приведены в табл. 4.

Таблица 4

**Потребление электроэнергии жителями Северо-Западного региона
за 16 кварталов 2003–06 гг. (млн кВт/ч, [4])**

i	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
y	6	4,4	5	9	7,2	4,8	6	10	8	5,6	6,4	11	9	6,6	7	10,8

Здесь i – это номер наблюдения, имеющий смысл «сквозного» (в пределах всего ВР, а не только одного цикла) дискретного времени. Предварительный анализ данных показывает, что данный ВР содержит возрастающий тренд, который будем считать линейным, а также циклическую

компоненту, содержащую $k_c=4$ периода по $k_s=4$ наблюдения в каждом.

Результаты для сравнения приведены в табл. 5. Данные табл. 4 и результатов расчета табл. 5 приведены на рис. 1.

Таблица 5

Результаты расчета по модели (1) и MBT

i	S	$T=5,715+0,186i$	E	$\hat{A}^2$	$T_{MBT}=5,813+0,175i$	S_{MBT}	E_{MBT}	E_{MBT}^2
1	0,518	5,901	-0,483	0,2333	5,9875	0,5125	-0,5	0,2500
2	-1,977	6,088	0,289	0,0835	6,1625	-1,8625	0,1	0,0100
3	-1,294	6,275	0,019	0,0004	6,3375	-1,2875	-0,05	0,0025
4	2,690	6,461	-0,151	0,0228	6,5125	2,6375	-0,15	0,0225
5		6,648	-0,029	0,0008	6,6875		0	0
6		6,834	-0,057	0,0032	6,8625		-0,2	0,0400
7		7,020	0,273	0,0745	7,0375		0,25	0,0625
8		7,207	0,104	0,0108	7,2125		0,15	0,0225
9		7,393	0,026	0,0007	7,3875		0,1	0,0100
10		7,580	-0,030	0,0009	7,5625		-0,1	0,0100
11		7,766	-0,072	0,0052	7,7375		-0,05	0,0025
12		7,952	0,358	0,1282	7,9125		0,45	0,2025
13		8,139	0,280	0,0784	8,0875		0,4	0,1600
14		8,325	0,252	0,0635	8,2625		0,2	0,0400
15		8,519	-0,218	0,0475	8,4375		-0,15	0,0225
16		8,698	-0,588	0,3457	8,6125		-0,45	0,2025
Σ	0	116,8	0	1,0994	116,8	0	0	1,0600

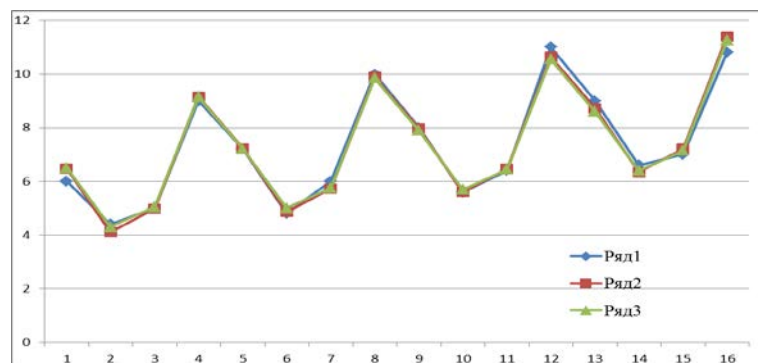


Рис. 1. Траектории временного ряда для примера [4]: ряд 1 – исходные данные, ряд 2 – расчет по модели (1), ряд 3 – расчет по MBT

Некоторые выводы:

1. Несмотря на то что результаты расчётов по известному и предложенному методам эквивалентны (предложенный всё же чуть эффективнее, и остаточная сумма квадратов уменьшилась на величину, превышающую возможную сумму погрешностей округлений), применимость метода можно считать обоснованной.

2. К достоинству нового метода можно отнести то, что он не так требователен к объёму исходных данных. При использовании в известном методе фильтрации методом скользящей средней информация о циклической компоненте требует как минимум одного дополнительного цикла, информации о котором может вообще не быть.

