

ПРОГНОЗУВАННЯ РОЗВИТКУ СКЛАДНИХ ІМОВІРНІСНИХ СИСТЕМ

У науці є щось таке, від чого захоплює дух.
Володіючи дріб'язковими запасами фактів,
за допомогою науки можна висунути
неймовірне число гіпотез та припущень.
Марк Твен. Життя на Міссісіпі

Сучасне прогнозування як специфічний вид наукового аналізу знаходить широке застосування в усіх сферах технічних, природних та соціальних наук. Передбачення стану техніко-технологічного устаткування, прогноз погоди, врожайності сільськогосподарських культур дозволяють інженеру, фермеру, товаровиробнику завчасно приготуватись до них, урахувати їх позитивні та негативні наслідки, а коли це можливо – втрутитись у процес розвитку, і що більш важливе – управляти ними. Рішення, які приймаються сьогодні, спираються на оцінки розвитку явищ у минулому, в свою чергу вони у більшому чи меншому ступеню впливають на майбутнє.

I. ОСНОВНІ ПОНЯТТЯ, КЛАСИФІКАЦІЇ ТА МЕТОДИ ПРОГНОЗУВАННЯ

Прогнозування (від грец. *prognosis* – знання наперед) є імовірнісне оцінювання майбутніх значень параметрів систем, що вивчаються, і є важливою попередньою стадією їх управління.

Прогнозування є головною методологічною базою прийняття рішень в процесі управління будь-якими складними імовірнісними системами – природними, технічними, економічними, екологічними тощо. Наукова обґрунтованість цієї бази стає особливо актуальною як при короткостроковому, так і при середньостроковому й довгостроковому прогнозуванні сучасних явищ та процесів.

Існує безліч методів прогнозування, які діляться на дві великі групи – інтуїтивні і формалізовані.

Інтуїтивні методи застосовуються при прогнозуванні розвитку дуже складних імовірнісних систем, коли врахувати всі важливі фактори неможливо. Серед найбільш популярних інтуїтивних методів прогнозування слід указати метод експертних оцінок – частіше за інших групових (метод мозкової атаки, метод Дельфі, методи нечіткої логіки та ін.).

Формалізовані методи застосовуються при прогнозуванні розвитку порівняно простих імовірнісних систем, коли всі важливі фактори піддаються обліку та виміру. Попереднім етапом прогнозування при цьому є формалізація

інформації про попередній розвиток процесу в період передісторії у вигляді певних кількісних показників і моделей. Довжина періоду передісторії (*ряду динаміки* або *часового ряду*) позначається через N (рис. 1), а довжина періоду упередження (прогнозування) – через L .

Інтервал або момент часу	Інтервал або момент часу 1	Інтервал або момент часу 2	...	Інтервал або момент часу N
Рівень Y_i	Y_1	Y_2	...	Y_N

Рис. 1 – Загальна схема ряду динаміки

Головними елементами ряду динаміки є рівні ряду Y_i та інтервали або моменти часу, до яких вони відносяться.

В основі прогнозування лежить екстраполяція знайдених у часових рядах тенденцій, тобто поширення виявлених у період передісторії (на відрізку часу 1, 2, ..., N) закономірностей і зв'язків на майбутнє – на період упередження L .

Екстраполяція полягає в гіпотезі про інерційності розвитку систем, на припущенні про те, що в недалекій перспективі виявлені в минулому зв'язки та закономірності кардинально не зміняться і діятимуть деякий час і в майбутньому. Слід зазначити, що таке припущення є досить реальним, особливо для короткострокового періоду, оскільки докорінні зміни показників більшості реально діючих систем зазвичай вимагають значних зусиль, коштів і тривалого часу.

Розрізняють точковий прогноз (у вигляді одного числа) та інтервальний прогноз (у вигляді двох чисел – нижньої та верхньої меж довірчого інтервалу). Достовірність довірчого інтервалу P зазвичай задається самим дослідником і показує імовірність попадання майбутнього фактичного значення прогнозованого параметра системи у інтервал від нижньої до верхньої межі. Найбільш популярними в практичних задачах прогнозування є 95-відсотковий довірчий інтервал (рівень достовірності $P = 0,95$; вірогідність помилки $1 - P = 0,05$), 99-відсотковий довірчий інтервал (рівень достовірності $P = 0,99$; вірогідність помилки $1 - P = 0,01$), 90-відсотковий довірчий інтервал (рівень достовірності $P = 0,90$; вірогідність помилки $1 - P = 0,10$).

