

ПРАКТИЧНА РОБОТА 8

ПАРНИЙ ТА МНОЖИННИЙ РЕГРЕСІЙНИЙ АНАЛІЗ У ПРОГНОЗУВАННІ

Завдання №1. В таблиці 3.1 наведені вхідні дані двох інтервальних часових рядів Y_1 та Y_2 й моментного часового ряду Y_3 . На їх основі необхідно:

1. Побудувати діаграми розсіювання для часових рядів Y_1 , Y_2 та Y_3 ;
 2. Оцінити часові ряди Y_1 , Y_2 та Y_3 на однорідність, за допомогою квадратичного коефіцієнта варіації;
 3. Оцінити часові ряди Y_1 , Y_2 та Y_3 за достатністю обсягу даних з рівнем довірчої вірогідності у 90% та 95%;
 4. Порівняти за обчисленими показниками часові ряди Y_1 з Y_2 та Y_1 з Y_3 .
- Зробити висновок про придатність вхідних даних до прогнозування.

Таблиця 3.1 – Вхідні дані часових рядів Y_1 , Y_2 та Y_3

Період	Y_1	Y_2	Дата	Y_3
1	8,0	$= 3,5 - N/10$	01.07.2020	8,0
2	11,7	$= 7,2 - N/10$	01.08.2020	11,7
3	8,8	$= 4,3 - N/10$	01.11.2020	8,8
4	8,4	$= 3,9 - N/10$	01.12.2020	8,4
5	12,5	$= 8,0 - N/10$	01.01.2021	12,5
6	8,0	$= 3,5 - N/10$	01.03.2021	8,0
7	9,3	$= 4,8 - N/10$	01.06.2021	9,3
8	10,8	$= 6,3 - N/10$	01.09.2021	10,8
9	9,9	$= 5,4 - N/10$	01.10.2021	9,9
10	12,5	$= 8,0 - N/10$	01.12.2021	12,5
11	12,1	$= 7,6 - N/10$	01.01.2022	12,1
12	12,5	$= 8,0 - N/10$	01.02.2022	12,5
13	10,3	$= 5,8 - N/10$	01.06.2022	10,3
14	13,5	$= 9,0 - N/10$	01.07.2022	13,5
15	12,0	$= 7,5 - N/10$	01.08.2022	12,0
16	9,9	$= 5,4 - N/10$	01.09.2022	9,9
17	14,1	$= 9,6 - N/10$	01.10.2022	14,1
18	11,2	$= 6,7 - N/10$	01.11.2022	11,2
19	10,9	$= 6,4 - N/10$	01.01.2023	10,9
20	15	$= 10,5 - N/10$	01.02.2023	15
21	10,6	$= 6,1 - N/10$	01.03.2023	10,6
22	11,9	$= 7,4 - N/10$	01.04.2023	11,9
23	13,5	$= 9,0 - N/10$	01.10.2023	13,5
24	12,6	$= 8,1 - N/10$	01.11.2023	12,6

Де N – порядковий номер студента по журналу

Рекомендації до виконання. За ознакою часу, ряди динаміки можуть бути моментними, або інтервальними. Часовий ряд є моментним, якщо значення його рівнів вимірюються на конкретну дату (наприклад, вартість основних виробничих фондів на 1 січня). Часовий ряд є інтервальним, якщо значення його рівнів утворюються шляхом агрегування (підсумовування) за певний період (наприклад, обсяг виробленої продукції за січень).

При обчисленні варіації часового ряду, виникає необхідність розраховувати його математичне сподівання (середній рівень). Для моментних рядів середню величину визначають за формулами середньої хронологічної простої, або зваженої:

1. Якщо проміжки часу між датами рівні – середня хронологічна проста:

$$\bar{y} = \frac{\frac{1}{2}y_1 + y_2 + \dots + y_{n-1} + \frac{y_n}{2}}{n-1}, \quad (3.1)$$

Де $y_t, t = 1 \dots n$ – рівні ряду на момент часу t ; n – кількість спостережень.

2. Якщо проміжки часу між датами нерівні – середня хронологічна зважена:

$$\bar{y} = \frac{\sum_t \bar{y}_t f_t}{\sum f_i}, \quad \text{де} \quad \bar{y}_t = \frac{y_{t-1} + y_t}{2}, \quad (3.2)$$

Де $y_t, t = 1 \dots n$ – рівні ряду на момент часу t ; $\bar{y}_t$ – проміжна середня на момент часу t ; f_t – тривалість проміжку часу між спостереженнями y_{t-1} та y_t .

3. Для інтервальних рядів (рівних та нерівних інтервалів) середню величину визначають за формулою середньої арифметичної простої:

$$\bar{y} = \frac{\sum y}{n}, \quad (3.3)$$

Де n – загальна тривалість всього досліджуемого періоду.

Придатність часового ряду до прогнозування визначається на основі варіації його рівнів, тобто, коливань індивідуальних значень ознаки навколо математичного сподівання: чим меншою є варіація, тим сукупність є більш однорідною й тим надійнішими є характеристики центру розподілу.