Предложенный метод может обойтись одним циклом предшествующей статистики, что и было использовано для прогнозирования деятельности малого предприятия, имеющего предысторию в один год. Других методов с таким сроком просто не существовало.

3. Оба метода предполагают неизменность циклической и трендовой компонент во времени, что не всегда соответствует наблюдаемым данным.

Для устранения последнего недостатка можно применить предлагаемый ниже метод фазовых трендов.

Метод фазовых трендов (МФТ) [6; 7]

Метод не предполагает выделение в явном виде циклической компоненты, но может быть применён лишь после её обнаружения. Факт наличия может быть обнаружен в траектории ВР, в периодограмме или спектрограмме анализа Фурье, а также в автокорреляционной функции наблюдаемого ВР.

Предполагается, что ВР содержит несколько (k_c) циклов (периодов), а в каждом цикле имеется несколько (k_s) равноотстоящих наблюдений, как это и было в предыдущем примере ($k_c = 4; k_s = 4$). Каждое наблюдение внутри цикла имеет номер $t = \overline{1, k_s}$. Каждое значение t отражает фазу внутри цикла. Одинаковые фазы разных циклов предполагаются коррелированными, причём для каждой фазы может быть найдено своё уравнение регрессии.

Т.к. предлагаемый метод прост, его реализацию рассмотрим сразу на числовом примере (том же). Пока выделим из данных табл.4 первую фазу (табл. 6).

Таблица 6

Данные по первой фазе числового примера

i	1	5	9	13
y_i	6	7,2	8	9

Если принять тренд линейным $T_1 = a_1 + b_1 i$, то МНК (и МВТ тоже) даст уравнение $T_1 = 5,835 + 0,245i$, где i – дискретное время в сквозной нумерации (не внутри цикла). Остаточная

сумма квадратов S_p для этого уравнения составила 0,028.

Эти результаты и результаты для других фазовых трендов сведены в табл. 7.

Таблица 7

Результаты расчета по методу МФТ

$T_1 = 5,835 + 0,245i$			$T_2 = 3,870 + 0,185i$			$T_3 = 4,660 + 0,160i$			$T_4 = 8,600 + 0,160i$		
$S_p = 0,028$			$S_p = 0,092$			$S_p = 0,072$			$S_p = 0,432$		
i	y_i	T_i	i	y_i	T_i	i	y_i	T_i	i	y_i	T_i
1	6	6,08	2	4,4	4,24	3	5,0	5,14	4	9,0	9,24
5	7,2	7,06	6	4,8	4,98	7	6,0	5,78	8	10,0	9,88
9	8	8,04	10	5,6	5,72	11	6,4	6,42	12	11,0	10,52
13	9	9,02	14	6,6	6,46	15	7,0	7,06	16	10,8	11,16

Можно заметить, что остаточная сумма квадратов для модели в целом $S_\Sigma = 0,028 + 0,092 + 0,072 + 0,432 = 0,624$ заметно меньше, чем для случая общего тренда и постоянной сезонной компоненты (1,099 из табл. 5), улучшатся также и другие характеристики надёжности – наблюдаемые значения коэффициентов детерминации, Тейла, критерия Фишера, средней ошибки аппроксимации.

Естественно, что МФТ можно использовать не только с линейными трендами.

Например, если все тренды выбрать в виде показательных функций, то остаточная сумма квадратов для этого примера несколько уменьшится. В принципе для трендов разных фаз можно использовать разные функции, например, для первой фазы линейную, для второй – показательную и т. д.

Примеры реализации МФТ на реальных данных

Можно, конечно, привести множество специально подобранных примеров, где лучшим будет предлагаемый разработчиками метод.

Поэтому для сравнения мы решили использовать результаты уже опубликованных методов (и с теми же данными), которые разработчики считают лучшими. Кроме того, мы рассматривали реальные данные в тех же временных интервалах, что и у

разработчиков для того, чтобы можно было сравнить результаты по характеристикам, выбранным разработчиками.

