

Пістунов І.М.

Прогнозування соціально- економічних процесів: навч. наоч. посіб. Дніпро : НТУ «ДП», 2023. 65 с.

В посібнику розглядається методи прогнозування, які базуються на використанні в якості змінних факторів попередні значення змінної, що прогнозується, поточного середнього цієї змінної, поточної похибки, тощо. Описані методи Бокса-Дженкінса Четверікова, експертних оцінок, Фостекра-Стюарта. Подана методика визначення похибки прогнозування та довірчого інтервалу прогнозування.

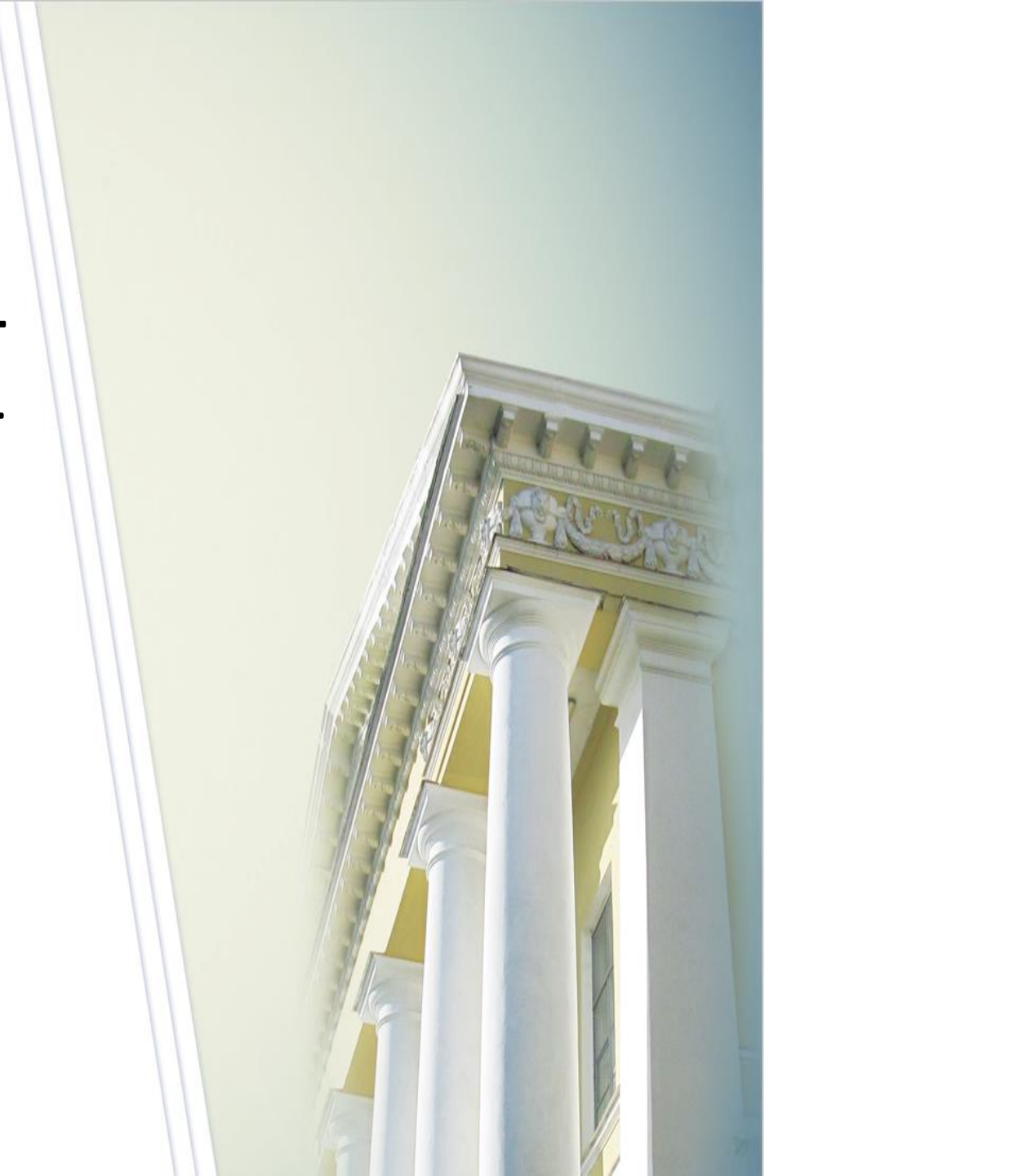
Наведено приклади розрахунків із застосування електронних таблиць Excel/

Призначено для студентів спеціальності 051 «Економіка»

Рецензенти:

Васильєва Н.К., завідувач каф. інформаційних систем ДДАЕУ, проф.

Алексєєв М.О., зав каф. програмних засобів комп'ютерних систем НТУ «ДП», проф.



Структура дисципліни

Тема 1.

Інформаційне уявлення динаміки розвитку соціально-економічних процесів.

Тема 2. Випадкові процеси та часові ряди.

Ідентифікація часових рядів.

Тема 3. Визначення типу нестационарності та ступеня інтеграції часового ряду.

Тема 4. Авторегресійні процеси (AR (p) -процеси)

Тема 5. Ідентифікація детермінованого тренду та сезонності



Метод прогнозування – це сукупність прийомів та способів міркування, які дають можливість на підставі аналізу ретроспективних даних, зовнішніх та внутрішніх зв'язків об'єкта прогнозування, а також їх вимірювань у рамках процесу чи явища, що розглядаються, зробити висновок щодо визначеної вірогідності відносно його (об'єкта) майбутнього розвитку.

Нині, за оцінками вчених, нараховується більше 150 різних методів прогнозування. Однак в якості основних практично використовуються всього 15...20 методів.



Економіко-математична модель — це математична модель економічного явища. Процес моделювання економічного явища передбачає такі етапи:

- формулювання предмета та мети дослідження;
- в економічній системі, що розглядається, виділяються структурні та функціональні елементи, які відповідають сформульованій меті, виявляються найбільш важливі характеристики елементів;
- якісно описуються взаємозв'язки між елементами моделі;
- вводяться символічні позначення для тих характеристик економічного об'єкта, що враховуються та формалізуються;
- наскільки можливо встановлюються взаємозв'язки між цими характеристиками, тобто будується математична модель, при цьому економічний об'єкт відображається у вигляді сукупності рівнянь, нерівностей, логічних відношень, графіків;
- виконуються розрахунки за математичною моделлю та аналіз отриманих результатів.



Моделі, які використовуються для прогнозування, називаються прогнозними моделями. Вони повинні задовольняти таким вимогам:

- **виражатись статистичними категоріями, тобто системою показників, таблицями чи системою угруповань, системою рівнянь, графіками в загальноприйнятих символах і термінах;**

- **відповідати сучасному етапу розвитку статистичної науки, підлягати перевірці на підставі статистичних критеріїв;**

- **для складних соціально-економічних явищ та процесів реалізовуватись на електронно-обчислювальних машинах за стандартними програмами та допускати можливість використання при аналізі додаткових чинникових ознак чи їх вилучення;**

- **будуватись на підставі достатньо великої кількості вірогідних даних, щоб яскраво могли проявлятися дійсні взаємозв'язки, тенденції та закономірності;**

- **розширювати межу пізнання, тобто забезпечувати отримання такої інформації, яку іншими засобами отримати неможливо.**



1. Випадкові процеси та часові ряди.
2. Ідентифікація часових рядів.

1. Інформаційне уявлення динаміки розвитку соціально-економічних процесів

Динамічний ряд - це сукупність спостережень одного показника, впорядкованих в залежності від значень іншого показника, які послідовно зростають або спадають.

Часовий ряд (time series) - це ряд динаміки, упорядкований за часом, або сукупність спостережень економічної величини в різні моменти часу. Складовими ряду спостережень є числові значення показника, які називаються *рівнями ряду*, і *моменти або інтервали часу*, до

яким належать рівні. Часовий ряд (ВР) можна записати в стислому вигляді:

$$Y_t, t=1,2,3, \dots, n$$

де t - рівновіддалені моменти спостережень (година, доба, місяць, квартал, рік і т.п.). Під довжиною часового ряду розуміють час, який минув від першого до останнього моменту спостереження. Часто довжиною ряду називають кількість рівнів n , які утворюють часовий ряд.



Залежно від характеру досліджуваних соціально-економічних показників тимчасові ряди поділяють на **моментні, інтервальні і похідні**. Часові ряди, утворені показниками, які характеризують економічне явище на певні моменти часу, називають **моментними**. Якщо рівні часового ряду утворюються шляхом агрегування за певний проміжок (інтервал) часу, такі ряди називають **інтервальними** часовими рядами. Тимчасові ряди можуть бути створені як з абсолютних значень економічних показників, так і з середніх або відносних величин - це **похідні ряди**.