Для вимірювання варіації рівнів часового ряду, використовуються показники: дисперсії, стандартного відхилення та квадратичного коефіцієнта варіації. Розглянемо методику їх розрахунку більш детально:

1. Дисперсія (σ^2) – це середнє значення квадрату відхилень рівнів часового ряду від його середнього рівня, формула (3.4).

$$\sigma^2 = \frac{\sum_{t=1}^n (y_t - \bar{y})^2}{n}, \quad (3.4)$$

Якщо вхідні дані представлені моментним часовим рядом з нерівними проміжками часу між вимірюваннями, тоді формула дисперсії обчислюється за середньою зваженою (3.5):

$$\sigma^2 = \frac{\sum (\bar{y}_t - \bar{y})^2 f_t}{\sum f_t}, \text{ де } \bar{y}_t = \frac{y_{t-1} + y_t}{2}, \quad (3.5)$$

Де $\bar{y}_t$ – проміжна середня величина; f_t – тривалість проміжку часу між спостереженнями y_{t-1} та y_t .

2. Стандартне відхилення (σ) – це корінь квадратний з дисперсії, формула (3.6).

$$\sigma = \sqrt{\sigma^2}, \quad (3.6)$$

Стандартне відхилення вимірює вплив випадкових факторів на коливання рівнів часового ряду навколо математичного сподівання. Чим меншим є значення σ , тим краще середній рівень $\bar{y}$ характеризує часовий ряд, тобто, дію головних факторів.

3. Квадратичний коефіцієнт варіації – це відношення стандартного відхилення до центру розподілу, формула (3.7).

$$K_{\sigma} = \frac{\sigma}{\bar{y}} \times 100\%, \quad (3.8)$$

Якщо $K_{\sigma} \leq 33,3\%$, то часовий ряд вважається однорідним й може використовуватись для подальшого прогнозування. Якщо ж $K_{\sigma} \geq 40,0\%$, то говорять про значні коливання рівнів ряду, а отже, побудований на їх основі прогноз не буде мати належної точності.

Поряд з варіацією, наявний часовий ряд повинний бути оцінений на достатність обсягу вхідних даних. Для цього, мінімально необхідний обсяг часового ряду буде розраховуватись за формулою (3.9).

$$n_{\min} = \frac{K_{\sigma}^2 * t_{\text{таблич}}^2}{m^2}, \quad (3.9)$$

Де m – показник, що відображає допустимий рівень похибки й становить 0,05 – 0,1 (5% - 10%); $t_{\text{таблич}}(n-1, P)$ – коефіцієнт Ст'юдента, що знаходиться з додатку А й залежить від параметрів: n – кількість рівнів часового ряду; P – рівень довірчої вірогідності ($P = 1 - m$).

Дані часового ряду можуть бути використані для подальшого прогнозування лише тоді, якщо виконуються обидві умови одночасно: ($K_{\sigma} \leq 33,3\%$) та ($n \geq n_{\min}$).

В протилежному випадку, статистичні дані вважають неоднорідними. Тобто, коливання рівнів часового ряду підпадають під значний вплив випадкових факторів, або їхній обсяг є недостатнім для побудови прогнозу. В будь якому випадку, кількість періодів упередження (прогнозних рівнів часового ряду) не може перевищувати $\frac{n}{3}$.

Завдання №2. В результаті порівняння в завданні №1 часових рядів Y_1 та Y_2 за однорідністю й достатністю обсягів було з'ясовано, що часовий ряд Y_1 є більш придатним для прогнозування. Тому, далі на його основі слід виконати наступні етапи:

1. Спрогнозувати динаміку часового ряду Y_1 на наступні 5 періодів, методами середнього абсолютного приросту та середнього темпу зростання. В якості базового рівня прогнозування, взяти середнє значення трьох останніх спостережень;

2. На основі ретроспективних даних оцінити точність обох прогнозів за допомогою показників: MAE , MSE , $RMSE$, $MAPE$ та $RMSPE$. Порівняти їх між собою;

3. Результати прогнозування показати на графіку. Зробити висновок про доцільність застосування методів середнього абсолютного приросту чи середнього темпу зростання для прогнозування вказаного часового ряду.

Рекомендації до виконання. Для відповіді на поставлені питання, необхідно:

При наявності ядро вираженого тренду в динаміці часового ряду, найбільш простими методами прогнозування є: метод середнього абсолютного приросту та середнього темпу зростання. Розглянемо їх більш детально.

1. Метод середнього абсолютного приросту. Прогнозний рівень економічного показника обчислюється у відповідності із середнім абсолютним приростом даної величини в минулому, формула (3.10).

$$y_{n+k} = y_0 + \bar{\Delta} \times k, \quad \text{де} \quad \bar{\Delta} = \frac{\Delta y}{n-1} = \frac{y_n - y_1}{n-1}, \quad (3.10)$$

Де y_{n+k} – прогнозне значення показника для k -ого періоду; y_0 – базовий рівень екстраполяції; $\bar{\Delta}$ – середній абсолютний приріст рівнів ряду; n – тривалість вхідного часового ряду (кількість періодів); k – номер періоду прогнозування.

Метод середнього абсолютного приросту застосовується у випадках, якщо досліджуваний показник має загальну лінійну тенденцію розвитку.

