2.6 Аналіз адитивної моделі часового ряду

Кожний рівень часового ряду формується під впливом великої кількості факторів, які умовно можна поділити на три групи:

– фактори, що формують тренд ряду,

– фактори, що формують циклічні коливання ряду,

– випадкові фактори.

При різних комбінаціях цих факторів залежність рівнів ряду від часу може набувати різноманітних форм.

Показник, що досліджується, крім тренду, може містити циклічні коливання. Вони можуть бути сезонними (залежати від пори року). За наявності даних за тривалі проміжки часу у ряді можна виявити довгострокові циклічні коливання, період яких становить кілька років. Деякі часові ряди не містять тенденції та циклічної компоненти: кожний їхній наступний рівень є сумою середнього рівня ряду та деякої випадкової компоненти. Найчастіше кожний рівень ряду формується під дією довгострокової тенденції, сезонних коливань та випадкової компоненти. При цьому фактичний рівень часового ряду можна подати як суму чи добуток трендової, циклічної та випадкової компонент. Відповідні моделі називають *адитивною та мультиплікативною моделями часового ряду*.

Основною задачею дослідження окремого часового ряду є знаходження кожної з його компонент для того, щоб використати отриману інформацію для прогнозування майбутніх значень ряду. Аналіз часового ряду полягає у виокремленні його компонент. Методика аналізу часового ряду залежить від характеру зв'язку між ними: адитивного чи мультиплікативного.

Адитивною моделлю часового ряду називають модель, у якій зміна значень показника подається у вигляді суми окремих компонент:

$$Y = T+S+E,$$

де Y– фактичне значення рівня ряду, T – трендове значення, S – сезонна компонента, E – випадкова компонента (похибка). Побудова адитивної моделі часового ряду складається з наступних етапів:

1. Обчислення значень сезонної компоненти S;
2. Віднімання сезонної компоненти S від фактичних значень Y (десезоналізація ряду);
3. Розрахунок тренду на основі десезоналізованих даних;
4. Розрахунок похибок (різниці між фактичними та трендовими значеннями), тобто обчислення $E = Y–T$;
5. Розрахунок похибки використання адитивної моделі прогнозування – середнього абсолютного відхилення (MAD) або середньої квадратичної

похибки (*MSE*) для перевірки відповідності моделі фактичним даним або для вибору з кількох моделей найкращої.

Процес побудови адитивної моделі часового ряду та її використання для короткострокового прогнозування значень показника ряду розглянемо на наступному прикладі.

Приклад. У таблиці 2.8 наведено дані про обсяги реалізації продукції підприємства за останні 13 кварталів. Побудувати адитивну модель цього часового ряду та спрогнозувати на її основі обсяг реалізації продукції підприємства у 2-му кварталі 2022 р.

Таблиця 2.8. Обсяги реалізації продукції підприємства за 1-й квартал 2019 р. – 1-й квартал 2022 р.

Квартал	2019				2020				2021				2022
	1	2	3	4	1	2	3	4	1	2	3	4	1
Обсяг реалізації (тис. од.)	239	201	182	297	324	278	257	384	401	360	335	462	481

Спочатку визначимо тренд часового ряду, усунувши вплив сезонних коливань методом ковзної середньої, розрахованої для 4 періодів часу (кварталів). Склавши перші чотири рівні ряду та поділивши знайдену суму на 4, отримаємо середній квартальний обсяг реалізації у 2019 р.:

$$\frac{239+201+182+297}{4}=229,75\ (\text{тис. од.})$$

Отримане значення відноситься до середини 2019 р., тобто точки між 2-м та 3-м кварталами цього року. Зсуваючись на один квартал по часовому ряду, знаходимо наступні ковзні середні, які розглядають як наближені значення тренду (таблиця 2.9). Для оцінки значень сезонної компоненти S використовують рівність $Y–T = S+E$.

Оцінки значень тренду, отримані у результаті розрахунку ковзних середніх за 4 точками, відносяться до інших моментів часу, ніж фактичні дані часового ряду. Вони знаходяться між фактичними рівнями ряду. Перше значення ковзної середньої (229,75) – це точка між фактичними значеннями 2-го та 3-го кварталів 2019 р. (таблиця 2.9).