МОМЕНТНИЙ ЧАСОВИЙ РЯД

ІНТЕРВАЛЬНИЙ ЧАСОВИЙ РЯД

Дата надання позички	01.10.	05.10.	12.10.	23.10.	03.11.	07.11.	Місяць	Січень	Лютий	Березень	Квітень	Травень
Розмір наданої позички, тис. грош. од.	3747	3710	3839	3783	3747	3710	Валовий внутрішній продукт, млн грн	6578	7016	7353	7353	7941



Характеристики динаміки розвитку соціально-економічних процесів і їх розрахунки

Характеристики	Розрахункові формули
1. Абсолютний приріст	$\Delta y_i = y_i - y_{i-k}$
2. Коефіцієнт зростання	$K_{i(зр)} = \frac{y_i}{y_{i-k}}$
3. Коефіцієнт приросту	$K_{i(пр)} = \frac{y_i - y_{i-k}}{y_{i-k}}$
4. Темп зростання%	$T_{i(зр)} = \frac{y_i}{y_{i-k}} * 100\%$
5. Темп приросту%	$T_{i(пр)} = \frac{y_i - y_{i-k}}{y_{i-k}} * 100\%$
6. Середній абсолютний приріст	$\overline{\Delta y_i} = \frac{y_i - y_{i-k}}{k}$
7. Середній темп росту	$\overline{T}_{(зр)} = \sqrt[n-1]{\frac{y_n}{y_1}}$
8. Середній темп приросту%	$\overline{T}_{(пр)} = \sqrt[n-1]{\frac{y_n}{y_1}} * 100\%$

Для отримання практичних оцінок часових рядів користуються такими формулами:

математичне сподівання:

$$\hat{\mu}_y = \frac{\sum_{t=1}^n y_t}{n} ;$$

дисперсія:

$$\hat{\sigma}_y^2 = \hat{\gamma}_0 = \frac{\sum_{t=1}^n (y_t - \hat{\mu}_y)^2}{n} ;$$

автоковаріація порядку τ :

$$\hat{\gamma}_\tau = \frac{\sum_{t=1}^{n-\tau} (y_t - \hat{\mu}_y)(y_{t+\tau} - \hat{\mu}_y)}{n - \tau} .$$

Зрушення в часі τ називають *часовим лагом*



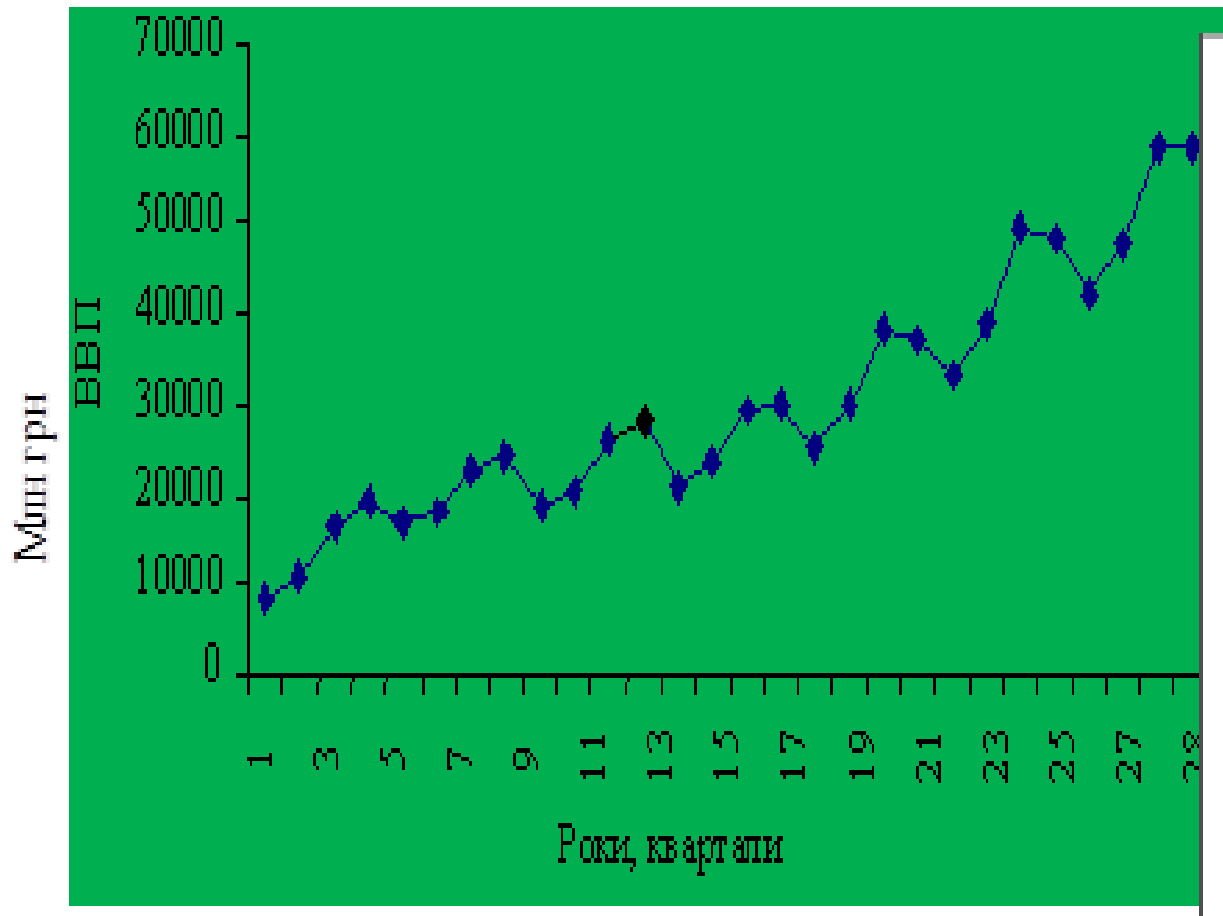


Рис. 2. Нестационарний часовий ряд



Рис. 3. Часовий ряд доходів консолідованого бюджету, виражений у відсотках в до ВВП



Прогнозування часового ряду методом ковзної середньої

для лінійної інтерполяції

$$y_t^* = \frac{\sum_{i=t-p}^{t+p} y_i}{m}, \quad t > p.$$

для нелінійної інтерполяції

$$y_t^* = \frac{\sum_{i=t-p}^{t+p} p_i y_i}{\sum_{i=t-p}^{t+p} p_i},$$

причому вага p_i визначається за допомогою методу найменших квадратів. Ця вага обчислюється для різних степенів поліному апроксимації і різних інтервалів згладжування. Так, для поліномів другого та третього степенів чисельна послідовність ваги при інтервалі згладжування $m = 5$ має значення $[-3, 12, 17, 12, -3]$, а при $m = 7$ — $[-2, 3, 6, 7, 6, 3, -2]$. Для поліномів четвертої та п'ятої степенів при $m = 7$ послідовність ваги буде $[5, -30, 75, 131, 75, -30, 5]$.



Експоненційне згладжування

основні формули

значення коефіцієнту α обирається в межах 0,1 - 0,3

В окремих випадках Браун запропонував визначати величину параметра згладжування залежно від довжини згладжування ряду за формулою

$$y_t^* = \alpha y_t + (1 - \alpha) y_{t-1}^*, \quad \alpha = \frac{2}{n+1}.$$

Відома також для визначення величини α формула Мейєра вигляду

$$y_t^* = \alpha \sum_{i=0}^{t-1} (1 - \alpha)^i y_{t-i} + (1 - \alpha)^t y_0, \quad \alpha \approx \frac{\sigma_n}{\sigma_\varepsilon},$$

$y_0 = \frac{y_1 + y_2 + y_3}{3}$ де $\sigma_n, \sigma_\varepsilon$ – середньоквадратичне відхилення моделі та вихідного ряду відповідно.

Модель Хольта-Брауна

Етап 1.

За першими п'ятьма точками часового ряду оцінюються початкові значення A_0 та A_1 , параметрів моделі за допомогою методу найменших квадратів для лінійної апроксимації:

$$Y_p(t) = A_0 + A_1 t$$

Етап 2.

За моделлю Брауна знаходимо прогноз на один крок ($k = 1$):

$$Y_p(t, k) = A_0(t) + A_1(t)k$$

Етап 3.

Розрахункове значення економічного показника порівнюють із фактичним і обраховується величина їх розбіжності (помилки). При $k = 1$ маємо:

$$e(t + 1) = Y(t + 1) - Y_p(t, 1)$$

Етап 4.

Відповідно до цієї величини коректуються параметри моделі.
У моделі Брауна модифікація здійснюється в такий спосіб:

$$A_0(t) = A_0(t-1) + A_1(t-1) + (1-\beta)^2 e(t);$$

$$A_1(t) = A_1(t-1) + (1-\beta)^2 e(t),$$

де β – коефіцієнт дисконтування даних, що змінюється в межах від 0 до 1, який характеризує знецінення даних за одиницю часу, а також відображає ступінь довіри більш пізнім спостереженням.

Оптимальне значення β знаходиться ітеративним шляхом, тобто багаторазовою побудовою моделі при різних β і вибором найкращої, або за формулою:

$$\beta = \frac{N-3}{N-1}$$

де N – довжина часового ряду. $e(t)$ – помилка прогнозування рівня $Y(t)$, обчислена в момент часу $(t-1)$ на один крок уперед.

Етап 5.

За моделлю зі скоректованими параметрами $\underline{A_0}$ і $\underline{A_1}$ знаходять прогноз на наступний момент часу.

Етап 6.

Повертаємося в пункт 3, якщо $t < N$.

Якщо $t = N$, то побудовану модель можна використовувати для прогнозування на майбутнє.

Методи Бокса-Дженкінса

AR(p) -авторегресивна модель порядку p

Модель має вигляд:

$$Y(t) = f_0 + f_1 \cdot Y(t-1) + f_2 \cdot Y(t-2) + \dots + f_p \cdot Y(t-p) + E(t)$$

де $Y(t)$ –залежна змінна у момент часу t . $f_0, f_1, f_2, \dots, f_p$ - оцінювані параметри. $E(t)$ - помилка від впливу змінних, які не враховуються в даній моделі. Завдання полягає в тому, щоб визначити $f_0, f_1, f_2, \dots, f_p$. Їх можна оцінити різними способами. Найправильніше шукати їх через систему рівнянь Юла-Уолкера, для складання цієї системи буде потрібно розрахунок значень автокореляційної функції. Можна поступити простішим способом - порахувати їх методом найменших квадратів.

MA(q) -модель з ковзаючим середнім порядку q

Модель має вигляд:

$$Y(t) = m + e(t) - w_1 \cdot e(t-1) - w_2 \cdot e(t-2) - \dots - w_p \cdot e(t-p)$$

Де $Y(t)$ -залежна змінна у момент часу t . $w_0, w_1, w_2, \dots, w_p$ - оцінювані параметри.