2. Метод середнього темпу зростання. Прогнозний рівень економічного показника обчислюється у відповідності із середнім темпом зростання даної величини в минулому, формула (3.11).

$$y_{n+k} = y_0 \times \bar{T}^k, \text{ де } \bar{T} = \sqrt[n-1]{\frac{y_n}{y_1}}, \quad (3.11)$$

Де $\bar{T}$ – середній темп зростання рівнів ряду.

Даний метод прогнозування застосовується у випадках, якщо загальна тенденція ряду динаміки характеризується показовою, або експоненційною кривою.

Очевидно, що прогнозні значення y_{n+k} будуть залежати від базового рівня екстраполяції y_0 , або від того, в який бік він буде відхилятися від тренду. В якості базового рівня екстраполяції може обиратись останній, або передостанній рівень ряду, однак такий підхід може призвести до зміщення прогнозної оцінки, рис. 3.1.

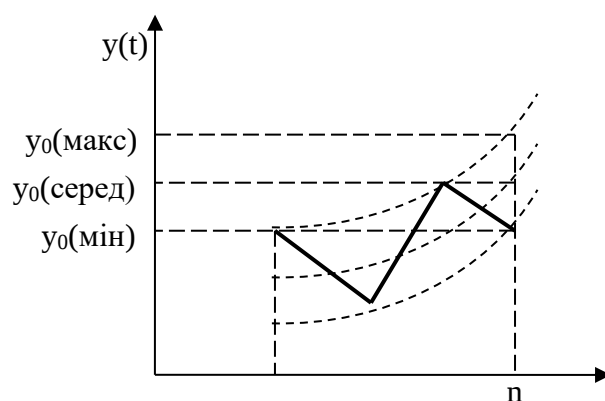


Рис. 3.1. Вибір базового рівня екстраполяції

Як бачимо з рис. 3.1, прийняття останнього рівня ряду в якості базового, призводить до зміщення прогнозної оцінки в бік зменшення; передостанній рівень ряду в якості базового, буде необґрунтовано збільшувати прогнозні

значення показника. Тому, на практиці, в якості y_0 рекомендується брати усереднене значення декількох останніх спостережень. Тому, в якості базового рівня прогнозування будемо брати середнє значення трьох останніх спостережень.

Слід пам'ятати, що розглянуті методи прогнозування є не лише найпростішими у застосуванні, але й самими наближеними. Тому, за точністю розрахунків, у більшості випадків вони поступаються іншим методам прогнозування.

Далі, на основі ретроспективних даних необхідно оцінити точність обох прогнозів за допомогою показників:

1. Середня абсолютна помилка прогнозування (MAE):

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_{fact,i} - y_i|, \quad (3.12)$$

Де $y_{fact,i}$, y_i – відповідно, фактичне та розрахункове значення досліджуваного показника в i -ому періоді; n – тривалість вхідного часового ряду.

2. Дисперсія (MSE) та середнє квадратичне відхилення ($RMSE$) помилки прогнозування:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_{fact,i} - y_i)^2, \quad (3.13)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_{fact,i} - y_i)^2}, \quad (3.14)$$

Зазначені індикатори (3.12)-(3.14) вимірюють похибку прогнозу в абсолютних одиницях виміру й тому їхня величина залежить від розмірності досліджуваного показника, що є їхнім недоліком.

3. Середня абсолютна помилка прогнозування у відсотках (*MAPE*):

$$MAPE = \frac{100\%}{n} \times \sum_{i=1}^n \left| \frac{y_{fact,i} - y_i}{y_{fact,i}} \right|, \quad (3.15)$$

4. Середнє квадратичне відхилення помилки прогнозування у відсотках (*RMSPE*):

$$RMSPE = 100\% \times \sqrt{\frac{1}{n} \sum_{i=1}^n \left(\frac{y_{fact,i} - y_i}{y_{fact,i}} \right)^2}, \quad (3.16)$$

Останні два індикатори (3.15)-(3.16) вимірюються у відносних одиницях. Тому, можна говорити про певний рівень адекватності прогнозної моделі на основі порівняння значень цих індикаторів з граничним рівнем, значення якого наведені в табл. 3.2.

Таблиця 3.2 – Критерії оцінки точності прогнозів

MAPE, RMSPE	Точність прогнозу
Менше 10%	Висока
10% – 20%	Добра
20% – 40%	Задовільна
40% – 50%	Погана
Більше 50%	Незадовільна

Завдання №3. На основі вхідних даних, додаток В та засобами візуального аналізу електронних таблиць Microsoft Excel, необхідно перевірити стаціонарність часових рядів Y_1 , Y_2 та Y_3 . Для цього, виконати наступні етапи:

1. Побудувати графіки динаміки кожного часового ряду: Y_1 , Y_2 та Y_3 ;
2. За допомогою ковзних середніх з інтервалом згладжування $m = 20$, побудувати графіки динаміки математичних сподівань та дисперсій для кожного часового ряду;

3. Побудувати графіки автокореляційних функцій кожного часового ряду й зробити висновки щодо їхньої стаціонарності.