Таблиця 2.9. Розрахунок за чотирма точками центрованих ковзних середніх значень тренду

Квартал	Обсяг реалізації, тис. од.	Всього, за 4 квартали	Ковзна середня за 4 квартали	Центрована ковзна середня	Оцінка сезонної компоненти
2019 р. 1	239				
2	201				
		919	229,75		
3	182			240,4	–58,4
		1004	251		
4	297			260,6	+36,4
		1081	270,25		
2020 р. 1	324			279,6	+44,4
		1156	289		
2	278			299,9	–21,9
		1243	310,75		
3	257			320,4	–63,4
		1320	330		
4	384			340,3	+43,8
		1402	350,5		
2021 р. 1	401			360,2	+40,8
		1480	360,2		
2	360			379,8	–19,8
		1558	379,8		
3	335			399,5	–64,5
		1638	399,5		
4	462				
2022 р. 1	481				

Наступна оцінка (251) відноситься до точки між 3-м та 4-м кварталами. Нам же потрібно отримати десезоналізовані значення, що відповідають тим же проміжкам часу, що й фактичні значення з квартал. Положення десезоналізованих середніх у адитивній моделі зміщується у часі шляхом розрахунку середніх для кожної пари значень ковзних середніх, розрахованих за 4 квартали. Знаходимо середню для перших двох ковзних середніх, центруючи їх на 3-й квартал 2019 р., тобто $\dfrac{229,75+251,2}{2}=240,4$. Отримали десезоналізовану середню за 3-й квартал 2019 р. Цю величину, яку називають *центрованою ковзною середньою*, можна порівнювати з фактичним значенням показника за цей період часу. У результаті віднімання трендового значення, отриманого методом ковзної середньої, від фактичного значення показника, отримаємо оцінку сезонної компоненти, у яку входить похибка E.

Знайдемо середні значення оцінок для кожного сезону (кварталу) (таблиця 2.10).

Таблиця 2.10. Розрахунок середніх оцінок сезонної компоненти

	Рік	Номер кварталу			
		1	2	3	4
	2019	–	–	–58,4	+36,4
	2020	+44,4	–21,9	–63,4	+43,8
	2021	+40,8	–19,8	–64,5	–
Всього		+85,2	–41,7	–186,3	+80,2
Середнє значення		$+\frac{85,2}{2}$	$-\frac{41,7}{2}$	$-\frac{186,3}{3}$	$+\frac{80,2}{2}$
Оцінка сезонної компоненти		+42,6	–20,8	–62,1	+40,1
Відкоригована сезонна компонента		+42,6	–20,7	–62,0	+40,1

Це дозволяє дещо зменшити величину похибки. Річна сума середніх квартальних оцінок сезонних компонент повинна дорівнювати нулю. Якщо ця умова не виконується, то потрібно відкоригувати кожну з таких оцінок. Найчастіше це коригування здійснюється шляхом віднімання від них їх середньої арифметичної за рік. Тоді сума скоригованих оцінок сезонних компонент дорівнюватиме нулю.

У даному прикладі сума оцінок сезонних компонент:

$$42,6–20,8–62,1+40,1 = –0,2$$

Вона відмінна від нуля. Середнє значення оцінки сезонної компоненти $-\frac{0,2}{4} = -0,05$. Цю величину потрібно відняти від оцінки сезонного компоненту для кожного кварталу, щоб отримати відкориговане значення. У даному прикладі середня оцінка сезонної компоненти –0,05 менша за одиницю останнього розряду після коми у показника, що досліджується. Тому для коригування оцінки сезонної компоненти округлення до десятих після віднімання –0,05 здійснювалось так, щоб сума скоригованих сезонних компонент дорівнювала нулю.

Виконаний аналіз свідчить, що у першому та четвертому кварталах обсяги реалізації перевищують середнє трендове значення близько на 40 тис. одиниць, обсяги реалізації у другому та третьому кварталах нижчі за середні відповідно на 21 тис. одиниць та 62 тис. одиниць.

За аналогічною схемою можна визначити сезонну компоненту не лише для кварталів, але й будь-яких проміжків часу (місяць, рік).

Розглянемо процес десезоналізації даних для розрахунку тренду. Тут від фактичних значень показника віднімають відповідні значення сезонної компоненти: $Y - S = T + E$ (таблиця 2.11). Отримуємо нові оцінки значень тренду, які ще містять похибку.