Авторегресійне ковзне середнє ARMA(p,q)

Під позначенням ARMA(p,q) [3] розуміється модель, р авторегресійних складових, що містить q, ковзаючих середніх.

Точніше модель ARMA(p,q) включає моделі AR(p) і MA(q):

$$X_t = c + e_t + \sum_{i=1}^q \theta_i e_{t-i} + \sum_{i=1}^p \phi_i X_{t-i},$$

Зазвичай значення помилки e_t вважають незалежними однаково розподіленими випадковими величинами, узятими з нормального розподілу з нульовим середнім: $e_t \sim N(0, \sigma^2)$, де σ^2 — дисперсія. Припущення можна ослабити, але це може привести до зміни властивостей моделі. Наприклад, якщо не припускати незалежності і однакового розподілу помилок, поведінка моделі суттєво міняється.

ARIMA (p,d,q)

У завданні аналізу тимчасового ряду з складною структурою часто використовуються моделі класу ARIMA(p,d,q)[2] (авторегресійне інтегрування ковзаючого середнього - Autoregressive Integrated Moving Average) порядку (p,d,q), які моделюють різні ситуації, що зустрічаються при аналізі стаціонарних і нестаціонарних рядів. Залежно від аналізованого ряду модель ARIMA (p,d,q) може трансформуватися до авторегресійної моделі AR(p), моделі ковзного середнього MA(q) або змішаній моделі ARMA (p,q). При переході від нестаціонарного ряду до стаціонарного значення параметра d, що визначає порядок різниці, приймається рівним 0 або 1, тобто цей параметр має тільки цілочисельні значення. Зазвичай обмежуються вибором між $d = 0$ і $d = 1$. Проте з поля зору дослідників випадає ситуація, коли параметр d може приймати дробові значення.

ARFIMA(p,d,q)

Для ситуації розгляду дробових значень порядку різниці, в роботах зарубіжних учених, в першу чергу, C.W.Granger, J.R.Hosking, P.M.Robinson, R. Beran, був запропонований новий клас моделей ARFIMA(p,d,q)[2] (F: fractional - дріб), що допускає можливість нецілого параметра d і авторегресійний дріб інтегрований процес ковзного середнього. Такі ряди володіють своєю специфікою: самоподібністю, дробовою розмірністю, поволі спадаючою кореляцією. Прогнозування тимчасових рядів за допомогою моделі ARFIMA(p,d,q) відкриває ширші перспективи для підвищення точності прогнозу.

Модель вигляду ARIMA (p,d,q)(P,D,Q)S

ARIMA (p,d,q)(P,D,Q)S [1],

де: p - авторегресійні доданки;

D - різниці;

q - доданки ковзаючого середнього;

P – сезонні авторегресійні доданки;

D – сезонні різниці на інтервалі S ;

Q – доданки сезонного ковзаючого середнього

Розрахунок довірчих інтервалів прогнозу

n - довжина часового ряду;

L - період випередження;

Y_{n+L} - точковий прогноз на момент $n + L$;

t_α - значення t -статистики Стюдента;

S_y - середня квадратична помилка ряду;

t - порядковий номер рівнів ряду, $t = 1, 2, \dots, n$.

Тоді, для поліному першого порядку, довірчий інтервал:

$$\text{MSE} = \frac{1}{n} \sum_{i=1}^n (Y_i - \hat{Y}_i)^2.$$

$$\hat{Y}_{n+L} \pm t_\alpha \cdot S_y \sqrt{\frac{n+1}{n} + \frac{(t_1 - \bar{t})^2}{\sum_{t=1}^n (t - \bar{t})^2}}$$

Для поліному другого порядку, довірчий інтервал:

$$\hat{y}_{n+1} \pm t_{\alpha} \cdot S_y \sqrt{1 + \frac{t_1^2}{\sum t_i^2} + \frac{\sum t_i^4 - (2 \sum t_i^3) t_1^2 + n t_1^4}{n \sum t_i^4 - (\sum t_i^2)^2}}$$



Розкладення (декомпозиція) часового ряду.

$$Y_t = f(t) + \varepsilon_t,$$

де $f(t)$ — не випадкова функція часу (детермінована частина);

ε_t — випадкова, недетермінована частина.

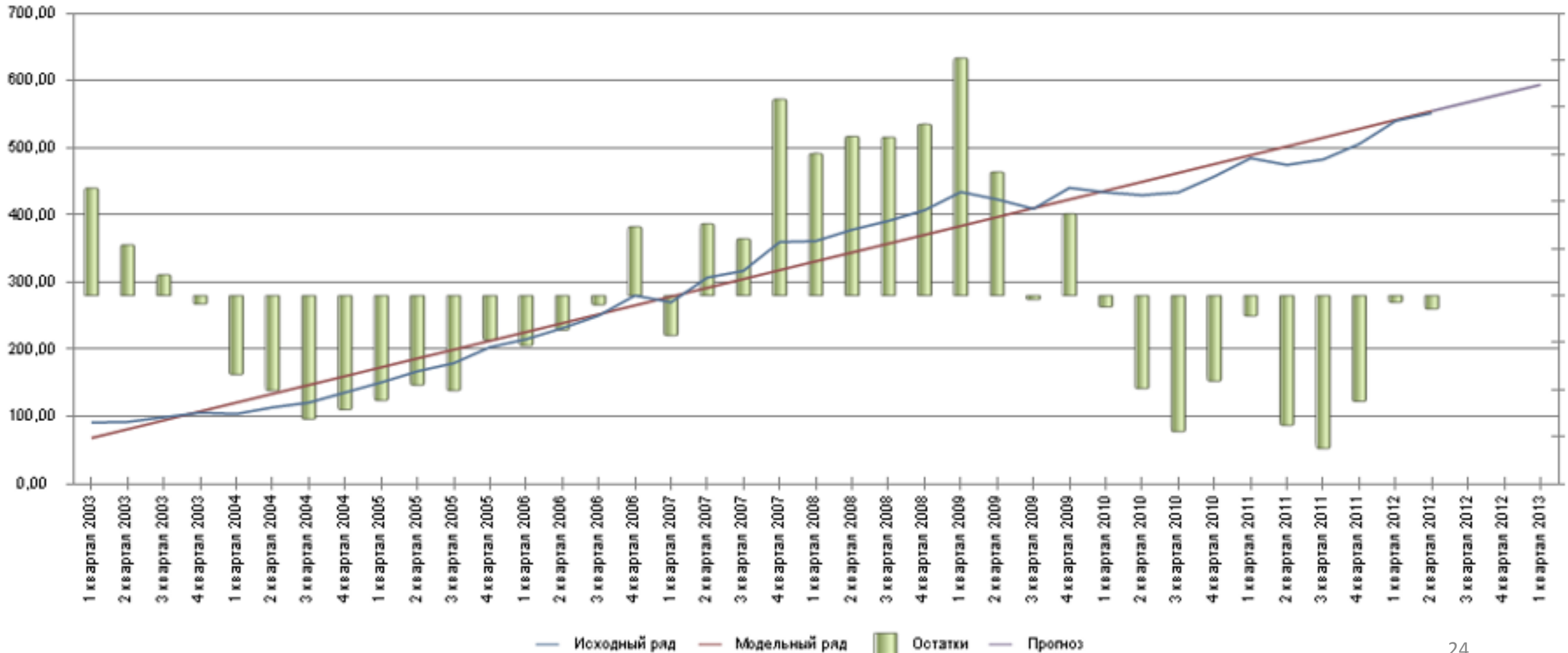
Еволюційні чинники визначають загальний напрям розвитку економічного показника, провідну його тенденцію.

Тенденція — це не випадкова складова часового ряду, яка змінюється повільно, і описується за допомогою певної функції V_t , яку називають функцією тренду або просто трендом. *Тренд відображає вплив на економічний показник деяких постійних чинників, дія яких акумулюється в часі.*

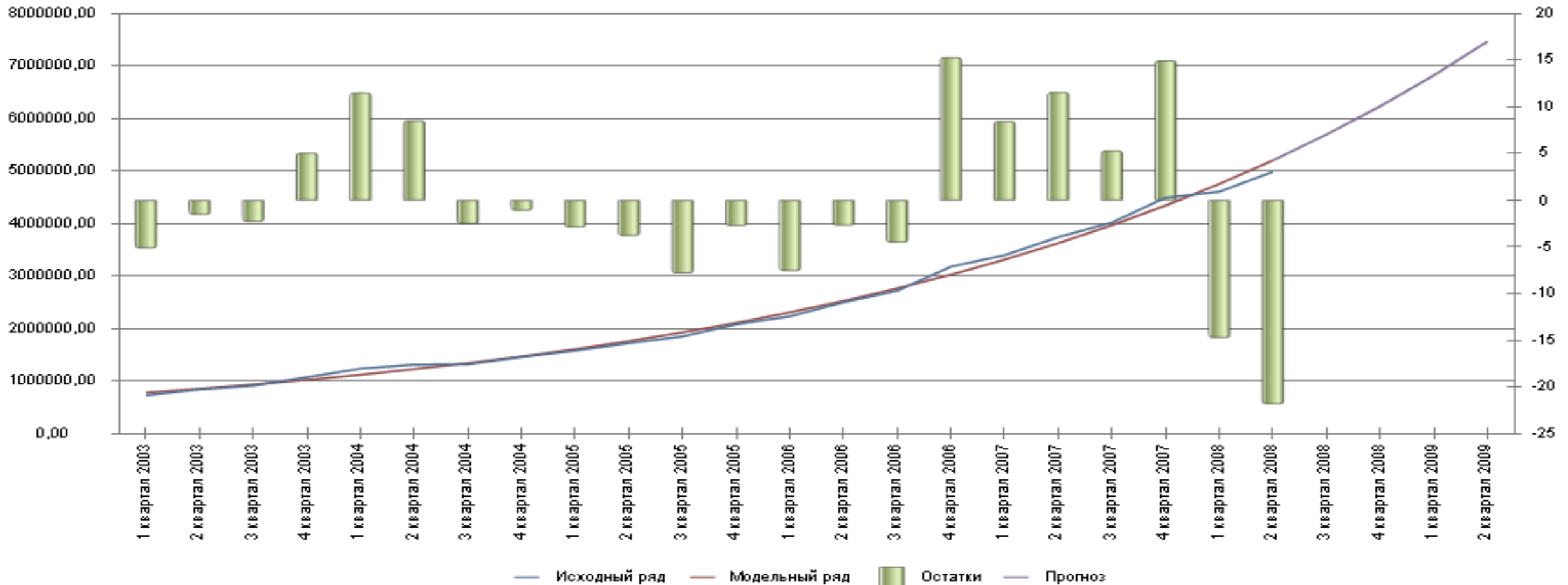
ПРИКЛАДИ ТРЕНДІВ

лінійний	$y_t = a + b \cdot t$
експоненціальний	$y_t = a \cdot k^t$
гіперболічний	$y_t = a + b/t$
степеневий	$y_t = a \cdot t^b$
поліноміальний	$y_t = a + b_1 \cdot t + b_2 \cdot t^2 + \dots + b_m \cdot t^m$
логіарифмічний	$y_t = a + b \cdot \log t$
логістичний	$y_t = \frac{1}{e^{a+b \cdot t} + 1}; y_t = \frac{y_{max} - y_{min}}{e^{a+b \cdot t} - 1} + y_{min}$