Рекомендації до виконання. Стаціонарний часовий ряд повинний відповідати наступним умовам:

1. Математичне сподівання (середній рівень часового ряду) є постійною величиною й з часом не змінюється;

2. Дисперсія рівнів часового ряду також є постійною величиною й з часом не змінюється;

3. Відсутність, або слабка автокореляція між рівнями часового ряду. Вимога про відсутність автокореляції означає відсутність узгодженості коливань рівнів часового ряду між різними проміжками часу. Дана умова унеможливорює наявність сезонних, або циклічних коливань.

Відповідно до вказаних умов, прогноз для стаціонарного часового ряду завжди буде дорівнювати його математичному сподіванню, або середньому рівню.

Розглянемо більш детально умову про відсутність автокореляції. Функція $f(\Delta) = \text{Corr}(y_t; y_{t+\Delta})$ називається автокореляційною функцією, де $\text{Corr}(y_t; y_{t+\Delta})$ – це коефіцієнт кореляції Пірсона між різними інтервалами одного й того самого часового ряду Y . Один з інтервалів починається з періоду t , а інший – з періоду $(t+\Delta)$. Для стаціонарного часового ряду зі зростанням Δ , автокореляційна функція $f(\Delta)$ наближається до нуля:

– $\text{Corr} \in [-1; -0,7)$, або $\text{Corr} \in (+0,7; +1]$ – сильна зворотна, або пряма лінійна залежність;

– $\text{Corr} \in [-0,7; -0,3)$, або $\text{Corr} \in (+0,3; +0,7]$ – середня зворотна, або пряма лінійна залежність. Зв'язок існує, але він менш сильний;

– $\text{Corr} \in [-0,3; -0,1)$, або $\text{Corr} \in (+0,1; +0,3]$ – слабка зворотна, або пряма лінійна залежність. Зв'язок слабкий;

– $\text{Corr} \in [-0,1; +0,1]$ – дуже слабка, або відсутня лінійна залежність.

– $Corr \rightarrow 0$ – кореляція відсутня, лінійного статистичного взаємозв'язку між змінними не спостерігається.

При обчисленні автокореляційної функції, виходять з наступних рекомендацій щодо ширини інтервалів, які порівнюються між собою:

- для коротких часових рядів (до 100 спостережень) слід використовувати ширину інтервалу від 5 до 10;
- для часових рядів середньої довжини (100-500 спостережень) слід використовувати ширину інтервалу від 10 до 20;
- для довгих часових рядів (більше 500 спостережень) слід використовувати ширину інтервалу від 20 до 50.

В нашому випадку довжина часового ряду дорівнює 100 періодам, тому в якості ширини інтервалів при побудові кореляційної функції, прийmemo $m = 20$. Для перевірки умов стаціонарності часових рядів, будемо використовувати вбудовані функції Microsoft Excel. На рис. 3.1 показаний приклад аналізу часового ряду Y_1 .

C21 ✕ ✓ f_x =AVERAGE(B2:B21)					
	A	B	C	D	E
1	Період	Часовий ряд Y_1	Середнє Y_1	Дисперсія Y_1	Автокореляція Y_1
2	1	54,47			
3	2	59,17			
4	3	62,09			
5	4	53,82			
6	5	49,69			
7	6	49,44			
8	7	64,47			
9	8	68,75			
10	9	74,50			
11	10	66,95			
12	11	60,73			
13	12	59,77			
14	13	74,15			
15	14	77,66			
16	15	83,94			
17	16	75,49			
18	17	67,25			
19	18	65,04			
20	19	80,48			
21	20	85,82	66,68	112,17	1,0000
22	21	90,94	68,51	130,80	0,7719

Рис. 3.1. Аналіз часового ряду Y_1

В стовпцях А:В, рис. 3.1, розташовані вхідні дані часового ряду Y_1 , які містять 100 спостережень, додаток В. Його динаміка наведена на рис. 3.2.

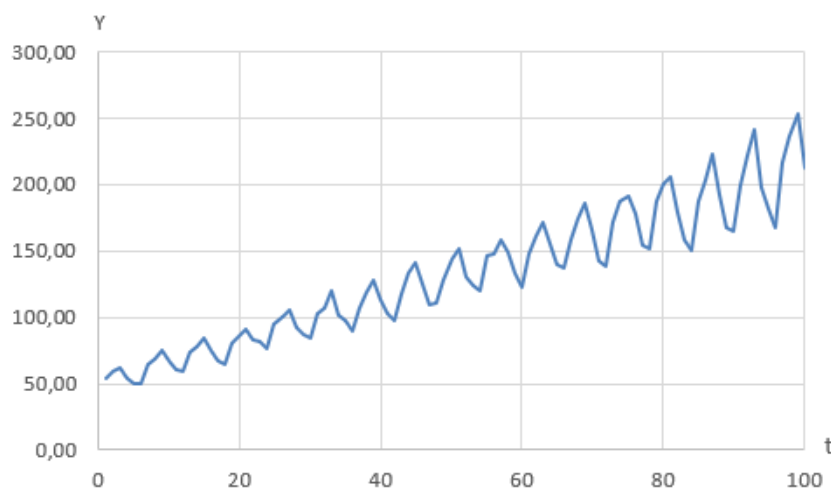


Рис. 3.2. Динаміка часового ряду Y_1

З рис. 3.2 видно, що часовий ряд Y_1 не відповідає жодному критерію стаціонарності:

- середній рівень ряду з часом має тенденцію до зростання;
- амплітуда коливань рівнів часового ряду, а отже й варіація навколо середнього рівня також з часом збільшуються. Відповідно, дисперсія повинна зростати;
- вхідний ряд містить циклічні коливання, а отже, умова про відсутність автокореляції також не задовольняється.