Таблиця 2.11. Розрахунок десезоналізованого обсягу продаж

Квартал	Обсяг реалізації *Y*, тис. од.	Сезонна компонента S	Десезоналізований обсяг продажу, *Y–S=T+E*
2019 1	239	+42,6	196,4
2	201	–20,7	221,7
3	182	–62,0	244,0
4	297	+40,1	256,9
2020 1	324	+42,6	281,4
2	278	–20,7	298,7
3	257	–62,0	319,0
4	384	+40,1	343,9
2021 1	401	+42,6	358,6
2	360	–20,7	380,7
3	335	–62,0	397,1
4	462	+40,1	421,9
2022 1	481	+42,6	438,4

За цими оцінками побудуємо рівняння лінійного тренду у вигляді $T = a + bt$, де *t* – порядковий номер кварталу у часовому ряді, *t* = 1,2,…,13. Коефіцієнти *a* та *b* визначаємо з застосуванням методу найменших квадратів за формулами:

$$b = \frac{n\sum_{t=1}^{n}(t \cdot y_t) - \sum_{t=1}^{n} t \cdot \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2},\ a = \frac{\sum_{t=1}^{n} y_t}{n} - b \cdot \frac{\sum_{t=1}^{n} t}{n}.$$

Тут y_t – значення величини *T+E* для відповідного порядкового номера *t* отримані у останньому стовпці таблиці 2.11.

Підставивши сюди номери кварталів *t* та відповідні значення y_t, отримаємо рівняння тренду у вигляді:

$$T = 180 + 20t.$$

Для розрахунку середньої абсолютної похибки *MAD* та середньої квадратичної похибки *MSE* використаємо дані таблиці 2.12.

Таблиця 2.12. Розрахунок відхилень для адитивної моделі

Квартал	Номер	Y	S	T	E=Y–S–T
2019 1	1	239	+42,6	200	–3,6
2	2	201	–20,7	220	+1,7
3	3	182	–62,0	240	+4,0
4	4	297	+40,1	260	–3,1
2020 1	5	324	+42,6	280	+1,4
2	6	278	–20,7	300	–1,3
3	7	257	–62,0	320	–1,0
4	8	384	+40,1	340	+3,9
2021 1	9	401	+42,6	360	–1,6
2	10	360	–20,7	380	+0,7
3	11	335	–62,0	400	–3,0
4	12	462	+40,1	420	+1,9
2022 1	13	481	+42,6	440	–1,6

Знаходимо:

$$MAD = \frac{\sum_{t=1}^{n} |E_t|}{n} = \frac{28,7}{13} = 2,2; \quad MSE = \sqrt{\frac{\sum_{t=1}^{n} |E_t|^2}{n}} = \sqrt{\frac{78,85}{13}} = \sqrt{6,1} = 2,5.$$

У нашому випадку ці похибки досить малі і не перевищують 2% від середнього фактичного значення показника часового ряду. Це свідчить про те, що модель можна використовувати для короткострокового прогнозування.

Розглянемо побудову прогнозних значень за моделлю з адитивною компонентою. Їх знаходимо за формулою:

$$F = T + S,$$

де $T = 180 + 20t$.

Сезонна компонента для кварталів: 1-й квартал: $S_1 = +42,6$, 2-й квартал: $S_2 = -20,7$, 3-й квартал: $S_3 = +62,0$, 4-й квартал: $S_4 = +40,1$. Порядковий номер кварталу, що відповідає 2-му кварталу 2022 р. t = 14. Прогнозне значення тренду $T_{14} = 180 + 20 \cdot 14 = 460$ (тис. од.). Прогноз на 2-й квартал 2022 р.: F = 460–20,7 = 439,3 (тис. од.).

Чим більш віддаленим є період прогнозування, тим меншою є точність прогнозу.

2.7 Побудова мультиплікативної моделі часового ряду

Розглянуту адитивну модель застосовують, якщо сезонна компонента близька до сталої. Якщо зі збільшенням тренду вона зростає, то кращий результат отримуємо з використанням *мультиплікативної моделі* $Y = T \cdot S \cdot E$, де T – трендова компонента, S – індекс сезонності, E – відносна похибка, обумовлена впливом випадкових факторів.