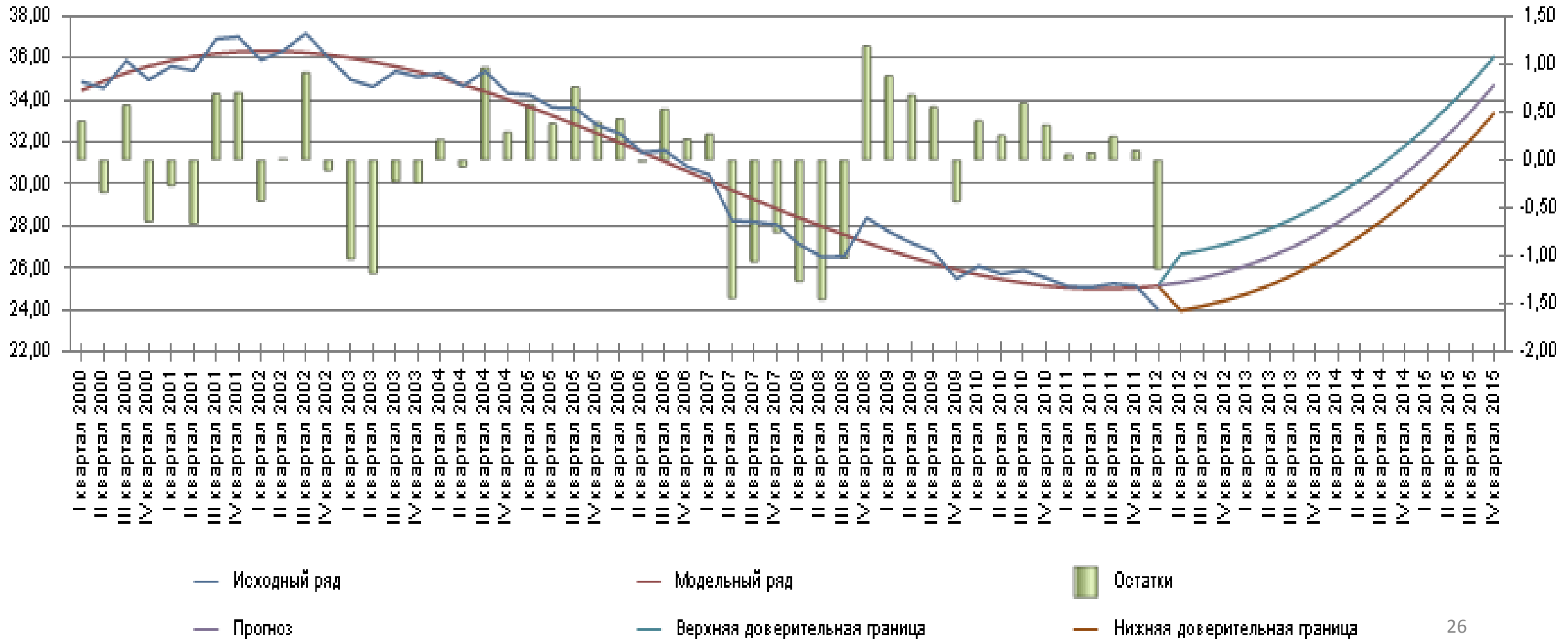
Лінійний тип тренду підходить для відображення тенденцій приблизно рівномірного зміни рівнів: рівних в середньому величинах абсолютного збільшення або абсолютного скорочення рівнів за рівних проміжків часу



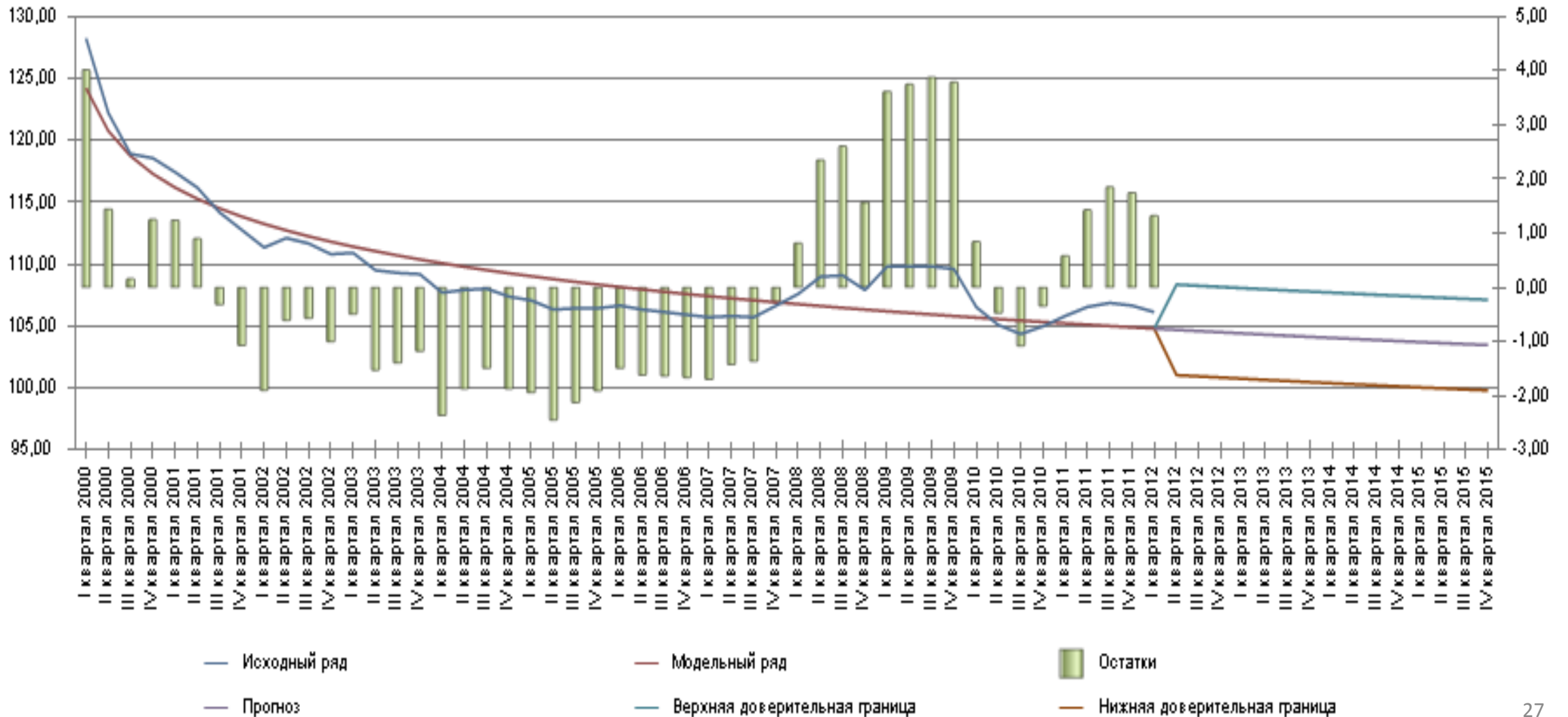
Експонентний тренд характерний процесам, що розвиваються в середовищі, що не створює ніяких обмежень для зростання рівнів.



Поліноміальний тренд описує дані, плавно змінюються в різних напрямках. При використанні поліноміального тренду користувачеві завжди необхідно задати порядок полінома



Рівняння логарифмічного тренду застосовують в тому випадку, коли досліджуваний процес призводить до уповільнення зростання показника, але при цьому зростання не припиняється, а прагне до якогось обмеженого межі. В цьому випадку гіперболічна форма тренду або парабола з негативним прискоренням не підходять. У логарифмічному тренді величини прискорення абсолютних змін мають знак, протилежний знаку самих абсолютних змін, а по модулю поступово зменшуються.



Спектральний аналіз

Ряд Фур'є — в математиці — спосіб представлення довільної складної функції сумою простіших. В загальному випадку кількість таких функцій може бути нескінченною, при цьому чим більше таких функцій враховується при розрахунку, тим вищою стає кінцева точність представлення даної функції. В більшості випадків в якості найпростіших використовуються тригонометричні функції синуса і косинуса. В цьому випадку ряд Фур'є називається *тригонометричним*, а обчислення такого ряду часто називають *розкладом на гармоніки*.