Перевіримо дані висновки за допомогою аналітичних й візуальних засобів Excel.

В колонці С, рис. 3.1, будемо обчислювати динаміку ковзної середньої з інтервалом згладжування $m = 20$. Для цього, в комірку С21 слід ввести формулу `=AVERAGE(B2:B21)`, або `=СРЗНАЧ(B2:B21)` й за правий нижній кут розтягнути її вниз до комірки С101. Відповідна динаміка ковзної середньої в графічному вигляді показана на рис. 3.3.

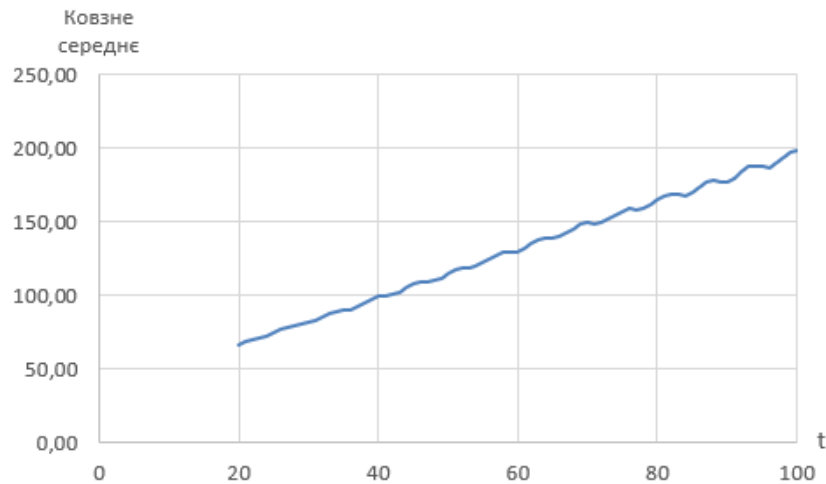


Рис. 3.3. Динаміка ковзної середньої часового ряду Y_1

Отже, рис. 3.3 наочно демонструє, що середній рівень ряду з часом має тенденцію до зростання. Тобто, математичне сподівання не є постійною величиною й перша умова стаціонарності не виконується.

В колонці D, рис. 3.1, будемо обчислювати динаміку ковзної дисперсії з інтервалом згладжування $m = 20$. Для цього, в комірку D21 слід ввести формулу `=VAR.P(B2:B21)`, або `=ДИСП.Г(B2:B21)` й за правий нижній кут розтягнути її вниз до комірки D101. Відповідна динаміка ковзної дисперсії часового ряду Y_1 показана на рис. 3.4.

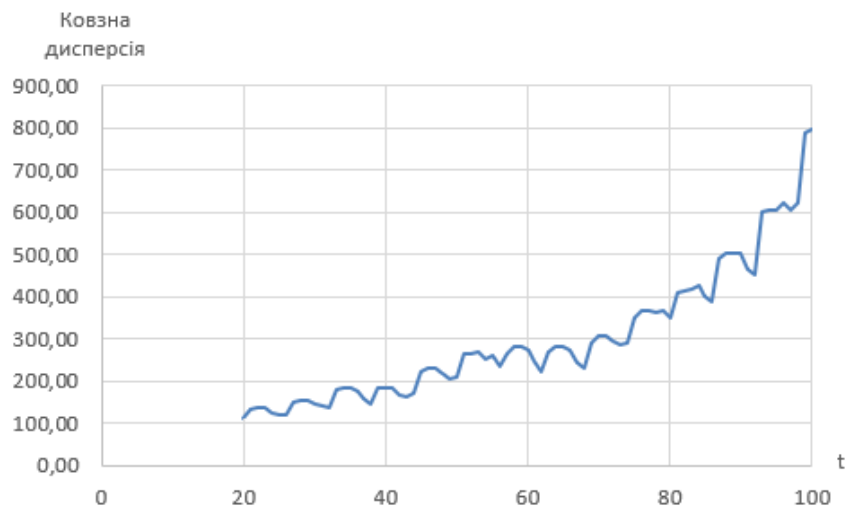


Рис. 3.4. Динаміка ковзної дисперсії часового ряду Y_1

З рис. 3.4 видно, що дисперсія рівнів часового ряду не є постійною величиною й з часом зростає. Тобто, друга умова стаціонарності також не виконується.

В колонці Е, рис. 3.1, будемо обчислювати значення автокореляційної функції з шириною інтервалів $m = 20$. Для цього, в комірку Е21 слід ввести формулу `=CORREL($B$2:$B$21;B2:B21)`, або `=КОРРЕЛ($B$2:$B$21;B2:B21)` й за правий нижній кут розтягнути її вниз до комірки Е101. Графік відповідної автокореляційної функції часового ряду Y_1 показаний на рис. 3.5.

