

4. Временные ряды

При построении эконометрической модели используются два типа данных:

- 1) данные, характеризующие совокупность различных объектов в определенный момент времени;
- 2) данные, характеризующие один объект за ряд последовательных моментов времени.

Модели, построенные по данным первого типа, называются *пространственными моделями*. Модели, построенные на основе второго типа данных, называются *моделями временных рядов*.

Временной ряд (ряд динамики) – это совокупность значений какого-либо показателя за несколько последовательных моментов или периодов времени. Каждый уровень временного ряда формируется под воздействием большого числа факторов, которые условно можно подразделить на три группы:

- 1) факторы, формирующие тенденцию ряда;
- 2) факторы, формирующие циклические колебания ряда;
- 3) случайные факторы.

Рассмотрим воздействие каждого фактора на временной ряд в отдельности.

Большинство временных рядов экономических показателей имеют тенденцию, характеризующую совокупное долговременное воздействие множества факторов на динамику изучаемого показателя. Все эти факторы, взятые в отдельности, могут оказывать разнонаправленное воздействие на исследуемый показатель. Однако в совокупности они формируют его возрастающую или убывающую тенденцию. На рис. 4.1 показан гипотетический временной ряд, содержащий возрастающую тенденцию.

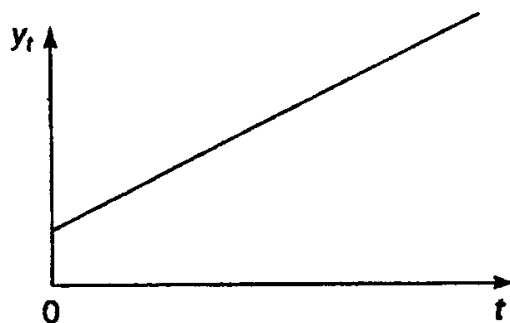


Рис. 4.1. Ряд, содержащий только тенденцию.

Также изучаемый показатель может быть подвержен циклическим колебаниям. Эти колебания могут носить сезонный характер, поскольку экономическая деятельность ряда отраслей экономики зависит от времени года (например, цены на сельскохозяйственную продукцию в летний период выше, чем в зимний; уровень безработицы в курортных городах в зимний период выше по сравнению с летним). При наличии больших массивов данных за длительные промежутки времени можно выявить циклические колебания, связанные с общей динамикой конъюнктуры рынка. На рис. 4.2 представлен гипотетический временной ряд, содержащий только сезонную компоненту.



Рис. 4.2. Ряд, содержащий только сезонную компоненту.

Некоторые временные ряды не содержат тенденции и циклической компоненты, а каждый следующий их уровень образуется как сумма среднего уровня ряда и некоторой (положительной или отрицательной) случайной компоненты. Пример ряда, содержащего только случайную компоненту, приведен на рис. 4.3.

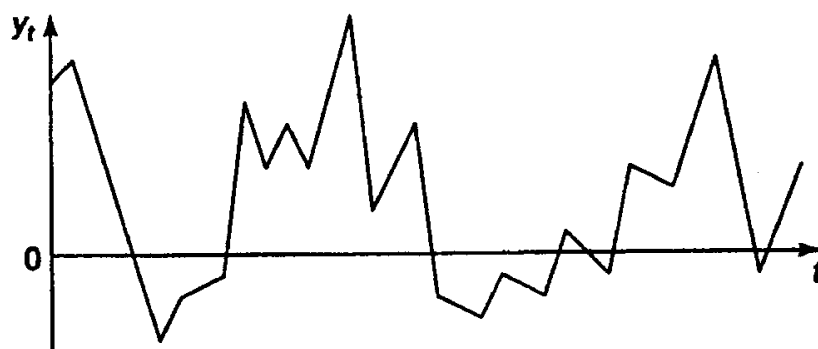


Рис. 4.3. Ряд, содержащий только случайную компоненту.

Очевидно, что реальные данные не следуют целиком и полностью из каких-либо описанных выше моделей. Чаще всего они содержат все три компоненты. Каждый их уровень формируется под воздействием тенденции, сезонных колебаний и случайной компоненты.

В большинстве случаев фактический уровень временного ряда можно представить как сумму или произведение трендовой, циклической и случайной компонент. Модель, в которой временной ряд представлен как сумма перечисленных компонент, называется *аддитивной моделью* временного ряда. Модель, в которой временной ряд представлен как произведение перечисленных компонент, называется *мультипликативной моделью* временного ряда. Основная задача эконометрического исследования отдельного временного ряда – выявление и придание количественного выражения каждой из перечисленных выше компонент с тем, чтобы использовать полученную информацию для прогнозирования будущих значений ряда или при построении моделей взаимосвязи двух или более временных рядов.

4.1. Автокорреляция уровней временного ряда

При наличии во временном ряде тенденции и циклических колебаний значения каждого последующего уровня ряда зависят от предыдущих. Корреляционную зависимость между последовательными уровнями временного ряда называют *автокорреляцией* уровней ряда.

Количественно ее можно измерить с помощью линейного коэффициента корреляции между уровнями исходного временного ряда и уровнями этого ряда, сдвинутыми на несколько шагов во времени.

Формула для расчета коэффициента автокорреляции имеет вид:

$$r_1 = \frac{\sum_{t=2}^n (y_t - \bar{y}_1)(y_{t-1} - \bar{y}_2)}{\sqrt{\sum_{t=2}^n (y_t - \bar{y}_1)^2 \sum_{t=2}^n (y_{t-1} - \bar{y}_2)^2}}, \quad (4.1)$$

где

$$\bar{y}_1 = \frac{1}{n-1} \sum_{t=2}^n y_t, \quad \bar{y}_2 = \frac{1}{n-1} \sum_{t=2}^n y_{t-1}.$$

Эту величину называют *коэффициентом автокорреляции уровней ряда первого порядка*, так как он измеряет зависимость между соседними уровнями ряда y_t и y_{t-1} .