Існують дві важливіші характеристики прогнозу: точність і достовірність. Точність визначається шириною довірчого інтервалу, тобто різницею між верхньою та нижньою межами інтервального прогнозу. Чим вища ця різниця (чим ширше довірчий інтервал), тим нижча точність прогнозу, і навпаки. Найбільшою точністю володіє точковий прогноз у вигляді одного числа, межі якого співпадають, а ширина дорівнює нулю. Інтервальний прогноз менш точний у порівнянні з точковим, однак більш достовірний, оскільки достовірність P визначається імовірністю попадання майбутнього значення прогнозованого параметра системи у довірчий інтервал. Достовірність точкового прогнозу близька до нуля, так як будь-яке відхилення майбутнього

фактичного значення від прогнозованого (точки на числовій осі) вважається помилкою прогнозу.

Отже, можна зробити висновок про те, що точність і достовірність прогнозу є характеристиками, що знаходяться у зворотному зв'язку між собою: з підвищенням одного параметра інший параметр знижується, і навпаки. Завдання дослідника полягає в знаходженні компромісу між точністю й достовірністю прогнозування. З указаних позицій точковий прогноз є максимально точним і мінімально достовірним. У процесі завдання потрібних рівнів достовірності прогнозу ($P = 90, 95, 99$ відсотків), дослідник підвищує ширину довірчого інтервалу, але при цьому його точність прогнозування падає. Тому завжди потрібно шукати компроміс між достатньою точністю та достовірністю оцінки майбутнього стану параметрів системи, що досліджується.

Прогнози класифікуються за різними ознаками. Одну з можливих класифікацій наведено на рис. 2.

Прогнози, побудовані на основі рядів динаміки параметрів складних імовірнісних систем, зазвичай належать до пошукових формалізованих прогнозів, що базуються на виявленій тенденції. Характерною особливістю попереджувальних прогнозів є використання схеми «якщо параметр системи досягне прогнозного рівня X , то система опиниться в стані Y ». Що ж до інших ознак класифікації, всі вони відрізнятимуться одна від одної у кожному конкретному випадку прогнозування.

У ході прогнозування можуть бути застосовувані абсолютні, відносні та середні характеристики часових рядів (табл. 1).

Таблиця 1 – Важливіші характеристики рядів динаміки

Показник	Абсолютний приріст (Π)	Темп росту (T_p)
Базисний	$Y_i - Y_{i-t}$	$\frac{Y_i}{Y_{i-t}}$
Ланцюговий	$Y_i - Y_{i-1}$	$\frac{Y_i}{Y_{i-1}}$
Середній	$\frac{\Pi_{БАЗ}}{N-1} = \frac{Y_N - Y_1}{N-1}$	$\sqrt[N-1]{T_{P\ БАЗ}} = \sqrt[N-1]{\frac{Y_N}{Y_1}}$

Тут i – номер порівнюваного рівня часового ряду; t – число періодів (моментів) часу на які база порівняння відстає від Y_i .

Так, знайдені на стадії попереднього аналізу середні показники динаміки можуть бути використані як найпростіші інструменти прогнозування.

1. Точковий прогноз Y_{N+L} на основі середнього абсолютного приросту $\bar{\Pi}$ (табл. 1):

$$Y_{N+L} = Y_N + \bar{\Pi} \times L, \quad (1)$$

де Y_N – останній рівень часового ряду; L – період упередження.