В табл. 8 приведен пример индекса реального ВВП по данным Росстата.

Таблица 8

Данные индекса реального ВВП

Год	2003				2004				2005			
Квартал	1	2	3	4	1	2	3	4	1	2	3	4
Индекс	100	106,9	121,4	120,7	107,0	115,2	130,2	128,7	112,8	122,3	138,3	138,4
МФТ	98	105,2	119,1	122,4	106,7	114,8	129,2	131,0	115,2	124,3	139,4	139,6
Год	2006				2007				2008			
Квартал	1	2	3	4	1	2	3	4	1	2	3	4
Индекс	120,2	131,7	138,3	138,4	129,3	142,2	160,5	163,6	140,5	153	170,2	165,5
МФТ	123,7	133,9	149,6	148,1	132,2	143,5	159,7	156,7	140,6	153	169,9	165,2

На рис. 2 приведены траектории временного ряда по индексу реального ВВП РФ на основе данных, представленных в табл. 8.

Построенная по этим данным ARMA-модель [9; 10] с использованием базиса Гребнера [3; 12] предполагала использование полиномиального уравнения 20-го порядка и 80-разрядных коэффициентов (чтобы не потерять точность). К сожалению, авторы не привели поквартальные данные по этой модели. Отмечено, что коэффициент детерминации составил

$R^2 = 0,99$, а оценка

$$MAPE = \frac{1}{l} \sum_{k=n+1}^{n+l} \left| \frac{(y_k - y_k^*)}{y_k} \right| = 2,74\%, \quad \text{что}$$

подтверждает высокую точность моделирования и прогнозирования. Здесь $n=12$ – количество кварталов, по которым строилась модель (обучающая выборка), $l=8$ – количество кварталов, по которым сравнивались реальные и прогнозные данные (контролирующая выборка).

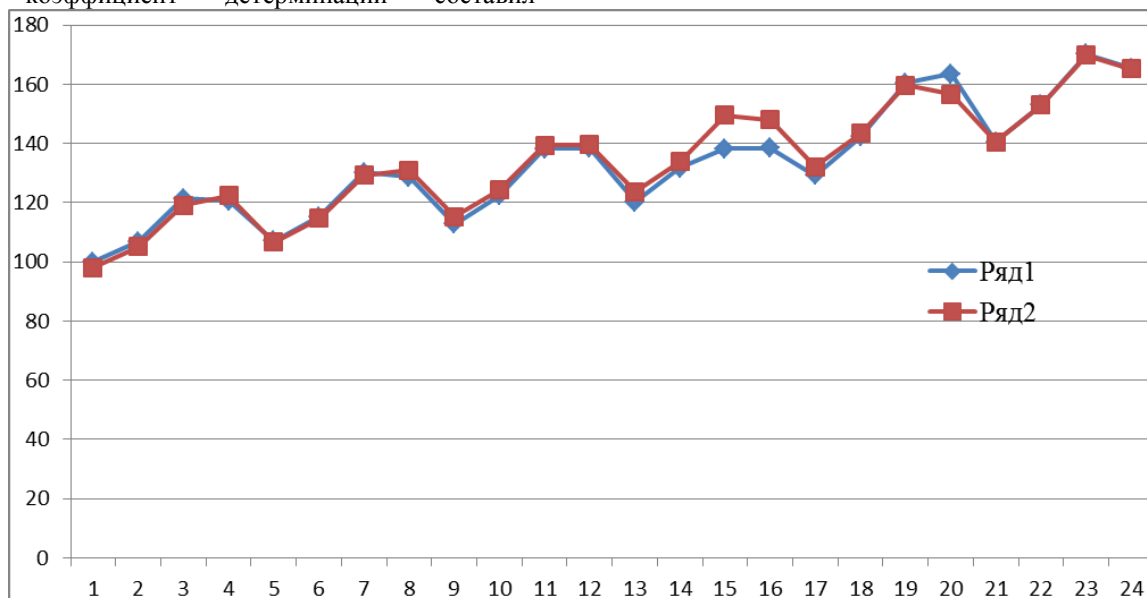


Рис. 2. Траектории временных рядов индекса ВВП РФ: ряд 1 – исходные данные [11], ряд 2 – расчет по МФТ