Аналогично можно определить коэффициенты автокорреляции второго и более высоких порядков. Так, коэффициент автокорреляции второго порядка характеризует тесноту связи между уровнями y_t и y_{t-2} и определяется по формуле:

$$r_2 = \frac{\sum_{t=3}^n (y_t - \bar{y}_3)(y_{t-2} - \bar{y}_4)}{\sqrt{\sum_{t=3}^n (y_t - \bar{y}_3)^2 \sum_{t=3}^n (y_{t-2} - \bar{y}_4)^2}}, \quad (4.2)$$

где

$$\bar{y}_3 = \frac{1}{n-2} \sum_{t=3}^n y_t, \quad \bar{y}_4 = \frac{1}{n-2} \sum_{t=3}^n y_{t-2}.$$

Число периодов, по которым рассчитывается коэффициент автокорреляции, называют *лагом*. С увеличением лага число пар значений, по которым рассчитывается коэффициент автокорреляции, уменьшается. Считается целесообразным для обеспечения

статистической достоверности коэффициентов автокорреляции использовать правило – максимальный лаг должен быть не больше $n/4$.

Свойства коэффициента автокорреляции.

1. Он строится по аналогии с линейным коэффициентом корреляции и таким образом характеризует тесноту только линейной связи текущего и предыдущего уровней ряда. Поэтому по коэффициенту автокорреляции можно судить о наличии линейной (или близкой к линейной) тенденции. Для некоторых временных рядов, имеющих сильную нелинейную тенденцию (например, параболу второго порядка или экспоненту), коэффициент автокорреляции уровней исходного ряда может приближаться к нулю.

2. По знаку коэффициента автокорреляции нельзя делать вывод о возрастающей или убывающей тенденции в уровнях ряда. Большинство временных рядов экономических данных содержат положительную автокорреляцию уровней, однако при этом могут иметь убывающую тенденцию.

Последовательность коэффициентов автокорреляции уровней первого, второго и т.д. порядков называют *автокорреляционной функцией* временного ряда. График зависимости ее значений от величины лага (порядка коэффициента автокорреляции) называется *коррелограммой*.

Анализ автокорреляционной функции и коррелограммы позволяет определить лаг, при котором автокорреляция наиболее высокая, а следовательно, и лаг, при котором связь между текущим и предыдущими уровнями ряда наиболее тесная, т.е. при помощи анализа автокорреляционной функции и коррелограммы можно выявить структуру ряда.

Если наиболее высоким оказался коэффициент автокорреляции первого порядка, исследуемый ряд содержит только тенденцию. Если наиболее высоким оказался коэффициент автокорреляции порядка k , то

ряд содержит циклические колебания с периодичностью в k моментов времени. Если ни один из коэффициентов автокорреляции не является значимым, можно сделать одно из двух предположений относительно структуры этого ряда: либо ряд не содержит тенденции и циклических колебаний, либо ряд содержит сильную нелинейную тенденцию, для выявления которой нужно провести дополнительный анализ. Поэтому коэффициент автокорреляции уровней и автокорреляционную функцию целесообразно использовать для выявления во временном ряде наличия или отсутствия трендовой компоненты и циклической (сезонной) компоненты.

Рассмотрим **пример**. Пусть имеются некоторые условные данные об общем количестве правонарушений на таможне одного из субъектов РФ (например, Республики Татарстан).

Таблица 4.1

Год	Квартал	t	Количество возбужденных дел, y_t
1999	I	1	375
	II	2	371
	III	3	869
	IV	4	1015
2000	I	5	357
	II	6	471
	III	7	992
	IV	8	1020
2001	I	9	390
	II	10	355
	III	11	992
	IV	12	905
2002	I	13	461
	II	14	454
	III	15	920
	IV	16	927

Построим поле корреляции:

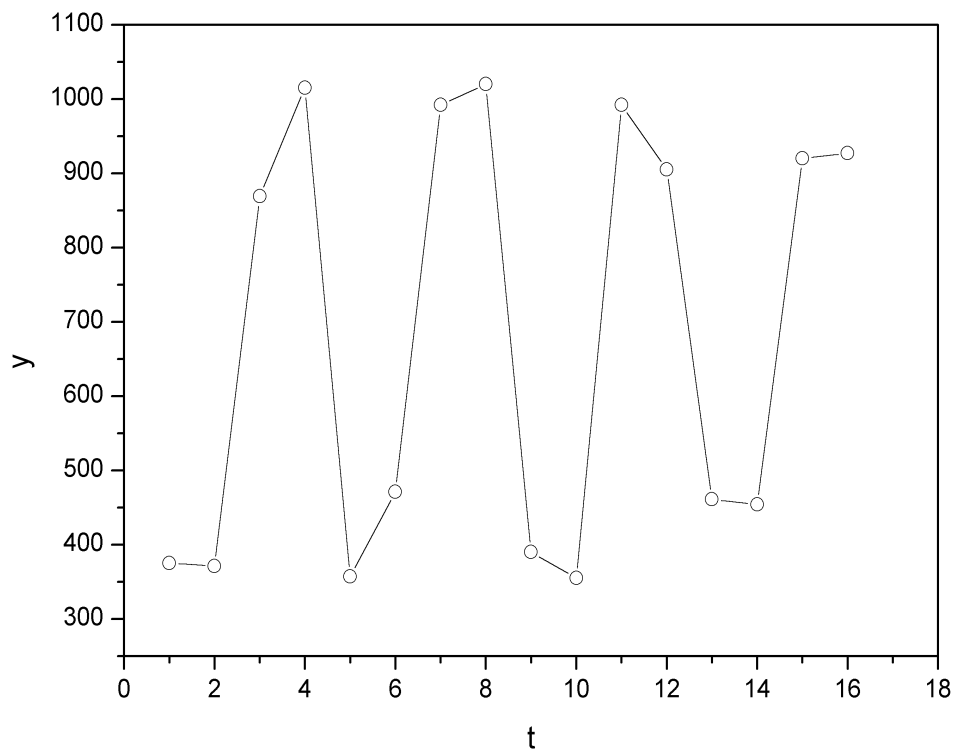


Рис. 4.4.