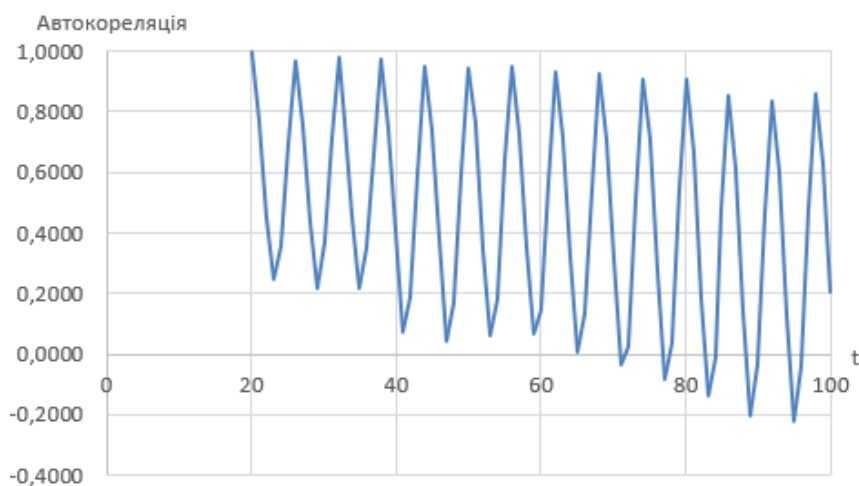


Рис. 3.5. Автокореляційна функція часового ряду Y_1

Умова про відсутність автокореляції означає, що відповідний графік, рис. 3.5, з часом повинний наближатись до нуля, що не відповідає дійсності. Таким чином, часовий ряд Y_1 не є стаціонарним: він має зростаючий тренд з циклічними коливаннями, амплітуда яких також збільшується.

Необхідно відтворити результати аналізу часового ряду Y_1 й аналогічним чином дослідити умови стаціонарності для часових рядів Y_2 та Y_3 .

Завдання №4. Використовуючи статистичні дані за 15 років про валовий випуск продукції Y , динаміку виробничих фондів K та трудових ресурсів L промислового підприємства необхідно:

1. Спрогнозувати динаміку використання основних виробничих фондів K на наступні 3 періоди, використовуючи: лінійну, степеневу та параболічну залежність другого порядку. Обрати найбільш достовірну з них;

2. Спрогнозувати динаміку використання трудових ресурсів L на наступні 3 періоди, використовуючи: експоненційну, логарифмічну та гіперболічну залежності. Обрати найбільш достовірну з них;

3. Побудувати виробничу функцію залежності випуску продукції від використання основних виробничих фондів й трудових ресурсів у вигляді лінійної регресії ($Y = a_0 + a_1K + a_2L$) та степеневі ($Y = a_0K^{a_1}L^{a_2}$). Оцінити їхню достовірність за допомогою критерію Фішера.

4. На основі прогнозних значень K та L розрахувати прогнозне значення Y , використовуючи найбільш адекватну виробничу функцію, отриману вище.

5. Побудувати графіки досліджуваних показників K , L та Y за вхідними та прогнозними даними. Додати на графіки розраховані трендові компоненти.

Таблиця 3.3 – Вхідні дані для побудови виробничої функції

Рік	Y	K	L
1	$= 400 + N$	311	92
2	485	305	100
3	490	348	122
4	$= 560 - N$	398	118
5	549	400	143
6	612	471	159
7	580	465	148
8	618	534	177
9	643	527	$= 153 + N$
10	685	544	200
11	717	586	238
12	769	592	266
13	$= 856 - N$	584	271
14	911	610	290
15	998	597	340

Де N – порядковий номер студента по журналу

Рекомендації до виконання. Достовірність всіх побудованих прогнозів в даному завданні будемо перевіряти за допомогою критерію Фішера, формула (3.17).

$$F_{\text{розн}} = \frac{\sum_{i=1}^n (y_{i,\text{розн}} - \overline{y_{\text{розн}}})^2}{m} \times \frac{n - m - 1}{\sum_{i=1}^n (y_i - y_{i,\text{розн}})^2}, \quad (3.17)$$

Де y_i – ретроспективний рівень часового ряду в i -ий період; $y_{i,\text{розн}}$ – розрахункове значення ретроспективного рівня часового ряду в i -ий період часу, згідно побудованої прогнозної моделі; $\overline{y_{\text{розн}}}$ – середнє значення розрахункових рівнів часового ряду за n періодів; m – кількість факторів, що впливають на розрахунковий показник.

При прогнозуванні, найчастіше єдиним фактором впливу на розрахунковий показник є час ($m = 1$). Дане твердження буде актуальним при прогнозуванні динаміки використання основних виробничих фондів K та трудових ресурсів L , завдання 4.1-4.2 (парний регресійний аналіз). При побудові виробничої функції $Y = f(K, L)$, завдання 4.3, кількість факторів $m = 2$ (множинний регресійний аналіз).