По методу фазовых трендов получено $R^2 = 0,9999$; $MAPE = 1,93\%$. Эти характеристики не хуже, чем у авторов в [9]. При этом фазовые тренды были выбраны линейными, и после расчётов приняли вид

$$\begin{aligned} T_1 &= 96,087 + 2,12129t; \quad T_2 = 100,397 + \\ &+ 2,392929t; \quad T_3 = 111,455 + 2,5405t; \\ T_4 &= 113,887 + 2,139357t. \end{aligned}$$

Здесь номер фазы совпадает с номером квартала.

Для иллюстрации применения нелинейных трендов мы взяли (у тех же авторов) поквартальную информацию об объёме инвестиций в основной капитал Самарской области за 9 лет (табл. 9).

Инвестиции в основной капитал Самарской области

Год	2000				2001				2002			
Квартал	1	2	3	4	1	2	3	4	1	2	3	4
Инвестиции	5768	10086	24855	22871	5898	4314	21407	24336	7020	4013	22369	26479
Модель Авторов[8]	3856	13560	28723	18601	1549	5044	25139	23720	5084	3557	22539	27422
МФТ	6439	7142	23202	25404	5837	6482	21569	22353	6672	6899	23532	24277
Год	2003				2004				2005			
Квартал	1	2	3	4	1	2	3	4	1	2	3	4
Инвестиции	9486	7625	27267	31512	13622	12408	36341	42283	18468	16188	51061	59461
Модель Авторов [8]	10574	8223	24735	31290	15033	16575	34104	38752	18027	23671	50405	53516
МФТ	8944	8395	29089	31182	12652	10970	38240	43065	17796	14622	50986	59927
Год	2006				2007				2008			
Квартал	1	2	3	4	1	2	3	4	1	2	3	4
Инвестиции	22807	20108	71071	80768	29570	25989	91039	106618	44885	30340	106568	142614
Модель Авторов [8]	22532	25313	70132	78101	33658	20410	86871	112662	56869	12626	93118	154037
МФТ	24378	19352	67327	81767	32396	25161	87262	108586	37110	35372	119013	136041