Приклад. Підприємство реалізує кілька видів продукції, обсяги реалізації одного з них наведені у таблиці 2.13.

Таблиця 2.13. Квартальні обсяги реалізації продукції підприємства

Квартал	**2019**, 1	2	3	4	**2020**, 1	2	3	4	**2021**, 1	2	3	4	**2022**, 1
Порядковий номер, t	1	2	3	4	5	6	7	8	9	10	11	12	13
Обсяг реалізації Y, тис. одиниць	70	66	65	71	79	66	67	82	84	69	72	87	94

З даних таблиці 2.13 видно, що обсяг реалізації продукції підприємства має сезонний характер. При цьому його значення взимку більші, ніж влітку, а відхилення фактичних значень від лінії тренду зростають.

Для прогнозування за наведеним часовим рядом використаємо мультиплікативну модель. Для побудови тренду вирівняємо часовий ряд методом ковзних середніх. Алгоритм вирівнювання такий же, як і для адитивної моделі. Для початкової оцінки трендових значень використовують центровані ковзні середні. Оцінки сезонної компоненти (індекси сезонності) знаходять за формулою:

$$\frac{Y}{T} = S \cdot E .$$

Результати розрахунку наведені у таблиці 2.14.

Таблиця 2.14. Розрахунок індексів сезонності

Період часу	Порядковий номер, t	Обсяг реалізації, Y, тис. од.	Ковзна середня за 4 квартали	Центрована ковзна середня, T	Індекс сезонності, $Y/T = S \cdot E$
2019, 1	1	70			
2	2	66			
			68,0		
3	3	65		69,13	0,940
			70,25		
4	4	71		70,25	1,011
			70,25		
2020, 1	5	79		70,5	1,121
			70,75		
2	6	66		72,13	0,915
			73,50		
3	7	67		74,13	0,904
			74,75		
4	8	82		75,13	1,092
			75,50		
2021, 1	9	84		76,13	1,103
			76,75		
2	10	69		77,38	0,892
			78,0		
3	11	72		79,25	0,909
			80,50		
4	12	87			
2022, 1	13	94			

Значення індексів сезонності отримуємо на основі квартальних оцінок, розрахунок яких наведено у таблиці 2.15. Оскільки значення індексу сезонності – це частки, а кількість сезонів (кварталів) дорівнює чотирьом, то сума індексів сезонності повинна дорівнювати 4. Якщо ця сума не дорівнює чотирьом, то значення індексів сезонності коригуються за таким же алгоритмом, як коригування сезонної компоненти у адитивній моделі. Якщо б вихідні дані були відносно тижня і передбачалося 7 сезонів по 1 дню на протязі тижня, то загальна сума індексів сезонності повинна була б дорівнювати 7.

У таблиці 2.15 скориговані оцінки індексів сезонності отримані шляхом множення отриманих оцінок на відношення $\frac{4}{3{,}984} = 1{,}004$, де 3,984 – сума оцінок індексів сезонності до коригування.

Таблиця 2.15. Розрахунок скоригованих квартальних оцінок індексів сезонності

	Рік	Номер кварталу				
		1	2	3	4	
	2019	–	–	0,940	1,011	
	2020	1,121	0,915	0,904	1,092	
	2021	1,103	0,892	0,904	–	
Загалом		2,224	1,807	2,753	2,103	
Середнє значення		$\frac{2,224}{2}$	$\frac{1,807}{3}$	$\frac{2,753}{3}$	$\frac{2,103}{2}$	
Оцінка індексів сезонності		1,112	0,903	0,918	1,05	$\Sigma = 3,984$
Скориговані індекси сезонності		1,116	0,904	0,922	1,055	$\Sigma = 4$

Отримані оцінки свідчать, що внаслідок сезонних коливань обсяги реалізації продукції у першому кварталі збільшуються у середньому на 11,6% у порівнянні з трендовим значенням, для четвертого кварталу маємо збільшення цього показника на 5,5% відповідного значення тренду. У другому та третьому кварталах спостерігаємо сезонне зменшення обсягів реалізації відповідно на 9,6% та 7,8% у порівнянні з трендовими значеннями.