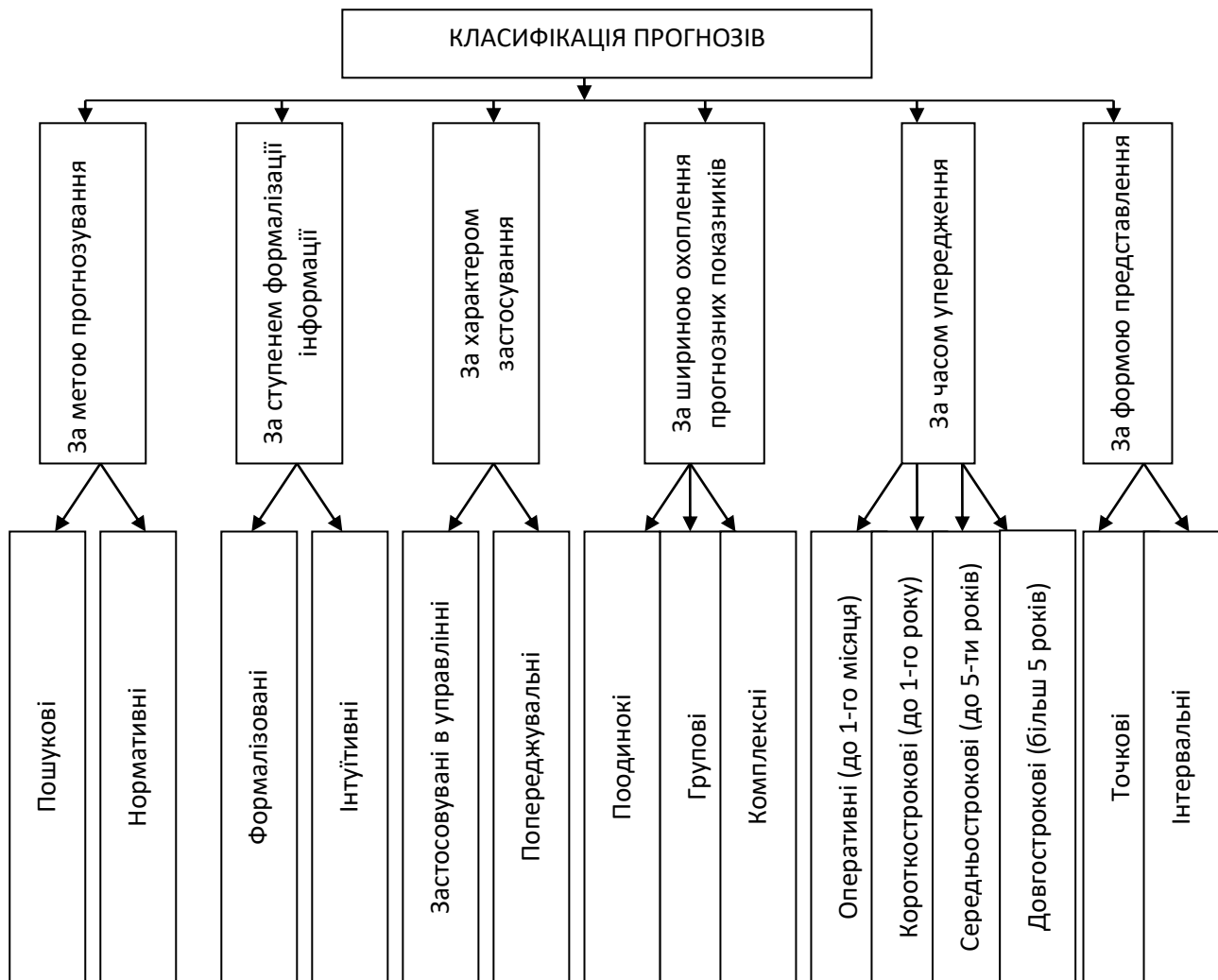


Рис. 2 – Класифікація прогнозів

Формула (1) дає досить точні значення прогнозу досліджуваного показники у разі рівномірного розвитку рівнів ряду динаміки, тобто коли він змінюється в арифметичній прогресії. Перевірка цієї умови зводиться до перевірки приблизної сталості ланцюгових абсолютних приростів ($\Pi_{\text{ЛАН}} \approx \text{const}$).

2. Точковий прогноз на основі середнього темпу зростання $\overline{Tp}$ з табл. 1:

$$Y_{N+L} = Y_N \times \overline{Tp}^L. \quad (2)$$

Формула (2) зазвичай застосовується і дає хороші прогнозні результати у разі рівноприскореного розвитку рівнів ряду динаміки, тобто коли він змінюється в геометричній прогресії. Перевірка цієї умови зводиться до перевірки приблизної сталості ланцюгових темпів зростання ($Tp_{\text{ЛАН}} \approx \text{const}$).