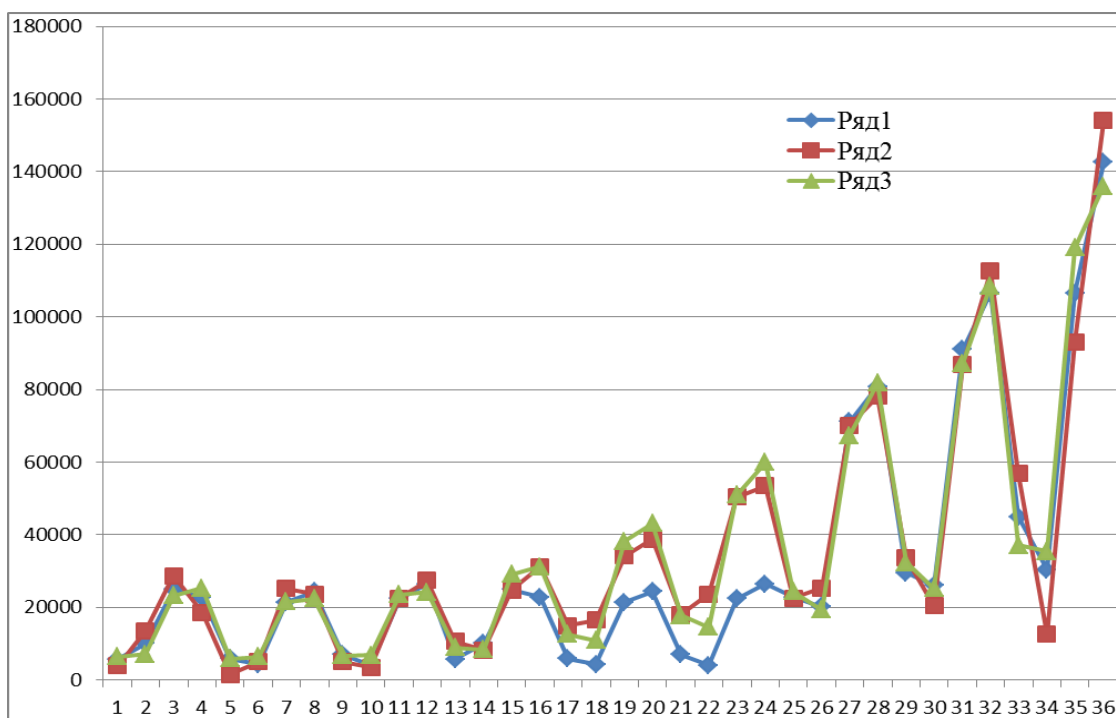


Рис. 3. Траектории временных рядов: ряд 1 – данные наблюдений [9], ряд 2 – расчет по модели [9], ряд 3 – расчет по МФТ

На рис. 3 приведены графики временных рядов по данным, представленными в табл. 9.

Здесь в авторском варианте [9] использовалась мультипликативно-аддитивная модель, в которой тренд был представлен полиномом второго порядка, циклическая компонента в мультипликативной части содержала 2 гармоники, а аддитивная часть также содержала гармонику:

$$y_k = (D_1(k\Delta)^2 + D_2k\Delta + D_3) \times (1 + A_1 \sin(\omega k\Delta) + 6B_1 \cos(\omega k\Delta) + A_2 \sin(2\omega k\Delta) + B_2 \cos(2\omega k\Delta)) + E_1 \sin(\omega k\Delta) + F_1 \cos(\omega k\Delta).$$

Качество модели и прогнозов характеризуют следующие показатели:

$$R^2 = 0,9828, \quad MAPE = 26\%.$$

По виду траектории для МФТ мы выбрали тренды в виде парабол 2-го порядка:
 $T = a + bx + cx^2$.

В результате получили:
 $T_1 = 6814,1 - 419,80t + 44,894t^2$;
 $T_2 = 7876,3 - 434,60t + 33,692t^2$;
 $T_3 = 26785,0 - 1531,41t + 112,332t^2$; $T_4 = 33434,7 - 2630,06t + 155,579t^2$.

При этом

$$R^2 = 0,9987; \quad MAPE = 1,15\%.$$

В табл. 9 за 2008 г. модельные данные приведены по прогнозам моделей, построенных по

данным до 2008 г. Реальные данные за 2008 г. использованы как контролирующая выборка.

По графикам рис. 3 видим, что МФТ в этих примерах даёт результаты не хуже, чем существенно более сложные модели [9]. Отметим, что на рис. 3 используется сквозная нумерация кварталов за 2000–2008 гг.

Выводы

В работе предложены два простых метода построения моделей временных рядов – метод взвешенных тангенсов и метод фазовых трендов. К основным достоинствам первого метода можно отнести то, что он позволяет построить модель при меньшем объёме статистической информации, чем для существующих методов. Основными достоинствами второго метода являются простота, возможность учёта эволюции циклической компоненты временного ряда, хорошие характеристики надёжности.