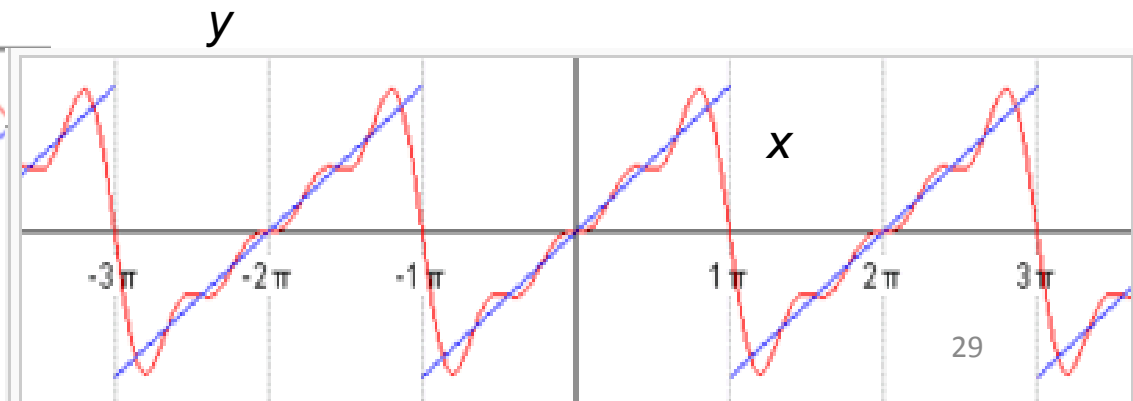
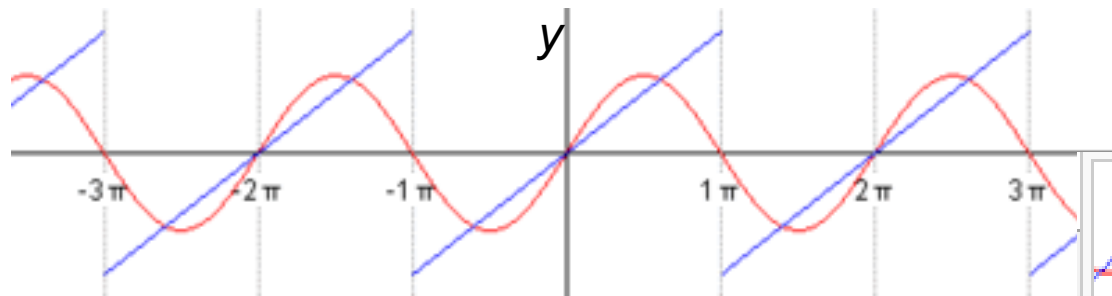
Тригонометричним рядом Фур'є називають ряд виду

$$y = \frac{a_0}{2} + \sum_{n=1}^{\infty} [a_n \cos(nx) + b_n \sin(nx)]$$

Сталі числа a_0, a_n, b_n ($n \in \mathbb{N}$) називаються *коефіцієнтами тригонометричного ряду* і знаходяться як

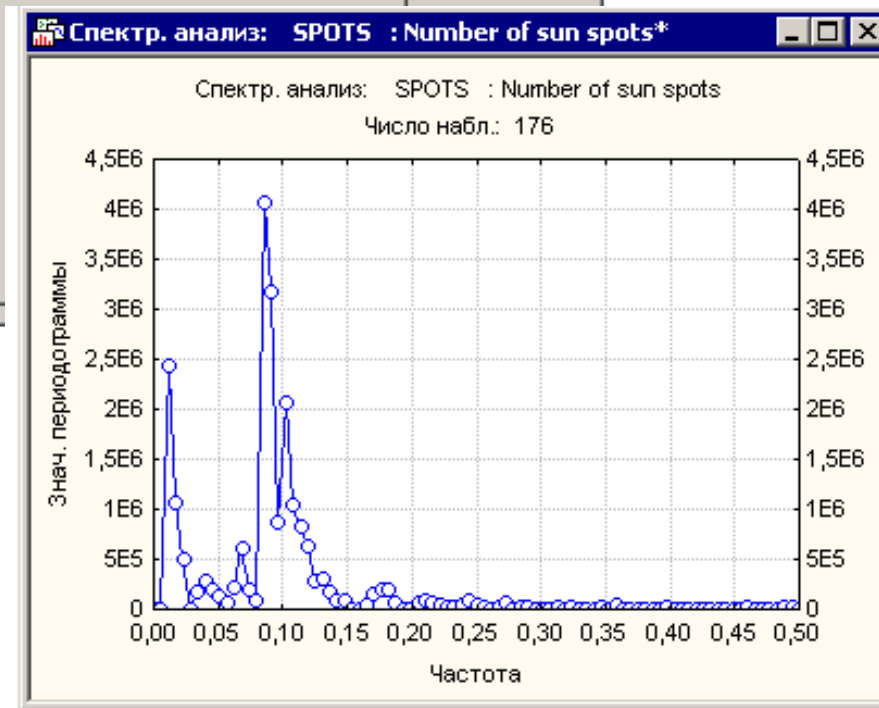
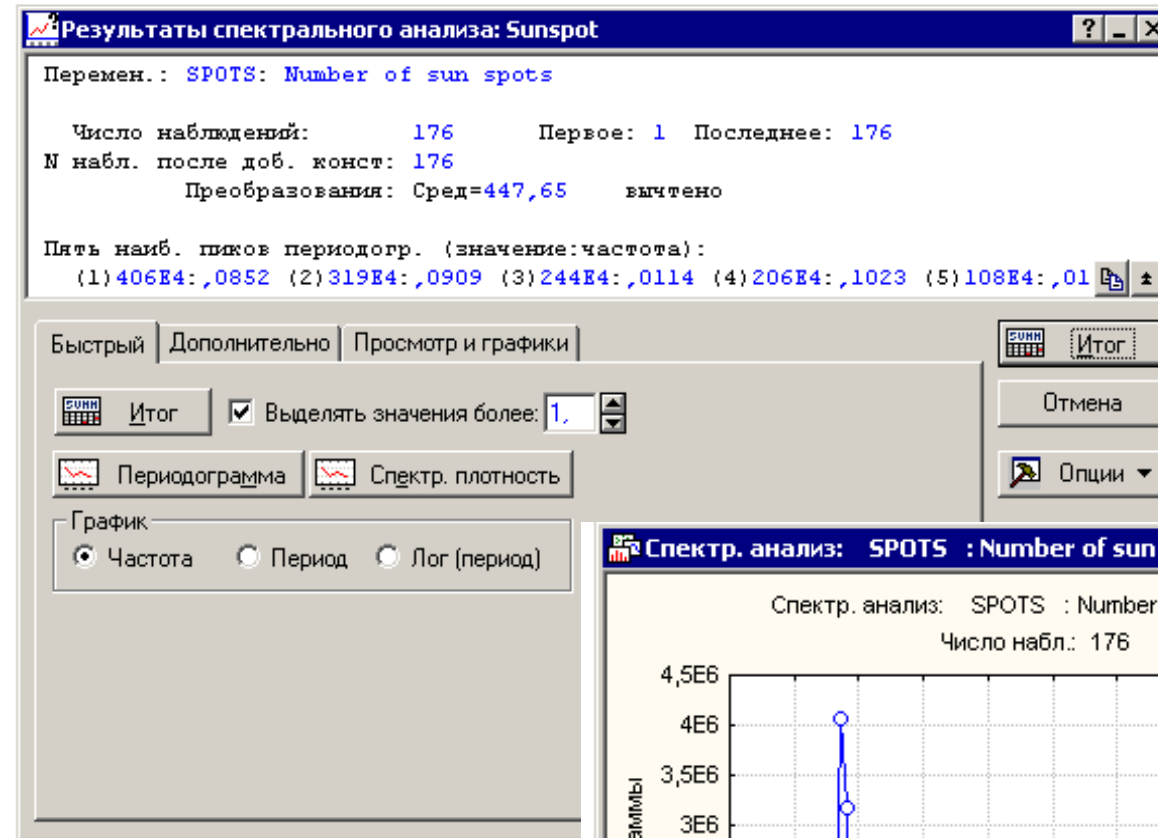
$$a_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx, \quad b_n = \frac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx.$$