2. ПРОГНОЗУВАННЯ НА ОСНОВІ ТРЕНДОВИХ МОДЕЛЕЙ

Побудовану на етапі аналітичного вирівнювання рівнів часового ряду *трендову* модель, що відображає основну тенденцію розвитку рівнів ряду динаміки у часі X , також часто використовують як інструмент прогнозування майбутніх значень досліджуваного параметра. Точкові прогнози знаходяться на базі рівняння тренду шляхом підстановки замість X майбутнього значення фактора часу $X = N + L$, яке відповідає періоду упередження L . Наприклад, для лінійного тренду $\hat{Y} = a_0 + a_1X$ точковий прогноз має такий загальний вигляд:

$$\hat{Y}_{N+L} = a_0 + a_1(N + L). \quad (3)$$

Використання трендових моделей у якості інструментів прогнозування відкриває можливість отримати поряд із точковим прогнозом також й інтервальний прогноз у вигляді двох чисел – нижньої та верхньої меж довірчого інтервалу із заздалегідь заданою достовірністю P потрапляння в нього фактичного значення Y_{N+L} . Межі довірчого інтервалу прогнозу розраховуються за такою загальною схемою:

$$\hat{Y}_{N+L} \pm \Delta, \quad (4)$$

де Δ – гранична похибка прогнозу.

Формула (4) означає, що в центрі довірчого інтервалу знаходиться точковий прогноз $\hat{Y}_{N+L}$, а ширина самого довірчого інтервалу дорівнює 2Δ . В результаті отримання довірчого інтервалу за схемою (4) точність прогнозу падає, оскільки розширюються його межі, але при цьому забезпечується наперед задана достовірність прогнозу P .

Виникає питання: як розрахувати величину граничної помилки прогнозу? Теорія статистичного моделювання та прогнозування дає на нього наступну відповідь. Для лінійного тренду формула граничної похибки прогнозу з достовірністю $P = (1 - \alpha) \times 100$ % попадання в довірчий інтервал майбутнього фактичного значення прогнозного параметру має такий вигляд:

$$\Delta = t_{\alpha;k} \times S_Y \times \sqrt{\frac{1}{N} + \frac{(N+L-\bar{X})^2}{\sum_{i=1}^N (X_i - \bar{X})^2}}, \quad (5)$$

де $t_{\alpha;k}$ – коефіцієнт довіри (α -квантиль t -розподілу Стюдента з рівнем значущості α і $k = N - m$ ступенями вільності; m – число оцінюваних коефіцієнтів тренда); S_Y – стандартна помилка тренду; $\bar{X}$ – середнє значення фактора часу.

Слід зазначити, що рівень значущості α по суті є вірогідністю помилки і визначається, виходячи з необхідної і заздалегідь заданої достовірності P (в %) з наступного співвідношення: $\alpha = 1 - P/100$. Наприклад, за необхідної

достовірності попадання в довірчий інтервал майбутнього фактичного значення Y_{N+L} $P = 95\%$ $\alpha = 0,05$; при $P = 90\% = 0,10$; при $P = 99\% = 0,01$ і т.д.

Число ступенів свободи $k = N - m$ для лінійного тренду ($m = 2$) дорівнює $N - 2$. А величина коефіцієнта довіри (α -квантиль t -розподілу Стюдента) знаходиться за допомогою редактора *Excel* (команда = *СТЬЮДРАСПОБР*(α ; $N - m$)) – *Enter*).

Стандартна помилка тренду S_Y визначається за формулою

$$S_Y = \sqrt{\frac{\sum_{i=1}^N (Y_i - \hat{Y}_i)^2}{N - m}} \quad (6)$$

і розраховується автоматично у процесі побудови моделі тренда за допомогою стандартної програми «Регресія» редактора *Excel*). Вона характеризує абсолютну точність знайденого рівняння тренду: тим менше S_Y , тим точніший тренд, і навпаки.

Третій співмножник (радикал) у виразі (5) відображає обрану форму математичної залежності між рівнями ряду динаміки та часом X . В даному випадку для лінійного тренду вираз під коренем не надто складний, але при переході до криволінійних функцій зв'язку він суттєво ускладнюється.