Список литературы

1. Айвазян С.А. Прикладная статистика. Основы эконометрики. М.: ЮНИТИ-ДАНА, 2001. 432 с.
2. Буторина О.В., Баталова Е.В., Фукалова Ю.С. Исследование циклических процессов экономики России с учётом межциклической рекуррентности // Вестник Пермского университета. Сер. Экономика. 2012. Вып. 4(15). С. 64–74.
3. Бухбергер Б. Базисы Гребнера. Алгоритмический метод в теории полиномиальных идеалов. Компьютерная алгебра. Символьные и алгебраические вычисления / под ред. Б. Бухбергера, Д. Коллинза, Р. Лоос. М.: Мир, 1986. С. 331–372.
4. Елисеева И.И., Курышева С.В., Костеева Т.В. и др. Эконометрика: учебник / под ред. И.И. Елисеевой. М.: Финансы и статистика, 2007. 576 с.
5. Кошкин Ю.Л., Сошникова Е.М. Метод взвешенных тангенсов // Общество, наука, инновации (НТК-2013): ежегод. открыт. всерос. науч.-техн. конф., 15–26 апреля г. Киров: Вят. гос. ун-т, 2013. Секция «Математические методы в экономике». (Электрон. опт. диск).
6. Кошкин Ю.Л., Сошникова Е.М., Тарлава М.Ю. Метод фазовых трендов // Общество, наука, инновации (НТК-2013): ежегод. открыт. всерос. науч.-техн. конф., 15–26 апреля г. Киров: Вят. гос. ун-т, 2013. Секция «Математические методы в экономике» (Электрон. опт. диск).
7. Кошкин Ю.Л., Тарлава М.Ю. Два метода анализа временных рядов // Математика, Компьютер, Образование. Тезисы доклада международной конференции, 3–7 февраля 2014 г. Дубна. Москва-Ижевск: Регулярная и хаотическая динамика, 2014. С. 210.
8. Мхитарян В.С., Архипова М.Ю., Балаш В.А. и др. Эконометрика / под ред. В.С. Мхитаряна. М.: Проспект, 2009. 384 с.
9. Семёнычев В.К., Семёнычев Е.В. Параметрическая идентификация рядов динамики: структуры, модели, эволюция: монография. Самара: СамНЦ РАН, 2011. 364 с.
10. Семёнычев В.К., Семёнычев Е.В., Коробецкая А.А. Метод параметрической итерационной декомпозиции тренд-сезонных рядов аддитивной структуры // Вестник Самарского муниципального института управления. 2010. №1(12). С. 63–72.
11. Индекс реального ВВП РФ. URL: <http://www.gks.ru> (дата обращения: 20.01.2014).
12. Adams W.W., Laustunau Ph. An introduction to Groebner Bases: Amer. Math. Soc. 1994. (Grad. Stud. in Math., Vol.3).

Получено: 02.02.2015

References

1. Aivazyan S.A. *Priladnaya statistica. Osnovy ekonometriki*. Moscow, UNITI-DANA., 2001. 432 p.
2. Butorina O.V., Batalova E.V., Fukalova Yu.S. Issledovanie tsiklicheskih protsessov ekonomiki Rossii s uchyotom mezhtsiklicheskoj rekurrentnosti // *Vestnik Permskogo universiteta, Ser. Ekonomika*. 2012. Vol. 4(15). pp. 64–74
3. Buhberger B. *Bazisy Grebnera. Algoritmicheskiy metod v teorii polinomialnyh idealov. Compyuternaya algebra. Simvolnye i algebbraicheskie vychisleniya*. Moscow, MIR., 1986, pp 331–372.
4. Eliseeva I.I., Kuryshva S.V., Kosteeva T.V. a.o. *Ekonometrika*. Moscow, Finansy i statistika, 2007. 576 p.
5. Koshkin Y.L., Soshnikova E.M. Metod vzveshennyh tangensov. Obschestvo, nauka, innovatsii (NTK-2013), *ezhegod. otkryt. vseros. nauch.-tehnich. konf., 15-26 apr., Kirov, Vyat. gos. un-t, 2013, sektsiya "Matematicheskie metody v ekonomike"*. (electron. opt. disk). 2013.
6. Koshkin Y.L., Soshnikova E.M., Tarlavina M. Y. Metod fazovyh trendov. Obschestvo, nauka, innovatsii (NTK-2013), *ezhegod. otkryt. vseros. nauch.-tehnich. konf., 15-26 apr., Kirov, Vyat. gos. un-t, 2013, sektsiya "Matematicheskie metody v ekonomike"*. (electron. opt. disk). 2013.
7. Koshkin Y.L., Tarlavina M.Y. Dva metoda analiza vremennyh ryadov. *Matematika, Compyuter, Obrazovanie. Tezisy doclada mezhdunarodnoy konferentsii., 3-7 fevr. 2014. Dubna. Moscow-Izhevsk, Reguljarnaya i haoticheskaya dinamika*. 2014, p.210.
8. Mhitaryan V.S., Arhipova M.Y. Balash V.A. a.o. *Ekonometrika*. Moscow, Prospekt, 2009. 384 p.
9. Semyonychev V.K., Semyonychev E.V. *Parametricheskaya identifikatsiya ryadov dinamiki: struktury, modeli, evolyutsiya: monografiya*. Samara, SamNC RAN, 2011. 364 p.
10. Semyonychev V.K., Semyonychev E.V., Korobeckaya A.A. Metod parametricheskoy i teratsionnoy dekompozitsii trend-sezonnyh ryadov additivnoy struktury. *Vestnik Samarskogo*