Збіжність ряду Фур'є

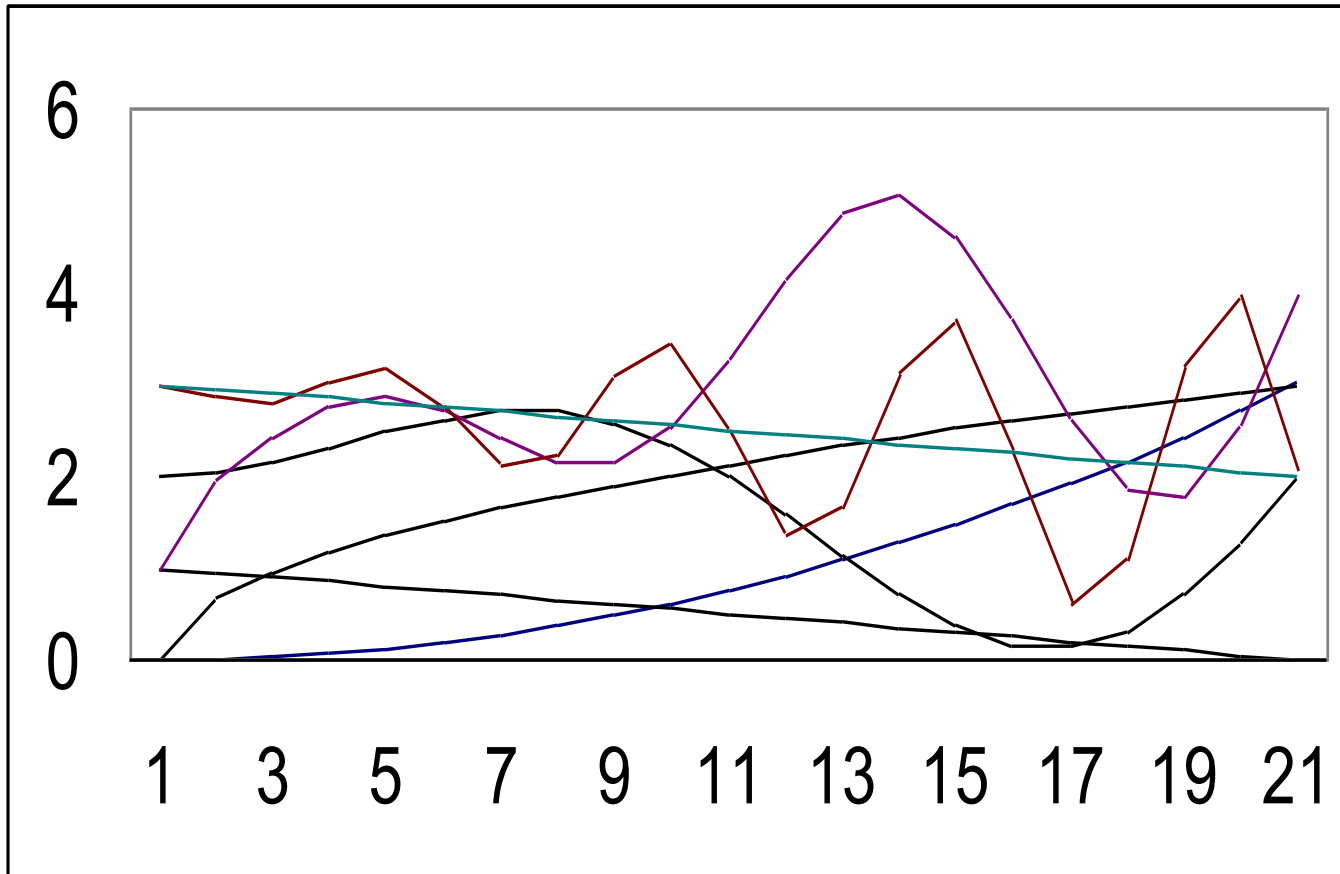


ПОБУДОВА АМПЛІТУДНО-ЧАСТОТНОЇ ХАРАКТЕРИСТИКИ ПРОЦЕСУ ПАКЕТОМ STATISTICA

Data: Sunspot*...	
Wolfer's sunspot data	
1 SPOTS	
1749	809
1750	834
1751	477
1752	478
1753	307
1754	122
1755	96
1756	102
1757	324
1758	476
1759	540
1760	629
1761	859
1762	612
1763	451
1764	364
1765	209
1766	114



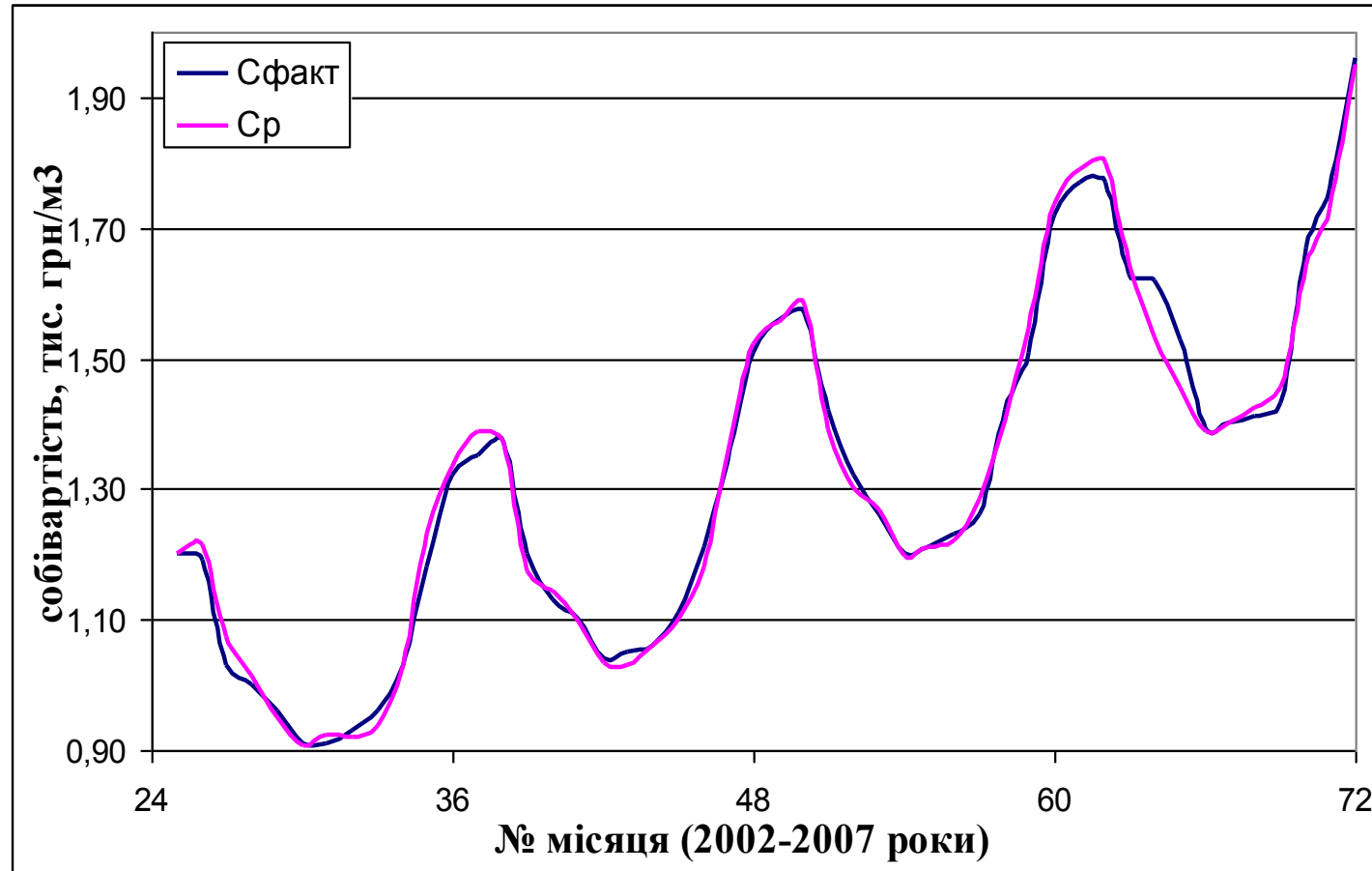
Прогноз періодичних моделей

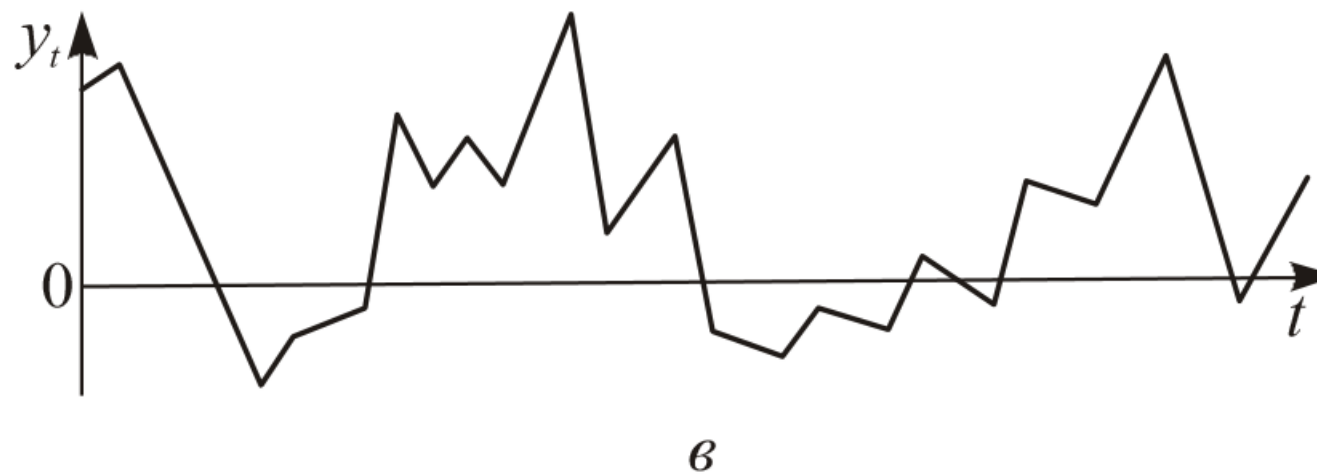
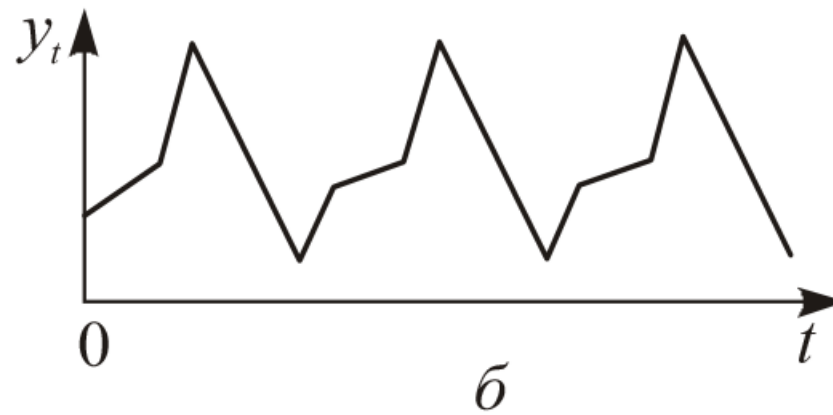
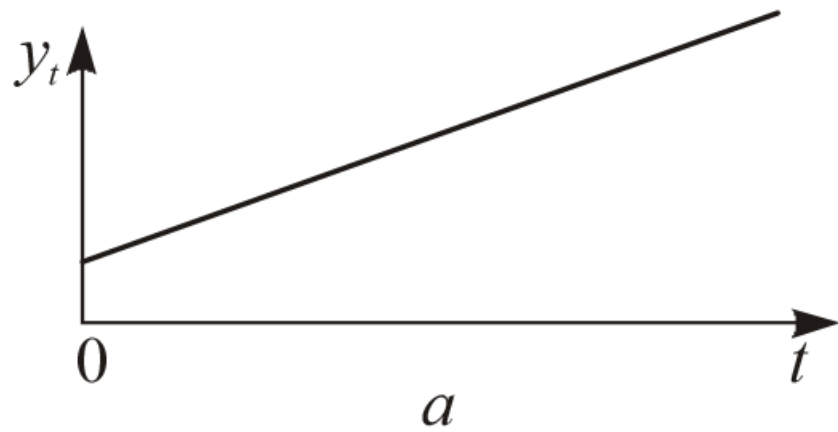


$$y = Ax^B + C(1 - e^{Dx})$$
$$\text{Sin}(Ex^F + G) + H$$

Приклад періодичної моделі

$$C_{t,1} = 1,342c_{t-12,1}^{0,887} + 0,525(1 - e^{-0,976c_{t-12,1}}) \sin(0,524c_{t-12,1}^{1,187} + 0,664) - 0,402c_{t-24,1}^{0,686} + 0,288(1 - e^{-0,106c_{t-24,1}}) \sin(0,524c_{t-24,1}^{0,808} + 0,195) - 0,146$$