Здійснимо десезоналізацію фактичних даних про обсяги реалізації продукції і на її основі побудуємо рівняння лінійного тренду. Десезоналізація даних здійснюється за формулою $\frac{Y}{S} = T \cdot E$. Результати розрахунку оцінок трендових значень шляхом десезоналізації наведені у таблиці 2.16.

Таблиця 2.16. Розрахунок оцінок трендових значень часового ряду

Період часу	Порядковий номер, t	Обсяг реалізації, Y	Індекс сезонності, S	Десезоналізоваий обсяг реалізації $Y/S = T \cdot E$
2019 1	1	70	1,116	62,7
2	2	66	0,907	72,8
3	3	65	0,922	70,6
4	4	71	1,055	67,3
2020 1	5	79	1,116	70,8
2	6	66	0,907	72,8

3	7	67	0,922	72,7
4	8	82	1,055	77,7
2021 1	9	84	1,116	75,2
2	10	69	0,907	76,1
3	11	72	0,922	78,2
4	12	87	1,055	82,4
2022 1	13	94	1,116	84,2

Отримані оцінки використаємо для знаходження коефіцієнтів рівняння лінійного тренду $T = a + bt$. Їх знаходять так же, як і у випадку адитивної моделі, за формулами:

$$b = \frac{n\sum_{t=1}^{n}(t \cdot y_t) - \sum_{t=1}^{n} t \cdot \sum_{t=1}^{n} y_t}{n\sum_{t=1}^{n} t^2 - \left(\sum_{t=1}^{n} t\right)^2}, \quad a = \frac{\sum_{t=1}^{n} y_t}{n} - b \cdot \frac{\sum_{t=1}^{n} t}{n}.$$

Тут y_t – значення величини $\frac{Y}{S}$ для відповідного порядкового номера кварталу *t,* отримані у останньому стовпці таблиці 2.16.

Для отриманих десезоналізованих значень обсягів реалізації продукції отримуємо: $b = 1{,}36; a = 64{,}6$, рівняння тренду набуває вигляду:

$$T = 64{,}6 + 1{,}36t.$$

Отримане рівняння тренду застосуємо для розрахунку трендових обсягів реалізації продукції для кожного порядкового номеру кварталу *t*. Використовуючи ці дані, розрахуємо похибки мультиплікативної моделі $E = \frac{Y}{T \cdot S}$. Відповідний розрахунок наведений у таблиці 2.17.

Таблиця 2.17. Розрахунок похибок мультиплікативної моделі

Період часу	Порядковий номер кварталу, *t*	Обсяг реалізації, *Y*	Індекс сезонності, *S*	Трендове значення, *T*	$T \cdot S$	Похибка, $Y/(T \cdot S)$
2019 1	1	70	1,116	66,0	73,7	0,95
2	2	66	0,907	67,3	61,0	1,08
3	3	65	0,922	68,7	63,3	1,03
4	4	71	1,055	70,0	73,9	0,96
2020 1	5	79	1,116	71,4	79,7	0,99

2	6	66	0,907	72,8	66,0	1,0
3	7	67	0,922	74,1	68,3	0,98
4	8	82	1,055	75,5	79,7	1,03
2021 1	9	84	1,116	76,8	85,7	0,98
2	10	69	0,907	78,2	70,9	0,97
3	11	72	0,922	79,6	73,3	0,98
4	12	87	1,055	80,9	85,4	1,02
2022 1	13	94	1,116	82,3	91,9	1,02

Величина відхилень прогнозних значень $T \cdot S$ від фактичних значень обсягів реалізації продукції є невеликою, вона не перевищує 5%. Тому ми можемо використовувати мультиплікативну модель часового ряду $F = T \cdot S$ для короткострокового прогнозування. Тут $T = 64,6 + 1,36t$, індекси сезонності: $S_1 = 1,116$ для першого кварталу, $S_2 = 1,097$ для другого кварталу, $S_3 = 0,922$ для третього кварталу, $S_4 = 1,055$ для четвертого кварталу. Прогноз на другий квартал 2022 року, що відповідає значенню $t = 14$, за мультиплікативною моделлю часового ряду становить:

$$F = T \cdot S = \left(64,6 + 1,36 \cdot 14\right) \cdot 0,907 = 75,9 \text{ (тис. од.).}$$