munitsipalnogo institute upravleniya. Samara, Samarskij munitsipalnyj institute upravleniya. 2010, №1(12), pp. 63–72.

11. *Index real'nogo VVP* [Electronic resource] //gks.ru [Official site]. URL: <http://www.gks.ru> (accessed 20.01.2014).

12. Adams W.W., Laustunau Ph. An introduction to Groebner Bases: Amer. Math. Soc. 1994. (Grad. Stud. in Math., Vol.3).

The date of the manuscript receipt:
02.02.2015

ON THE QUESTION OF MODELLING TIME SERIES TRENDS

Yury L. Koshkin, Candidate of Technical Sciences, Associate Professor

E-mail: koshyuri@gmail.com

Anatoly V. Shatrov, Doctor of Mathematics and Physics, Professor, Head of the Department of Mathematical Modelling in Economics

E-mail: avshatrov1@yandex.ru

Vyatka State University; 36, Moskovskaya st., Kirov, 610000, Russia

The paper presents the analysis of modern methods of modelling time series trends. In economy and other spheres of scientific and practical activity we see objects of our interest developing over time. In order to model these objects, econometric methods and presentation of initial and resulting data in the form of time series are usually used. At present there is a great number of methods of time series modelling. Many of the methods, developed for solving specific problems, are not universal. Very often researchers use dynamic decomposition of time series into several components. Commonly a trend component, a cyclical component and a random component are singled out. In this paper the first method (method of weighted tangents - MWT) involves decomposition into the trend and cyclical components. The second method does not involve decomposition containing a cyclical component. Instead, the method of phase trends (MPT) uses the concept of "phases", which can be found in the initial form of time series. Application of the phase trends method allows for performing a piecewise approximation of time series. Modern methods are not aimed at work with short time series as some part of statistical data is lost in the preliminary smoothing. The MWT can be applied for short time series in case there is at least one cycle. Many methods do not consider development of time series over time (evolution). That is why the authors suggest using the MPT, which in many cases gives results that are not inferior in quality compared to complicated modern methods.

Keywords: time series, trend, cyclical component, additive model, prediction.

Пробьба ссылаться на эту статью в русскоязычных источниках следующим образом:

Кошкин Ю.Л., Шатров А.В. К вопросу о моделировании трендов временных рядов // Вестник Пермского университета. Сер. «Экономика» = Perm University Herald. Economy. 2015. № 3(26). С. 32–41.

Please cite this article in English as:

Koshkin Yu.L., Shatrov A.V. On the question of modelling time series trends // Vestnik Permskogo universiteta. Seria Ekonomika = Perm University Herald. Economy. 2015. № 3(26). P. 32–41.