Отримана прогнозна модель вважається адекватною ретроспективним даним, якщо $F_{\text{розн}} > F_{\text{табл}}(P, m, n)$. Табличне значення $F_{\text{табл}}$ можна розрахувати в Microsoft Excel за допомогою функції:

$$= \text{FINV}(P; m; n-m-1), \text{ або} \\ = \text{FPACПОБР}(P; m; n-m-1).$$

Тут P – це допустимий рівень похибки. При $P = 0,05$ рівень надійності складає 0,95, або 95%.

Прогнозування динаміки використання основних виробничих фондів K на наступні 3 періоди, згідно поставленого завдання, відбувається на основі лінійної, степеневі та параболічної залежностей другого порядку. Причому, побудована трендова компонента повинна максимально наближатись до ретроспективних рівнів вхідного часового ряду. Для вирішення даного завдання

використовується метод найменших квадратів. Розглянемо специфіку його застосування для вказаних трьох випадків.

1. Лінійна залежність $y_{\text{розн}} = a_0 + a_1 x$. Невідомі параметри лінійної регресії a_0 та a_1 знаходяться за формулами:

$$a_0 = \frac{\sum y_i \sum x_i^2 - \sum x_i \sum x_i y_i}{n \sum x_i^2 - (\sum x_i)^2}; \quad a_1 = \frac{n \sum x_i y_i - \sum x_i \sum y_i}{n \sum x_i^2 - (\sum x_i)^2}, \quad (3.18)$$

2. Степенева залежність $y_{\text{розн}} = a_0 x^{a_1}$. Шляхом послідовного логарифмування та заміни змінних, спочатку перетворюємо степеневу залежність до лінійного вигляду:

$$\ln(y_{\text{розн}}) = \ln(a_0 x^{a_1}), \text{ звідки отримуємо:}$$

$$\ln(y_{\text{розн}}) = \ln(a_0) + \ln(x^{a_1}) = \ln(a_0) + a_1 \times \ln(x).$$

Далі проводимо заміну змінних:

$$\ln(y_{\text{розн}}) = \hat{y};$$

$$\ln(a_0) = \hat{a}_0;$$

$$\ln(x) = \hat{x}.$$

Тоді остаточне рівняння набуває лінійного вигляду:

$$\hat{y} = \hat{a}_0 + a_1 \hat{x}.$$

Після знаходження методом найменших квадратів невідомих параметрів $\hat{a}_0$ та a_1 , необхідно розрахувати початкове a_0 , згідно попередньої заміни:

$$a_0 = e^{\hat{a}_0}.$$

3. Парабола другого порядку $y_{\text{розн}} = a_0 + a_1 x + a_2 x^2$. Введемо заміну змінних: $x_1 = x^2$. Тоді, рівняння параболи прийме вигляд:

$$y_{\text{розн}} = a_0 + a_1 x + a_2 x_1.$$

Як бачимо, початкова нелінійна залежність з однією змінною x перетворилась на лінійну залежність з двома змінними x та x_1 . Тобто, маємо випадок множинної регресії, який буде розглядатись нижче.

Кожну з побудованих регресій слід перевірити на достовірність вхідним даним за допомогою критерію Фішера й обрати ту, де $F_{\text{розр}}$ приймає найбільше значення.

Прогнозування динаміки використання трудових ресурсів L на наступні 3 періоди, згідно поставленого завдання, слід виконувати за допомогою експоненційної, логарифмічної та гіперболічної залежностей. Для застосування з цією метою методу найменших квадратів, спочатку слід перетворити вказані залежності до лінійного вигляду.

1. Експоненційна залежність $y_{\text{розр}} = a_0 e^{a_1 x}$. Перш ніж провести заміну змінних, виконаємо логарифмування лівої й правої частин нашого рівняння:

$$\ln(y_{\text{розр}}) = \ln(a_0 e^{a_1 x}), \text{ звідки отримуємо:}$$

$$\ln(y_{\text{розр}}) = \ln(a_0) + \ln(e^{a_1 x}) = \ln(a_0) + a_1 x \times \ln(e) = \ln(a_0) + a_1 x.$$

Далі проводимо заміну змінних:

$$\ln(y_{\text{розр}}) = \hat{y};$$

$$\ln(a_0) = \widehat{a_0}.$$

Тоді остаточне рівняння набуває лінійного вигляду:

$$\hat{y} = \widehat{a_0} + a_1 x.$$

Після знаходження методом невідомих параметрів $\widehat{a_0}$ та a_1 , необхідно розрахувати початкове a_0 :

$$a_0 = e^{\widehat{a_0}}.$$

2. Логарифмічна залежність $y_{\text{розр}} = a_0 + a_1 \ln(x)$. Введемо заміну змінних:

$$\ln(x) = \hat{x}.$$

Тоді, початкове рівняння знову приймає лінійну форму:

$$y_{\text{розр}} = a_0 + a_1 \hat{x}.$$

3. Гіперболічна залежність $y_{\text{розр}} = a_0 + \frac{a_1}{x}$. Введемо заміну змінних:

$$\frac{1}{x} = \hat{x}.$$

Тоді, початкове рівняння прийме лінійну форму:

$$y_{\text{розр}} = a_0 + a_1 \hat{x}.$$

Так само, з наведених залежностей слід обирати ту, яка надає найвищий рівень достовірності за критерієм Фішера, формула (3.17).