Так, при параболічній (2-го ступеня) формі тренда третій співмножник у формулі (5) набуває вигляд:

$$\sqrt{\frac{(N+L)^2 + \frac{\sum_{i=1}^N X_i^4 - (2\sum_{i=1}^N X_i^2)(N+L)^2 + N(N+L)^4}{\sum_{i=1}^N X_i^2 - \frac{N^2 \sum_{i=1}^N X_i^2 - (\sum_{i=1}^N X_i)^2}{N}}}{\sum_{i=1}^N X_i^2 - \frac{N^2 \sum_{i=1}^N X_i^2 - (\sum_{i=1}^N X_i)^2}{N}}}. \quad (7)$$

Слід пам'ятати, що автоматичний розрахунок величини граничної помилки прогнозу Δ у стандартних програмах кореляційно-регресійного аналізу редактора *Excel* не передбачено. А вручну здійснити вказані розрахунки за великих m практично неможливо. Тому у реальних дослідженнях при прогнозуванні з цією метою користуються системою програм *STATISTICA*.

При прогнозуванні майбутніх значень рівнів досліджуваного ряду економічної динаміки необхідно стежити, щоб серед значень Y_i , що спостерігаються, не було так званих *аномальних* спостережень у вигляді піків або, навпаки, різких падінь рівнів ряду. Справа в тому, що аномальні спостереження зазвичай свідчать про дію якихось випадкових факторів, не характерних для інших періодів (моментів) часу. Тому за їх наявності середня величина Y буде спотворена і, як наслідок, коефіцієнти лінійного тренду a_0 , a_1 міститимуть систематичну помилку. Це зрештою спричинить неправильну прогнозну оцінку, наприклад, сильно завищену чи занижену.

Існують об'єктивні взаємозв'язки між числом оцінюваних коефіцієнтів тренду m і довжиною періоду передісторії N , а також між величиною N та періодом упередження L . Доведено, що число N має у 3-4 рази перевищувати кількість оцінюваних коефіцієнтів тренду m . Для лінійного рівняння $m = 2$, тому, щоб побудувати тренд $\hat{Y} = a_0 + a_1 X$, необхідно мати щонайменше 6-8

рівнів часового ряду. А при безпосередньому прогнозуванні має виконуватися умова $L \leq N/3$, тобто довжина періоду упередження не повинна перевищувати третини довжини періоду передісторії. Порушення зазначених співвідношень призводить до отримання ненадійних трендових моделей та, відповідно, до помилкових прогнозних оцінок.

У якості прикладу розглянемо прогнозування на 3 моменту часу ($L = 3$) одного з параметрів хімічного агрегату на основі часового ряду наступних 12 спостережень ($N = 12$), що наведені в табл. 2.

Таблиця 2 – Значення досліджуваного параметру, одиниць

Момент часу X_i	1	2	3	4	5	6	7	8	9	10	11	12
Параметр Y_i	681	615	592	511	487	499	452	436	380	337	334	321

На базі даних табл. 2 знайдемо ланцюгові, базисні та середні абсолютні прирости і темпи зростання рівня досліджуваного параметру (див. табл. 3 розрахунки під нею).

Таблиця 3 – Важливіші характеристики ряду динаміки

Момент часу X_i	Параметр Y_i	Абсолютні прирости Π		Темпи росту T_p	
		лан.	баз.	лан.	баз.
1	681	-	-	-	1
2	615	-66	-66	0,90308	0,90308
3	592	-23	-89	0,96260	0,86931
4	511	-81	-170	0,86318	0,75037
5	487	-24	-194	0,95303	0,71512
6	499	12	-182	1,02464	0,73275
7	452	-47	-229	0,90581	0,66373
8	436	-16	-245	0,96460	0,64023
9	380	-56	-301	0,87156	0,55800
10	337	-43	-344	0,88684	0,49486
11	334	-3	-347	0,99110	0,49046
12	321	-13	-360	0,96108	0,47137

Середній абсолютний приріст:

$$\overline{\Pi} = \frac{\Pi_{БАЗ}}{N-1} = \frac{Y_N - Y_1}{N-1} = \frac{321 - 681}{12-1} = \frac{-360}{11} = -32,727;$$