Мал. 1. Головні компоненти часового ряду: а - тренд, який зростає; б - сезонна компонента; в - випадкова компонента



Загальне розкладання часового ряду в адитивну модель

$$y_t = v_t + s_t + c_t + \varepsilon_t$$

модель тренду

$$y_t = v_t + \varepsilon_t \quad t = 1, 2, \dots, n$$

модель сезонності

$$y_t = s_t + \varepsilon_t \quad t = 1, 2, \dots, n$$

тренд-сезонна модель

$$y_t = v_t + s_t + \varepsilon_t \quad t = 1, 2, \dots, n$$

Іще бувають: мультиплікативна модель часового ряду:

$$y_t = v_t * s_t * c_t *$$

• комбінована модель часового ряду:

$$y_t = v_t * s_t * c_t +$$

Тренд U_t , сезонну V_t , та циклічну C_t компоненти називають регулярними чи систематичними компонентами часового ряду. Складова частина часового ряду, що залишилась після вилучення із нього регулярних компонент, становить собою випадкову, нерегулярну компоненту e_t . Тому в загальному випадку часовий ряд y_t , $t = 1, 2, \dots, n$ можна записати у вигляді суми

$$y_t = U_t + V_t + C_t + e_t$$



Виявлення наявності в часовому ряду сезонних коливань зводиться до перевірки на випадковість залишкового ряду l_t ,

$$l_t = y_t - U_t$$

або шляхом побудови та аналізу графіка вихідного ряду.

Для вилучення сезонної компоненти найчастіше використовуються ітераційні методи.

Основна ідея ітераційних процедур полягає в багатократно-му використанні ковзної середньої

$$y_t^* = \frac{\frac{y_{t-T_0/2}}{2} + y_{t-T_0/2+1} + \dots + y_t + \dots + y_{t+T_0/2} + \frac{y_{t+T_0/2}}{2}}{T_0}$$

і одночасній оцінці сезонної компоненти в кожному циклі. При цьому перехід від одного циклу ітераційної процедури до другого може супроводжуватись зміною параметрів ковзної середньої³⁶

Метод Четверикова

1. Емпіричний ряд y_t згладжується ковзною середньою з періодом T_0 , тобто враховується $(T_0 + 1)$ рівнів вихідного ряду, із яких перший та останній використовуються з вагою 0,5. При цьому $T_0/2$ рівнів з обох кінців ряду, що втрапились, або відновлюються екстраполяцією вирівняного ряду, або вилучаються з подальшої обробки. Як наслідок будемо мати попередню оцінку тренда

$$y_t^{(1)} = U_t^{(1)}$$

та перше відхилення вихідного ряду від згладженого

$$l_t^{(1)} = y_t - y_t^{(1)}$$

2. Знаходиться середньоквадратичне відхилення σ_1

$$\sigma_1 = \sqrt{\frac{\sum_{t=1}^{T_0} I_{1t}^2 - \left(\sum_{t=1}^{T_0} I_{1t}\right)^2 / T_0}{T_0 - 1}},$$

на яке далі діляться окремі місячні (квартальні) відхилення

$$I_t^{(1)*} = \frac{I_t^{(1)}}{\sigma_i^{(1)}}.$$

3. Із нормованих таким чином відхилень обчислюється попередня середня сезонна хвиля

$$v_t^{(1)} = \frac{\sum_{t=1}^m I_t^{(1)*}}{m},$$

де $m = n/T_0$.

4. Середня попередня сезонна хвиля множиться на середнє квадратичне відхилення кожного року (кварталу) і результат вилучається із вихідного ряду

$$U_{1t} = y_t - v_t^{(1)} \cdot \sigma_t^{(1)}.$$

5. Отриманий таким чином ряд без попередньої сезонної хвилі знову згладжується ковзною середньою (для місячних даних використовується від п'яти до семи рівнів залежно від інтенсивності мілких кон'юнктурних коливань і тривалості більш крупних). Унаслідок будемо мати нову оцінку тренда $U_t^{(2)}$. Відхилення емпіричного ряду від тренда

$$l_{2t} = y_t - U_t^{(2)}.$$

знову оброблюється за пунктами 2 і 3 з метою виявлення остаточної сезонної хвилі.

6. Вилучення остаточної середньої сезонної хвилі здійснюється після множення середньої сезонної хвилі на коефіцієнт напруженості сезонної хвилі k

$$v_t = v_t^{(2)} k,$$

$$k = \frac{\sum_{t=1}^{T_0} l_{2t} \varepsilon_{2t}}{\sum_{t=1}^{T_0} \varepsilon_{2t}^2}; \quad (2.16)$$

$$\varepsilon_{2t} = l_{2t} - v_{2t}^{(2)}. \quad (2.17)$$

Перевагою цього методу є відносна простота. Однак його недоліком є те, що ковзна середня не забезпечує використання $T_0/2$ рівнів з обох кінців вихідного ряду.



Модель Тейла-Вейджа

Прогноз здійснюється за формулами

$$\hat{a}_{1,t} = \alpha 1 \cdot (x_t - \hat{g}_{v_t, k_{t-1}}) + (1 - \alpha 1) \cdot (\hat{a}_{1,t-1} + \hat{a}_{2,t-1})$$

$$\hat{g}_{v_t, k_t} = \alpha 2 \cdot (x_t - \hat{a}_{1,t}) + (1 - \alpha 2) \cdot \hat{g}_{v_t, k_{t-1}}$$

$$\hat{a}_{2,t} = \alpha 3 \cdot (\hat{a}_{1,t} - \hat{a}_{1,t-1}) + (1 - \alpha 3) \cdot \hat{a}_{2,t-1}$$

$$\hat{x}_t^* = \hat{a}_{1,t-\tau} + \tau \cdot \hat{a}_{2,t-\tau} + \hat{g}_{v_t, k_{t-1}}$$

Порядок виконання розрахунку:

1. Знайти коефіцієнти лінійної регресії виду $x_t = a_{1,0} + a_{2,0} \cdot t$
2. Розрахувати середнє для декількох перших точок різниці
Це буде параметр $\hat{g}_{v,0}$ $\Delta_t = x_t - \hat{x}_t$
3. Для першої точки взяти значення a_{ij} з розрахунку моделі.
4. Знаходити нове значення \hat{x}_t^* і порівнювати його зі справжнім значенням, визначаючи похибку.
5. Для зручності, сформуванати наступну таблицю

\hat{x}_t	$\Delta_t = x_t - \hat{x}_t$	\hat{x}_t^*	$a_{1,k}$	$a_{2,k}$	$g_{1,k}$	Похибка, %
-------------	------------------------------	---------------	-----------	-----------	-----------	------------

ПРОГНОЗУВАННЯ БАГАТОВИМІРНИХ ПРОЦЕСІВ

Тоді лінійну багатофакторну модель регресії можна записати аким чином

$$y_i = b_0 x_{i0} + b_1 x_{i1} + b_2 x_{i2} + \dots + b_p x_{ip} + \varepsilon_i,$$
$$\bar{y}_n = X_n^T b,$$

де X_n – матриця прогнозних значень факторів.

Довірчі інтервали точкового прогнозу обчислюються за формулами:

$$y_{n \max} = \bar{y}_n + \Delta y_n; \quad y_{n \min} = \bar{y}_n - \Delta y_n,$$

де

$$\Delta y_n = t_{\alpha p} S \left[1 + X_n (X^T X)^{-1} X_n^T \right]$$

Наявність мультиколінеарності призводить до того, що незначна зміна статистичних даних викликає значну зміну значень коефіцієнтів b_j . Отже, перед побудовою багатofакторної лінійної моделі регресії необхідно перевірити відсутність мультиколінеарності.

Одним із найбільш простих способів вилучення мультиколінеарності є виключення з подальшої обробки одного з факторів, що має кореляційну залежність з іншим фактором.



Методи кластерного аналізу використовуються для класифікації (розбиття) сукупності об'єктів на групи „подібних” об'єктів, які називають кластерами, таким чином, щоб об'єкти одного класу знаходились на „близьких” відстанях між собою, а об'єкти різних класів – на відносно „віддалених” відстанях один від одного. При цьому кожний об'єкт X_j ($j = 1, 2, \dots, m$) розглядається як точка в m -мірному просторі.