Уже исходя из графика видно, что значения y образуют пилообразную фигуру. Рассчитаем несколько последовательных коэффициентов автокорреляции. Для этого составляем первую вспомогательную таблицу.

Таблица 4.2

t	y_t	y_{t-1}	$y_t - \bar{y}_1$	$y_{t-1} - \bar{y}_2$	$(y_t - \bar{y}_1) \times (y_{t-1} - \bar{y}_2)$	$(y_t - \bar{y}_1)^2$	$(y_{t-1} - \bar{y}_2)^2$
1	2	3	4	5	6	7	8
1	375	—	—	—	—	—	—
2	371	375	-328,93	-288,13	94774,60	108194,94	83018,90
3	869	371	169,07	-292,13	-49390,42	28584,66	85339,94
4	1015	869	315,07	205,87	64863,46	99269,10	42382,46
5	357	1015	-342,93	351,87	-120666,78	117600,98	123812,50
6	471	357	-228,93	-306,13	70082,34	52408,94	93715,58
7	992	471	292,07	-192,13	-56115,41	85304,88	36913,94
8	1020	992	320,07	328,87	105261,42	102444,80	108155,48
9	390	1020	-309,93	356,87	-110604,72	96056,60	127356,20

1	2	3	4	5	6	7	8
10	355	390	-344,93	-273,13	94210,73	118976,70	74600,00
11	992	355	292,07	-308,13	-89995,53	85304,88	94944,10
12	905	992	205,07	328,87	67441,37	42053,70	108155,48
13	461	905	-238,93	241,87	-57790,00	57087,54	58501,10
14	454	461	-245,93	-202,13	49709,83	60481,56	40856,54
15	920	454	220,07	-209,13	-46023,24	48430,80	43735,36
16	927	920	227,07	256,87	58327,47	51560,78	65982,20
Сумма	10499	9947	0,05	0,05	74085,13	1153760,93	1187469,73
Среднее значение	699,93	663,13	–	–	–	–	–

Следует заметить, что среднее значение получается путем деления не на 16, а на 15, т.к. у нас теперь на одно наблюдение меньше.

Теперь вычисляем коэффициент автокорреляции первого порядка по формуле (4.1):

$$r_1 = \frac{74085,13}{\sqrt{1153760,39 \cdot 1187469,73}} = 0,063294.$$

Составляем вспомогательную таблицу для расчета коэффициента автокорреляции второго порядка.

Таблица 4.3

t	y_t	y_{t-2}	$y_t - \bar{y}_3$	$y_{t-2} - \bar{y}_4$	$(y_t - \bar{y}_3) \times (y_{t-2} - \bar{y}_4)$	$(y_t - \bar{y}_3)^2$	$(y_{t-2} - \bar{y}_4)^2$
1	2	3	4	5	6	7	8
1	375	–	–	–	–	–	–
2	371	–	–	–	–	–	–
3	869	375	145,57	-269,79	-39273,33	21190,62	72786,64
4	1015	371	291,57	-273,79	-79828,95	85013,06	74960,96
5	357	869	-366,43	224,21	-82157,27	134270,94	50270,12
6	471	1015	-252,43	370,21	-93452,11	63720,90	137055,44
7	992	357	268,57	-287,79	-77291,76	72129,84	82823,08
8	1020	471	296,57	-173,79	-51540,90	87953,76	30202,96
9	390	992	-333,43	347,21	-115770,23	111175,56	120554,78
10	355	1020	-368,43	375,21	-138238,62	135740,66	140782,54
11	992	390	268,57	-254,79	-68428,95	72129,84	64917,94
12	905	355	181,57	-289,79	-52617,17	32967,66	83978,24
13	461	992	-262,43	347,21	-91118,32	68869,50	120554,78
14	454	905	-269,43	260,21	-70108,38	72592,52	67709,24
15	920	461	196,57	-183,79	-36127,60	38639,76	33778,76
16	927	454	203,57	-190,79	-38839,12	41440,74	36400,82
Сумма	10128	9027	-0,02	-0,06	-1034792,71	1037835,43	1116776,36
Среднее значение	723,43	644,79	–	–	–	–	–

Следовательно

$$r_2 = \frac{-1034792,71}{\sqrt{1037835,43 \cdot 1116776,36}} = -0,961183.$$

Аналогично находим коэффициенты автокорреляции более высоких порядков, а все полученные значения заносим в сводную таблицу.

Таблица 4.4

Лаг	Коэффициент автокорреляции уровней
1	0,063294
2	-0,961183
3	-0,036290
4	0,964735
5	0,050594
6	-0,976516
7	-0,069444
8	0,964629
9	0,162064
10	-0,972918
11	-0,065323
12	0,985761

Коррелограмма:

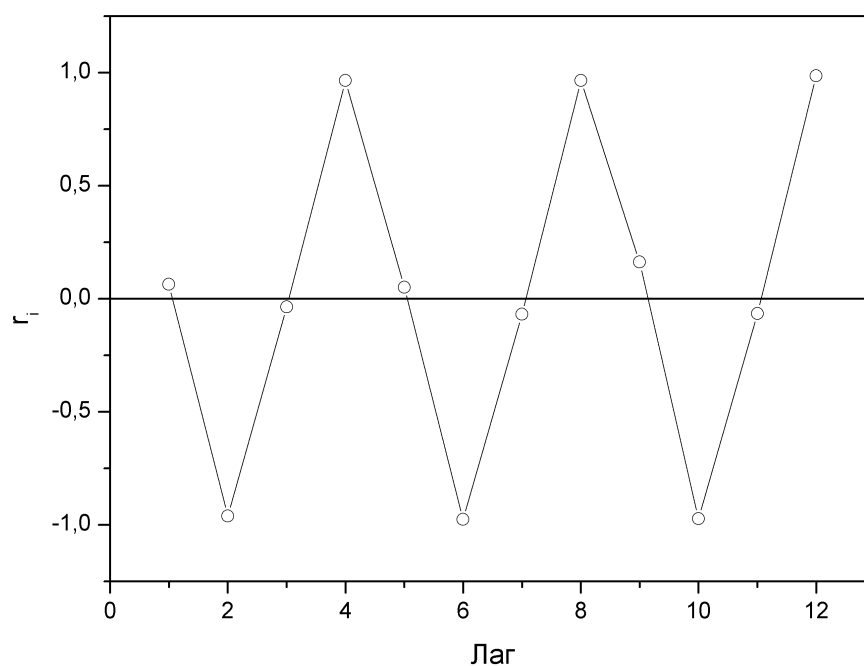


Рис. 4.5.