Середній темп зростання:

$$\overline{T_p} = \sqrt[N-1]{T_{p\ БАЗ}} = \sqrt[N-1]{\frac{Y_N}{Y_1}} = \sqrt[12-1]{\frac{321}{681}} = \sqrt[11]{0,471} = 0,934.$$

Здійснимо точкове прогнозування досліджуваного параметру на 3 моменту часу ($L = 3$) за допомогою середнього абсолютного приросту $\overline{\Pi}$ на основі формули (1):

$$Y_{N+L} = Y_N + \overline{\Pi} \times L = 321 + (-32,727) \times 3 = 222,819.$$

Розрахуємо точковий прогноз досліджуваного параметру на 3 моменту часу ($L = 3$) за допомогою середнього темпу зростання $\overline{Tp}$ на основі формули (2):

$$Y_{N+L} = Y_N \times \overline{Tp}^L = 321 \times (0,934)^3 = 261,545.$$

Побудуємо точковий та інтервальний (з достовірністю 95%) прогнози досліджуваного параметру на 3 моменту часу ($L = 3$) за допомогою лінійного тренду (рис. 3).

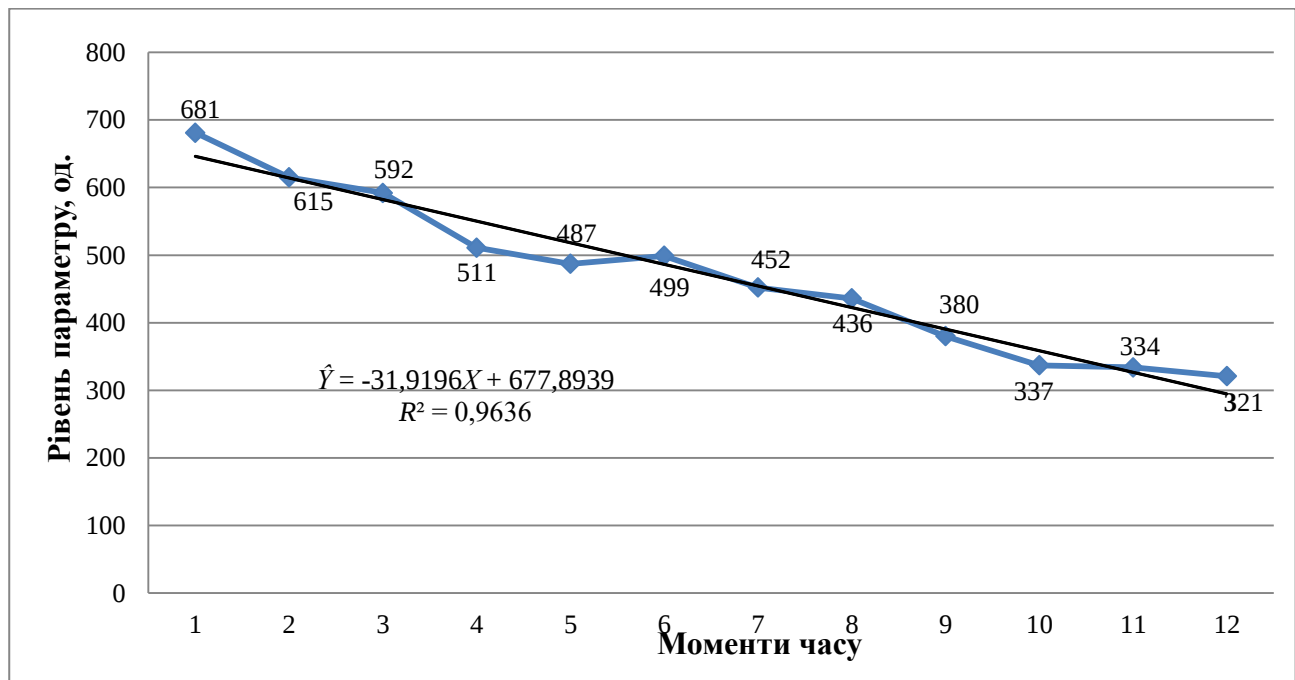


Рис. 3 – Фактичні та вирівняні за лінійним трендом значення досліджуваного параметру, побудовані в редакторі *Excel*

Отриманий лінійний тренд є достатньо точним: коефіцієнт детермінації $R^2 = 0,9636$ указує на те, що більш 96% варіації досліджуваного параметру пояснюється саме його лінійною залежністю від часу.