Наступним етапом нашого аналізу є побудова виробничої функції залежності випуску продукції Y від використання основних виробничих фондів K й трудових ресурсів L у вигляді лінійної та степеневі регресії. Маємо задачу з побудови множинної регресії, оскільки результативний показник залежить від двох факторів одночасно.

По аналогії з парним регресійним аналізом, метод найменших квадратів застосовується для визначення невідомих параметрів лише для множинної регресії лінійного типу. Тому, у випадку нелінійної залежності, вона по аналогії повинна спочатку бути приведена до лінійного вигляду шляхом алгебраїчних перетворень: заміни змінних та логарифмування.

Нехай, лінійна залежність має вигляд:

$$y_{\text{розн}} = b_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n, \quad (3.19)$$

Де n – кількість факторів; $b_0, b_1, \dots, b_n$ – невідомі параметри множинної регресії, які необхідно знайти.

Для того, щоб МНК повернув нам всі невідомі параметри, включаючи b_0 , до складу регресії необхідно включити фіктивну змінну $x_0=1$. З урахуванням цього, рівняння (3.19) запишеться у вигляді:

$$y_{\text{розн}} = b_0x_0 + b_1x_1 + b_2x_2 + \dots + b_nx_n, \quad (3.20)$$

Представимо вхідні дані у матричному вигляді:

$$Y = \begin{bmatrix} Y_1 \\ Y_2 \\ \dots \\ Y_m \end{bmatrix}, X = \begin{bmatrix} x_{10} & x_{11} & \dots & x_{1n} \\ x_{20} & x_{21} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ x_{m0} & x_{m1} & \dots & x_{mn} \end{bmatrix} = \begin{bmatrix} 1 & x_{11} & \dots & x_{1n} \\ 1 & x_{21} & \dots & x_{2n} \\ \dots & \dots & \dots & \dots \\ 1 & x_{m1} & \dots & x_{mn} \end{bmatrix}, B = \begin{bmatrix} b_0 \\ b_1 \\ \dots \\ b_n \end{bmatrix}$$

Тоді, розрахунок невідомих параметрів рівняння (3.20) буде відбуватись шляхом наступних матричних перетворень:

$$B = [X]^T [X]^{-1} [X]^T Y, \quad (3.21)$$

Параметри рівняння регресії (3.19) у Microsoft Excel також можна знайти за допомогою вбудованої функції =LINEAR(), або =ЛИНЕЙН(). Дана функція має наступні аргументи:

- I аргумент – вектор-стовпець відомих значень результативного показника Y;
- II аргумент – матриця відомих значень факторів X. Дана матриця **не повинна** містити колонку фіктивних змінних $x_0=1$;
- III, IV аргументи – логічні параметри, повинні приймати значення «істина», або 1.

Після введення даної функції, виділити діапазон комірок, куди повинні повернутись розраховані значення. Розмірність цього діапазону $5 \times (n + 1)$. Далі слід натиснути <F2> та одночасно <CTRL>+<SHIFT>+<ENTER>. Розраховані значення параметрів регресії, які необхідно було знайти, будуть розташовані у першому рядку, таблиця 3.4.

Таблиця 3.4 – Формат даних вбудованої функції =LINEAR()

b_n	...	b_2	b_1	b_0
...	...	...	...	...
Fрозр				
...	...	...	...	...

У випадку степеневій множинній регресії, її також спочатку слід привести до лінійної форми. Нехай, необхідно побудувати регресію виду:

$$y_{\text{розр}} = b_0 x_1^{b_1} x_2^{b_2} \times \dots \times x_n^{b_n}, \quad (3.22)$$

Після логарифмування даного рівняння, отримаємо:

$$\ln(y_{\text{розр}}) = \ln(b_0 x_1^{b_1} x_2^{b_2} \times \dots \times x_n^{b_n}), \text{ звідки будемо мати}$$

$$\ln(y_{\text{розр}}) = \ln(b_0) + b_1 \ln(x_1) + b_2 \ln(x_2) + \dots + b_n \ln(x_n).$$

Дали виконуємо заміну змінних:

$$\ln(y_{\text{розр}}) = \hat{y};$$

$$\ln(b_0) = \widehat{b_0};$$

$$\ln(x_i) = \widehat{x_i}.$$

Тоді остаточне рівняння набуває лінійного вигляду:

$$\hat{y} = \widehat{b_0} + b_1 \widehat{x_1} + b_2 \widehat{x_2} + \dots + b_n \widehat{x_n}.$$

Тепер, на основі вхідних даних $\hat{y}$, $\widehat{x_1}$, $\widehat{x_2}$, ..., $\widehat{x_n}$ за допомогою рівняння (3.21), або вбудованої функції =LINEAR() можна обчислити невідомі параметри регресії $\widehat{b_0}$, b_1 , b_2 , ..., b_n .

Достовірність обох отриманих моделей, (3.19) та (3.21) також слід перевірити за допомогою критерію Фішера, формула (3.17) й візуалізувати результати прогнозування у вигляді відповідних діаграм.