Анализ коррелограммы и графика исходных уровней временного ряда позволяет сделать вывод о наличии в изучаемом временном ряде сезонных колебаний периодичностью в четыре квартала.

4. 2. Моделирование тенденции временного ряда

Распространенным способом моделирования тенденции временного ряда является построение аналитической функции, характеризующей зависимость уровней ряда от времени, или тренда. Этот способ называют *аналитическим выравниванием временного ряда*.

Поскольку зависимость от времени может принимать разные формы, для ее формализации можно использовать различные виды функций. Для построения трендов чаще всего применяются следующие функции:

линейный тренд: $\hat{y}_t = a + b \cdot t$;

гипербола: $\hat{y}_t = a + \frac{b}{t}$;

экспоненциальный тренд: $\hat{y}_t = e^{a+b \cdot t}$ (или $\hat{y}_t = a \cdot b^t$);

степенная функция: $\hat{y}_t = a \cdot t^b$;

полиномы различных степеней: $\hat{y}_t = a + b_1 \cdot t + b_2 \cdot t^2 + \dots + b_m \cdot t^m$.

Параметры каждого из перечисленных выше трендов можно определить обычным МНК, используя в качестве независимой переменной время $t = 1, 2, \dots, n$, а в качестве зависимой переменной – фактические уровни временного ряда \hat{y}_t . Для нелинейных трендов предварительно проводят стандартную процедуру их линеаризации.

Существует несколько способов определения типа тенденции. К числу наиболее распространенных способов относятся качественный анализ изучаемого процесса, построение и визуальный анализ графика зависимости уровней ряда от времени. В этих же целях можно

использовать и коэффициенты автокорреляции уровней ряда. Тип тенденции можно определить путем сравнения коэффициентов автокорреляции первого порядка, рассчитанных по исходным и преобразованным уровням ряда. Если временной ряд имеет линейную тенденцию, то его соседние уровни \hat{y}_t и \hat{y}_{t-1} тесно коррелируют. В этом случае коэффициент автокорреляции первого порядка уровней исходного ряда должен быть высоким. Если временной ряд содержит нелинейную тенденцию, например, в форме экспоненты, то коэффициент автокорреляции первого порядка по логарифмам уровней исходного ряда будет выше, чем соответствующий коэффициент, рассчитанный по уровням ряда. Чем сильнее выражена нелинейная тенденция в изучаемом временном ряде, тем в большей степени будут различаться значения указанных коэффициентов.

Выбор наилучшего уравнения в случае, когда ряд содержит нелинейную тенденцию, можно осуществить путем перебора основных форм тренда, расчета по каждому уравнению скорректированного коэффициента детерминации и средней ошибки аппроксимации. Этот метод легко реализуется при компьютерной обработке данных.

4.3. Моделирование сезонных колебаний

Простейший подход к моделированию сезонных колебаний – это расчет значений сезонной компоненты методом скользящей средней и построение аддитивной или мультипликативной модели временного ряда.

Общий вид аддитивной модели следующий:

$$Y = T + S + E. \quad (4.3)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как сумма трендовой (T), сезонной (S) и случайной (E) компонент.

Общий вид мультипликативной модели выглядит так:

$$Y = T \cdot S \cdot E. \quad (4.4)$$

Эта модель предполагает, что каждый уровень временного ряда может быть представлен как произведение трендовой (T), сезонной (S) и случайной (E) компонент.

Выбор одной из двух моделей осуществляется на основе анализа структуры сезонных колебаний. Если амплитуда колебаний приблизительно постоянна, строят аддитивную модель временного ряда, в которой значения сезонной компоненты предполагаются постоянными для различных циклов. Если амплитуда сезонных колебаний возрастает или уменьшается, строят мультипликативную модель временного ряда, которая ставит уровни ряда в зависимость от значений сезонной компоненты.

Построение аддитивной и мультипликативной моделей сводится к расчету значений T , S и E для каждого уровня временного ряда.

Процесс построения модели включает в себя следующие шаги.

- 1) Выравнивание исходного ряда методом скользящей средней.
- 2) Расчет значений сезонной компоненты S .
- 3) Устранение сезонной компоненты из исходных уровней ряда и получение выровненных данных $(T + E)$ в аддитивной или $(T \cdot E)$ в мультипликативной модели.
- 4) Аналитическое выравнивание уровней $(T + E)$ или $(T \cdot E)$ и расчет значений T с использованием полученного уравнения тренда.
- 5) Расчет полученных по модели значений $(T + E)$ или $(T \cdot E)$.
- 6) Прогноз будущих значений уровней временного ряда на основе построенной модели.

Методику построения каждой из моделей рассмотрим на примерах.

Пример. Построение аддитивной модели временного ряда. Обратимся к данным об объеме правонарушений на таможне за четыре года, представленным в табл. 4.1.

Было показано, что данный временной ряд содержит сезонные колебания периодичностью 4, т.к. количество правонарушений в первый-второй кварталы ниже, чем в третий-четвертый. Рассчитаем компоненты аддитивной модели временного ряда.

Шаг 1. Проведем выравнивание исходных уровней ряда методом скользящей средней. Для этого:

1.1. Просуммируем уровни ряда последовательно за каждые четыре квартала со сдвигом на один момент времени и определим условные годовые объемы потребления электроэнергии (гр. 3 табл. 4.5).