Точковий прогноз за лінійним трендом знайдемо за формулою (3) :

$$\hat{Y}_{N+L} = a_0 + a_1(N + L) = 677,8939 - 31,9196 \times (12 + 3) = 199,1.$$

95-відсотковий довірчий інтервал прогнозу досліджуваного параметру на 3 моменту часу ($L = 3$) розрахуємо за схемою (4) з використанням системи *STATISTICA*, версія 10 (табл. 4).

Таблиця 4 – Точковий та 95-відсотковий інтервальний прогноз досліджуваного параметру на 3 моменту часу

	b-Weight	Value	b-Weight * Value
Var1	-31,9196	15,00000	-478,794
Intercept			677,894
Predicted			199,100
-95,0%CL			158,982
+95,0%CL			239,218

У табл. 4 точковий прогноз (*Predicted*), нижня межа (*-95,0%CL*) і верхня межа (*+95,0%CL*) наведені в трьох виділених рядках (*CL* – *confidence level* – рівень довіри). На рис. 4 наведено графічне зображення отриманих результатів прогнозування.

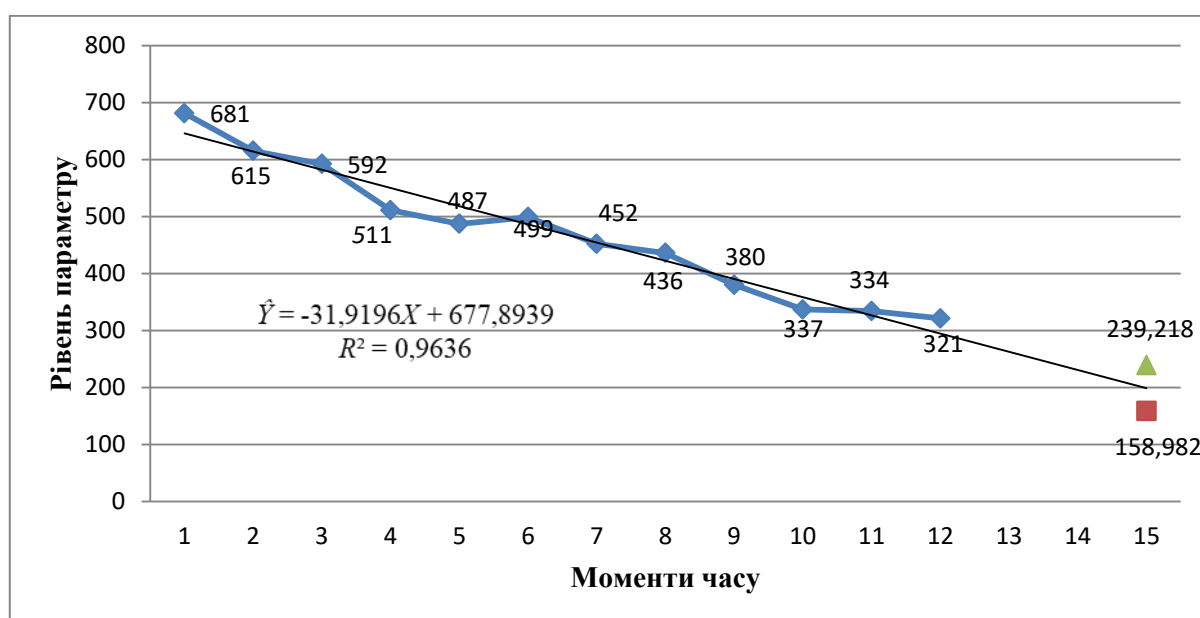


Рис. 4 – Фактичні, вирівняні за лінійним трендом значення досліджуваного параметру та точковий і 95-відсотковий довірчий інтервал прогнозу на 3 моменти часу, побудовані в редакторі *Excel*

Таким чином, якщо закономірності й зв'язки, виявлені в періоді передісторії, збережуться певний час у майбутньому, то досліджуваний параметр хімічного агрегату слід очікувати у 15-й момент часу на рівні 199,1 од. При цьому з достовірністю 95% можна стверджувати, що даний рівень буде знаходитись у межах від 158,98 до 239,22 од.