1.2. Разделив полученные суммы на 4, найдем скользящие средние (гр. 4 табл. 4.5). Полученные таким образом выровненные значения уже не содержат сезонной компоненты.

1.3. Приведем эти значения в соответствие с фактическими моментами времени, для чего найдем средние значения из двух последовательных скользящих средних – центрированные скользящие средние (гр. 5 табл. 4.5).

Таблица 4.5

№ квартала, t	Количество правонарушений, y_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	375	–	–	–	–
2	371	2630	657,5	–	–
3	869	2612	653	655,25	213,75
4	1015	2712	678	665,5	349,5
5	357	2835	708,75	693,375	-336,375
6	471	2840	710	709,375	-238,375
7	992	2873	718,25	714,125	277,875
8	1020	2757	689,25	703,75	316,25
9	390	2757	689,25	689,25	-299,25
10	355	2642	660,5	674,875	-319,875
11	992	2713	678,25	669,375	322,625
12	905	2812	703	690,625	214,375
13	461	2740	685	694	-233
14	454	2762	690,5	687,75	-233,75
15	920	–	–	–	–
16	927	–	–	–	–

Шаг 2. Найдем оценки сезонной компоненты как разность между фактическими уровнями ряда и центрированными скользящими средними (гр. 6 табл. 4.5). Используем эти оценки для расчета значений сезонной компоненты S_i (табл. 4.6). Для этого найдем средние за каждый квартал (по всем годам) оценки сезонной компоненты \bar{S}_i .

Таблица 4.6

Показатели	Год	№ квартала, i			
		I	II	III	IV
	1999	–	–	213,75	349,5
	2000	–336,375	–238,375	277,875	316,25
	2001	–299,25	–319,875	322,625	214,375
	2002	–233	–233,75	–	–
Всего за i -й квартал		–868,625	–792	814,25	880,125
Средняя оценка сезонной компоненты для i -го квартала, \bar{S}_i		–289,542	–264	271,417	293,375
Скорректированная сезонная компонента, S_i		–292,355	–266,813	268,604	290,562

В моделях с сезонной компонентой обычно предполагается, что сезонные воздействия за период взаимопогашаются. В аддитивной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна нулю.

Для данной модели имеем:

$$-289,542 - 264 + 271,417 + 293,375 = 11,25.$$

Корректирующий коэффициент: $k = 11,25/4 = 2,813$.

Рассчитываем скорректированные значения сезонной компоненты ($S_i = \bar{S}_i - k$) и заносим полученные данные в таблицу 4.6.

Проверим равенство нулю суммы значений сезонной компоненты:

$$-292,355 - 266,813 + 268,604 + 290,562 = 0,00.$$

Шаг 3. Исключим влияние сезонной компоненты, вычитая ее значение из каждого уровня исходного временного ряда. Получим

величины $T + E = Y - S$ (гр. 4 табл. 4.7). Эти значения рассчитываются за каждый момент времени и содержат только тенденцию и случайную компоненту.

Таблица 4.7

t	y_t	S_t	$y_t - S_t$	T	$T + S$	$E = y_t - (T + S)$	E^2
1	2	3	4	5	6	7	8
1	375	-292,355	667,355	672,684	380,329	-5,329	28,3982
2	371	-266,813	637,813	673,610	406,797	-35,797	1281,425
3	869	268,604	600,396	674,535	943,139	-74,139	5496,591
4	1015	290,562	724,438	675,461	966,023	48,977	2398,747
5	357	-292,355	649,355	676,386	384,031	-27,031	730,675
6	471	-266,813	737,813	677,312	410,499	60,501	3660,371
7	992	268,604	723,396	678,237	946,841	45,159	2039,335
8	1020	290,562	729,438	679,163	969,725	50,275	2527,576
9	390	-292,355	682,355	680,088	387,733	2,267	5,139289
10	355	-266,813	621,813	681,014	414,201	-59,201	3504,758
11	992	268,604	723,396	681,939	950,543	41,457	1718,683
12	905	290,562	614,438	682,865	973,427	-68,427	4682,254
13	461	-292,355	753,355	683,790	391,435	69,565	4839,289
14	454	-266,813	720,813	684,716	417,903	36,097	1302,993
15	920	268,604	651,396	685,641	954,245	-34,245	1172,720
16	927	290,562	636,438	686,567	977,129	-50,129	2512,917

Шаг 4. Определим компоненту T данной модели. Для этого проведем аналитическое выравнивание ряда $(T + E)$ с помощью линейного тренда. Результаты аналитического выравнивания следующие:

$$\hat{T} = 671,759 + 0,9255 \cdot t.$$

Подставляя в это уравнение значения $t = 1, 2, \dots, 16$, найдем уровни T для каждого момента времени (гр. 5 табл. 4.7).

Шаг 5. Найдем значения уровней ряда, полученные по аддитивной модели. Для этого прибавим к уровням T значения сезонной компоненты для соответствующих кварталов (гр. 6 табл. 4.7).

На одном графике отложим фактические значения уровней временного ряда и теоретические, полученные по аддитивной модели.

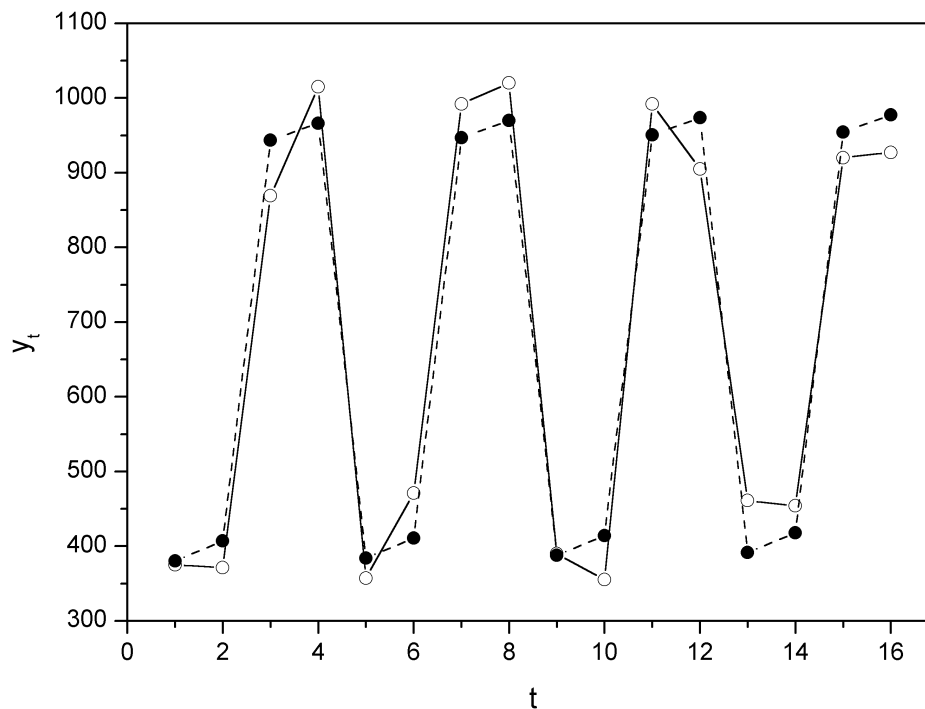


Рис. 4.6.

Для оценки качества построенной модели применим сумму квадратов полученных абсолютных ошибок.

$$R^2 = 1 - \frac{\sum E^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{37901,872}{1252743,75} = 0,970.$$

Следовательно, можно сказать, что аддитивная модель объясняет 97% общей вариации уровней временного ряда количества правонарушений по кварталам за 4 года.

Шаг 6. Прогнозирование по аддитивной модели. Предположим, что по нашему примеру необходимо дать прогноз об общем объеме правонарушений на I и II кварталы 2003 года. Прогнозное значение F_t уровня временного ряда в аддитивной модели есть сумма трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 671,777 + 0,9233 \cdot t.$$

Получим

$$T_{17} = 671,777 + 0,9233 \cdot 17 = 687,473;$$

$$T_{18} = 671,777 + 0,9233 \cdot 18 = 688,396.$$

Значения сезонных компонент за соответствующие кварталы равны: $S_1 = -292,448$ и $S_2 = -266,781$. Таким образом,

$$F_{17} = T_{17} + S_1 = 687,473 - 292,448 \approx 395;$$

$$F_{18} = T_{18} + S_2 = 688,396 - 266,781 \approx 422.$$

Т.е. в первые два квартала 2003 г. следовало ожидать порядка 395 и 422 правонарушений соответственно.

Построение мультипликативной модели рассмотрим на данных предыдущего примера.

Шаг 1. Методика, применяемая на этом шаге, полностью совпадает с методикой построения аддитивной модели.

Таблица 4.8

№ квартала, t	Количество правонарушений, \hat{y}_t	Итого за четыре квартала	Скользящая средняя за четыре квартала	Центрированная скользящая средняя	Оценка сезонной компоненты
1	2	3	4	5	6
1	375	—	—	—	—
2	371	2630	657,5	—	—
3	869	2612	653	655,25	1,3262
4	1015	2712	678	665,5	1,5252
5	357	2835	708,75	693,375	0,5149
6	471	2840	710	709,375	0,6640
7	992	2873	718,25	714,125	1,3891
8	1020	2757	689,25	703,75	1,4494
9	390	2757	689,25	689,25	0,5658
10	355	2642	660,5	674,875	0,5260
11	992	2713	678,25	669,375	1,4820
12	905	2812	703	690,625	1,3104
13	461	2740	685	694	0,6643
14	454	2762	690,5	687,75	0,6601
15	920	—	—	—	—
16	927	—	—	—	—

Шаг 2. Найдем оценки сезонной компоненты как частное от деления фактических уровней ряда на центрированные скользящие

средние (гр. 6 табл. 4.8). Эти оценки используются для расчета сезонной компоненты S (табл. 4.9). Для этого найдем средние за каждый квартал оценки сезонной компоненты S_i . Так же как и в аддитивной модели считается, что сезонные воздействия за период взаимопогашаются. В мультипликативной модели это выражается в том, что сумма значений сезонной компоненты по всем кварталам должна быть равна числу периодов в цикле. В нашем случае число периодов одного цикла равно 4.

Таблица 4.9

Показатели	Год	№ квартала, i			
		I	II	III	IV
	1999	–	–	1,3262	1,5252
	2000	0,5149	0,6640	1,3891	1,4494
	2001	0,5658	0,5260	1,4820	1,3104
	2002	0,6643	0,6601	–	–
Всего за i -й квартал		1,7450	1,8501	4,1973	4,2850
Средняя оценка сезонной компоненты для i -го квартала, \bar{S}_i		0,5816	0,6167	1,3991	1,4283
Скорректированная сезонная компонента, S_i		0,5779	0,6128	1,3901	1,4192

Имеем

$$0,5816 + 0,6167 + 1,3991 + 1,4283 = 4,0257.$$

Определяем корректирующий коэффициент:

$$k = 4 / 4,0257 = 0,9936.$$

Скорректированные значения сезонной компоненты S_i получаются при умножении ее средней оценки \bar{S}_i на корректирующий коэффициент k .

Проверяем условие равенство 4 суммы значений сезонной компоненты:

$$0,5779 + 0,6128 + 1,3901 + 1,4192 = 4.$$

Шаг 3. Разделим каждый уровень исходного ряда на соответствующие значения сезонной компоненты. В результате получим величины $T \cdot E = Y/S$ (гр. 4 табл. 4.10), которые содержат только тенденцию и случайную компоненту.

Таблица 4.10

t	y_t	S_i	y_t/S_i	T	$T \cdot S$	$E = y_t/(T \cdot S)$
1	2	3	4	5	6	7
1	375	0,5779	648,9012	654,9173	378,4767	0,9908
2	371	0,6128	605,4178	658,1982	403,3439	0,9198
3	869	1,3901	625,1349	661,4791	919,5221	0,9451
4	1015	1,4192	715,1917	664,7600	943,4274	1,0759
5	357	0,5779	617,7539	668,0409	386,0608	0,9247
6	471	0,6128	768,6031	671,3218	411,3860	1,1449
7	992	1,3901	713,6177	674,6027	937,7652	1,0578
8	1020	1,4192	718,7148	677,8836	962,0524	1,0602
9	390	0,5779	674,8572	681,1645	393,6450	0,9907
10	355	0,6128	579,3081	684,4454	419,4281	0,8464
11	992	1,3901	713,6177	687,7263	956,0083	1,0377
12	905	1,4192	637,6832	691,0072	980,6774	0,9228
13	461	0,5779	797,7159	694,2881	401,2291	1,1490
14	454	0,6128	740,8616	697,5690	427,4703	1,0621
15	920	1,3901	661,8229	700,8499	974,2515	0,9443
16	927	1,4192	653,1849	704,1308	999,3024	0,9277

Шаг 4. Определим компоненту T в мультипликативной модели. Для этого рассчитаем параметры линейного тренда, используя уровни $T \cdot E$. В результате получим уравнение тренда:

$$\hat{T} = 651,6364 + 3,2809 \cdot t.$$

Подставляя в это уравнение значения $t = 1, 2, \dots, 16$, найдем уровни T для каждого момента времени (гр. 5 табл. 4.10).

Шаг 5. Найдем уровни ряда, умножив значения T на соответствующие значения сезонной компоненты (гр. 6 табл. 4.10). На одном графике откладываем фактические значения уровней временного ряда и теоретические, полученные по мультипликативной модели.

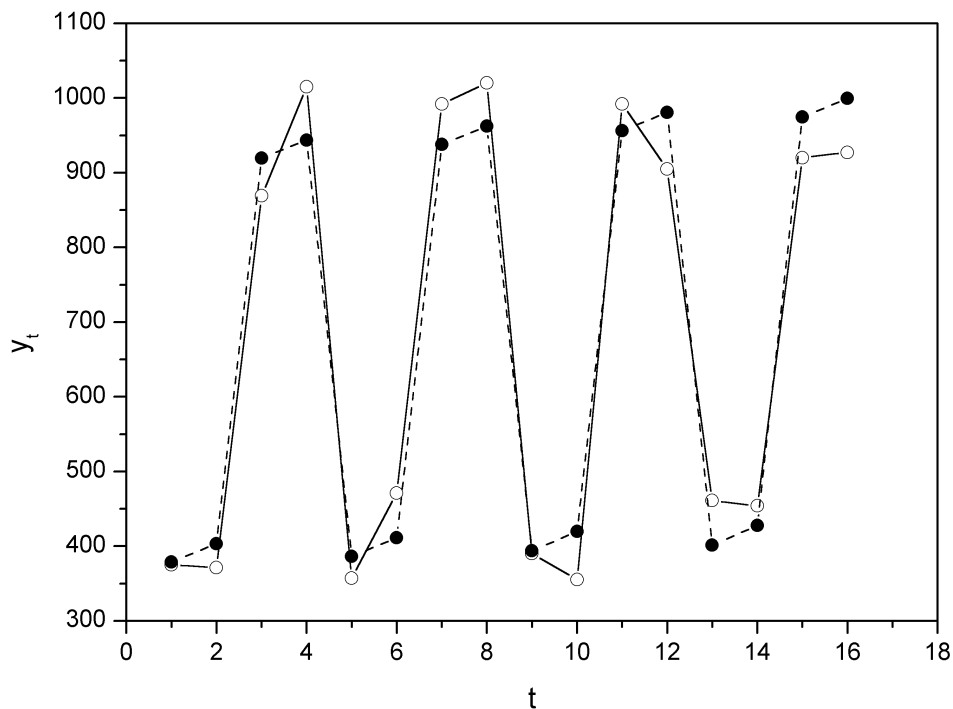


Рис. 4.7.

Расчет ошибки в мультипликативной модели производится по формуле:

$$E = Y / (T \cdot S).$$

Для сравнения мультипликативной модели и других моделей временного ряда можно, по аналогии с аддитивной моделью, использовать сумму квадратов абсолютных ошибок $(y_t - T \cdot S)^2$:

$$R^2 = 1 - \frac{\sum (y_t - T \cdot S)^2}{\sum (y_t - \bar{y})^2} = 1 - \frac{43065,02}{1252743,75} = 0,966.$$

Сравнивая показатели детерминации аддитивной и мультипликативной моделей, делаем вывод, что они примерно одинаково аппроксимируют исходные данные.

Шаг 6. Прогнозирование по мультипликативной модели. Если предположить, что по нашему примеру необходимо дать прогноз об общем объеме правонарушений на I и II кварталы 2003 года, прогнозное

значение F_t уровня временного ряда в мультипликативной модели есть произведение трендовой и сезонной компонент. Для определения трендовой компоненты воспользуемся уравнением тренда

$$T = 651,6364 + 3,2809 \cdot t.$$

Получим

$$T_{17} = 651,6364 + 3,2809 \cdot 17 = 707,4117;$$

$$T_{18} = 651,6364 + 3,2809 \cdot 18 = 710,6926.$$

Значения сезонных компонент за соответствующие кварталы равны: $S_1 = 0,5779$ и $S_2 = 0,6128$. Таким образом

$$F_{17} = T_{17} \cdot S_1 = 707,4117 \cdot 0,5779 \approx 409;$$

$$F_{18} = T_{18} \cdot S_2 = 710,6926 \cdot 0,6128 \approx 436.$$

Т.е. в первые два квартала 2003 г. следовало ожидать порядка 409 и 436 правонарушений соответственно.

Таким образом, аддитивная и мультипликативная модели дают примерно одинаковый результат по прогнозу.

4.4. Автокорреляция в остатках. Критерий Дарбина-Уотсона

Автокорреляция в остатках может быть вызвана несколькими причинами, имеющими различную природу.

1. Она может быть связана с исходными данными и вызвана наличием ошибок измерения в значениях результативного признака.

2. В ряде случаев автокорреляция может быть следствием неправильной спецификации модели. Модель может не включать фактор, который оказывает существенное воздействие на результат и влияние которого отражается в остатках, вследствие чего последние могут оказаться автокоррелированными. Очень часто этим фактором является фактор времени t .

От истинной автокорреляции остатков следует отличать ситуации, когда причина автокорреляции заключается в неправильной

спецификации функциональной формы модели. В этом случае следует изменить форму модели, а не использовать специальные методы расчета параметров уравнения регрессии при наличии автокорреляции в остатках.

Один из более распространенных методов определения автокорреляции в остатках – это расчет критерия Дарбина-Уотсона:

$$d = \frac{\sum_{t=2}^n (\varepsilon_t - \varepsilon_{t-1})^2}{\sum_{t=1}^n \varepsilon_t^2}. \quad (4.5)$$

Т.е. величина d есть отношение суммы квадратов разностей последовательных значений остатков к остаточной сумме квадратов по модели регрессии.

Можно показать, что при больших значениях n существует следующее соотношение между критерием Дарбина-Уотсона d и коэффициентом автокорреляции остатков первого порядка r_1 :

$$d \cong 2 \cdot (1 - r_1). \quad (4.6)$$

Таким образом, если в остатках существует полная положительная автокорреляция и $r_1 = 1$, то $d = 0$. Если в остатках полная отрицательная автокорреляция, то $r_1 = -1$ и, следовательно, $d = 4$. Если автокорреляция остатков отсутствует, то $r_1 = 0$ и $d = 2$. Т.е. $0 \leq d \leq 4$.

Алгоритм выявления автокорреляции остатков на основе критерия Дарбина-Уотсона следующий. Выдвигается гипотеза H_0 об отсутствии автокорреляции остатков. Альтернативные гипотезы H_1 и H_1^* состоят, соответственно, в наличии положительной или отрицательной автокорреляции в остатках. Далее по специальным таблицам (см. приложение Е) определяются критические значения критерия Дарбина-Уотсона d_L и d_U для заданного числа наблюдений n , числа независимых переменных модели m и уровня значимости α . По этим значениям числовой промежуток $[0; 4]$ разбивают на пять отрезков.

Принятие или отклонение каждой из гипотез с вероятностью $1 - \alpha$ осуществляется следующим образом:

$0 < d < d_L$ – есть положительная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1 - \alpha$ принимается H_1 ;

$d_L < d < d_U$ – зона неопределенности;

$d_U < d < 4 - d_U$ – нет оснований отклонять H_0 , т.е. автокорреляция остатков отсутствует;

$4 - d_U < d < 4 - d_L$ – зона неопределенности;

$4 - d_L < d < 4$ – есть отрицательная автокорреляция остатков, H_0 отклоняется, с вероятностью $P = 1 - \alpha$ принимается H_1^* .

Если фактическое значение критерия Дарбина-Уотсона попадает в зону неопределенности, то на практике предполагают существование автокорреляции остатков и отклоняют гипотезу H_0 .

Пример. Проверим гипотезу о наличии автокорреляции в остатках для аддитивной модели нашего временного ряда. Исходные данные и промежуточные расчеты заносим в таблицу:

Таблица 4.11

t	y_t	$\varepsilon_t = E$	ε_{t-1}	$(\varepsilon_t - \varepsilon_{t-1})^2$	ε_t^2
1	375	-5,252	–	–	27,584
2	371	-35,843	-5,252	935,8093	1284,7
3	869	-74,183	-35,843	1469,956	5503,1
4	1015	48,937	-74,183	15158,53	2394,8
5	357	-26,946	48,937	5758,23	726,09
6	471	60,464	-26,946	7640,508	3655,9
7	992	45,124	60,464	235,3156	2036,2
8	1020	50,244	45,124	26,2144	2524,5
9	390	2,361	50,244	2292,782	5,574
10	355	-59,229	2,361	3793,328	3508,1
11	992	41,431	-59,229	10132,44	1716,5
12	905	-68,450	41,431	12073,83	4685,4
13	461	69,668	-68,45	19076,58	4853,6
14	454	36,078	69,668	1128,288	1301,6
15	920	-34,263	36,078	4947,856	1174
16	927	-50,143	-34,263	252,1744	2514,3
Сумма		-0,002	50,141	84921,85	37911,97

Фактическое значение критерия Дарбина-Уотсона для данной модели составляет:

$$d = \frac{84921,85}{37911,97} = 2,24.$$

Сформулируем гипотезы: H_0 – в остатках нет автокорреляции; H_1 – в остатках есть положительная автокорреляция; H_1^* – в остатках есть отрицательная автокорреляция. Зададим уровень значимости $\alpha = 0,05$. По таблице значений критерия Дарбина-Уотсона определим для числа наблюдений $n = 16$ и числа независимых параметров модели $k = 1$ (мы рассматриваем только зависимость от времени t) критические значения $d_L = 1,10$ и $d_U = 1,37$. Фактическое значение d -критерия Дарбина-Уотсона попадает в интервал $d_U < d < 4 - d_U$ ($1,37 < 2,24 < 2,63$). Следовательно, нет основания отклонять гипотезу H_0 об отсутствии автокорреляции в остатках.

Существует несколько ограничений на применение критерия Дарбина-Уотсона.

1. Он неприменим к моделям, включающим в качестве независимых переменных лаговые значения результативного признака.
2. Методика расчета и использования критерия Дарбина-Уотсона направлена только на выявление автокорреляции остатков первого порядка.
3. Критерий Дарбина-Уотсона дает достоверные результаты только для больших выборок.