Функції, що дозволяють віднести новий об'єкт до певного кластеру – це такі функції, що приймають найбільше значення для даних того класу, до якого вони відносяться



МЕТОДИ ЕКСПЕРТНИХ ОЦІНОК

Рівень відповідності експерта

$$k = 0,5 \left(\frac{\sum_{j=1}^m v_j}{\sum_{j=1}^m v_{j \max}} + \frac{\lambda}{P} \right),$$

де v_j – вага градації складової характеристики з номером j в анкеті в балах, перекресленої експертом;

$v_{j \max}$ – максимальна вага (межа шкали) цієї складової характеристики в балах;

m – загальна кількість складових характеристики в анкеті;

λ – вага чарунки, перекресленої експертом у шкалі самооцінки в балах;

P – межа шкали самооцінки експерта в балах.

Перша умова використовується для визначення максимальної чисельності експертної групи n_{\max} за формулою

$$n_{\max} \leq \frac{\sum_{i=1}^n k_i}{ck_{\max}},$$

де c – константа;

k_{\max} – максимально можлива компетентність за використаною шкалою компетентності;

k_i – компетентність окремого експерта з номером i .

Далі визначається мінімальна кількість членів експертної групи n_{\min} . Це здійснюється шляхом використання умови стабілізації середньої оцінки прогнозної характеристики, яка формулюється у вигляді

$$\frac{V - V'}{V_{\max}} < \varepsilon,$$

де V – середня оцінка прогнозної величини в балах, за даними експертної групи;

V' – середня оцінка за даними експертної групи, із якої виведено (або в яку введено) одного експерта;

V_{\max} – максимально можлива оцінка прогнозної величини за прийнятою шкалою оцінок;

ε – задана величина зміни середньої похибки при додаванні чи вилученні одного експерта.

Аналіз експертних висновків

За Кендалом

$$W = \frac{12 \sum_{i=1}^m \left(\sum_{j=1}^d \rho_{ij} - \frac{d(m+1)}{2} \right)^2}{d^2 (m^3 - m)}$$



Якщо оцінки експертів показали достатній рівень узгодженості, для кожного значення змінної розраховується не тільки середнє (V), але й дисперсія D .

Приблизне значення довірчого інтервалу Δ обчислюється за формулою

$$\Delta = \pm t_{\alpha} \sqrt{D \left(1 + \frac{1}{n} \right)},$$

де t_{α} – параметр, який визначається за таблицями Ст'юдента для заданого рівня довірчої ймовірності і ступенів свободи $k = (n - 2)$.

Коефіцієнт варіації оцінок, даних експертами, визначається залежністю

$$g = \frac{\sigma}{V},$$

де $\sigma = \sqrt{D}$.

Індекс сезонності

Індекс сезонності розраховується за формулою:

$$I_c = \langle y \rangle_t / \langle y \rangle_c$$

де $\langle y \rangle_t$ - середній місячний рівень показника за три і більше років,

$\langle y \rangle_c$ - середньомісячне за всі роки значення показника.



Ідентифікація часових рядів

- Перевірка стаціонарності / не стаціонарності часового ряду:

1.1. Метод перевірки різниць середніх рівнів

Крок 1. Вхідний часовий ряд $y_1, y_2, y_3, \dots, y_n$ розподіляють на дві приблизно однакові за кількістю спостережень частини: в першій частині n_1 першої половини рівнів вхідного ряду, у другій — решта рівнів n_2 ($n_1 + n_2 = n$)

Крок 2. Для кожної з цих частин розраховують середні значення й дисперсії:

$$\bar{y}_1 = \frac{\sum_{t=1}^{n_1} y_t}{n_1} \quad \sigma_1^2 = \frac{\sum_{t=1}^{n_1} (y_t - \bar{y}_1)^2}{(n_1 - 1)} \quad \bar{y}_2 = \frac{\sum_{t=1}^{n_2} y_t}{n_2} \quad \sigma_2^2 = \frac{\sum_{t=1}^{n_2} (y_t - \bar{y}_2)^2}{(n_2 - 1)}$$

- *Крок третій.* Перевірка рівності (однорідності) дисперсій обох частин ряду за допомогою F -критерію із табличним (критичним) значенням критерію Фішера F_α із заданим рівнем значущості α . Якщо розрахункове значення $F < F_\alpha$, то гіпотезу про рівність дисперсій приймають, і можна переходити до четвертого кроку. Якщо $F \geq F_\alpha$, гіпотезу про рівність дисперсій відхиляють і доходять висновку, що цей метод не дає відповіді щодо наявності тренду.
- На *четвертому кроці* перевіряють гіпотезу про відсутність тренду за допомогою t -критерію Стюдента. Для цього визначають розрахункове значення критерію Стюдента за формулою:

$$t = \frac{|\bar{y}_1 - \bar{y}_2|}{\hat{\sigma} \sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$

- де $\hat{\sigma}$ — оцінка середньоквадратичного відхилення різниць середніх

Приклад

Доходи	Роки	Середнє значення	Дисперсія	F		t
Млн грн	1999—2000 2001—2002	9860,2 13695,8	8349206 4459451	1,87	2733,44	2,8
% до ВВП	1999—2000 2001—2002	26,0 24,5	6,41 1,92	3,34	2,2	1,37

Метод Форстера-Стьюарта

Крок перший. Порівнюють кожен рівень вхідного часового ряду, починаючи з другого рівня, з усіма попередніми, при цьому визначають дві числові послідовності:

$$k_t = \begin{cases} 1, & \text{якщо } y_t \text{ більше всіх попередніх рівнів} \\ 0, & \text{в іншому разі} \end{cases}$$

$$l_t = \begin{cases} 1, & \text{якщо } y_t \text{ менше всіх попередніх рівнів} \\ 0, & \text{в іншому разі} \end{cases}$$

$t = 2, 3, \dots, n.$

Крок другий. Розраховують величини c і d :

$$c = \sum_{t=2}^n (k_t + l_t) \qquad d = \sum_{t=2}^n (k_t - l_t)$$

Крок третій Перевіряється гіпотеза стосовно того, чи можна вважати випадковими: 1) відхилення величини c від математичного сподівання ряду, в якому рівні розташовані випадково, 2) відхилення величини d від нуля.

$$t_c = \frac{|c - \hat{\mu}|}{\hat{\sigma}_1}$$

$$\hat{\sigma}_1 = \sqrt{2 \ln n - 3,4253}$$

$$t_d = \frac{|d - 0|}{\hat{\sigma}_2}$$

$$\hat{\sigma}_2 = \sqrt{2 \ln n - 0,8456}$$

Метод Форстера-Стьюарта

Крок четвертий. Розрахункові значення t_c і t_d порівнюють із табличним значенням t -критерію із заданим рівнем значущості t_{α} . Якщо розрахункове значення t менше за табличне t_{α} , то гіпотезу про відсутність відповідного тренду приймають, в іншому разі тренд існує. Наприклад, якщо t_c більше табличного значення t_{α} , а t_d менше t_{α} , то для заданого часового ряду існує тренд у середньому, а тренду дисперсії рівнів ряду немає.



Застосування методу Форстера-Стьюарта для двох часових рядів: доходів консолідованого бюджету (млн грн) та доходів консолідованого бюджету (% до ВВП) дає розрахунки, наведені в табл. 3

Таблиця 3

Доходи	Σk_t	Σl_t	c	d	t_c	t_d
Млн грн	8	0	8	8	3,28	4,07
% до ВВП	4	1	5	3	0,9	1,53

На рівні значущості , тобто з імовірністю 0,95 та з числом ступенів волі $n - 2 = 16 - 2 = 14$ табличне значення критерія Стюдента дорівнює $= 2,145$.

Для часового ряду доходів, виражених у млн грн, розрахункові значення t_c і t_d перевищують табличне значення , тобто нульова гіпотеза не приймається, існує тренд як середнього, так і дисперсії ряду.

Для часового ряду доходів, виражених у відсотках до ВВП, розрахункові значення t_c і t_d менші за табличне значення , тобто приймається гіпотеза про відсутність тренду в тенденції й дисперсії ряду

Дисперсія залишків

$$\frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{n}$$

Спрощена схема періодів прогнозування



ОЦІНЮВАННЯ ЯКОСТІ ПРОГНОЗІВ. Абсолютні показники

середню похибку прогнозу, яка розраховується за формулою

$$\bar{e} = \frac{\sum_{t=1}^n e_t}{n}$$

Цей критерій характеризує величину зміщення прогнозу і для правильних прогнозів повинен прямувати до нуля за умови великої кількості спостережень, тобто $\bar{e} \rightarrow 0$ при $n \rightarrow \infty$;

дисперсію похибки

$$D(e) = \frac{\sum_{t=1}^n (e_t - \bar{e})^2}{n}$$

та її *стандартне відхилення*

$$\sigma_e = \sqrt{\frac{\sum_{t=1}^n (e_t - \bar{e})^2}{n}}$$

Цей критерій характеризує величину розкиду значень прогнозу навколо його середнього значення.

абсолютне середнє відхилення

$$|\bar{e}| = \frac{\sum_{t=1}^n |e_t - \bar{e}|}{n};$$

середній квадрат похибки, який для лінійної моделі збігається з дисперсією похибки, —

$$\bar{e}^2 = \frac{\sum_{t=1}^n e_t^2}{n};$$

абсолютну середню відносну похибку

$$|\bar{e}_s| = \frac{1}{n} \sum_{t=1}^n \frac{|e_t|}{y_t} 100\%.$$

Цей критерій використовують при порівнянні точності прогнозів різних об'єктів, оскільки він характеризує відносну точність прогнозу;

середню відносну похибку

$$\bar{e}_s = \frac{1}{n} \sum_{t=1}^n \frac{e_t}{y_t} 100\%.$$

Порівняльні показники

$$k = \sqrt{\frac{\sum_{t=1}^n (y_t^* - y_t)^2}{\sum_{t=1}^n (P_t^* - y_t)^2}},$$

де P_t^*, y_t^* – прогнозне значення еталонного та даного прогнозу відповідно.

В якості еталонного прогнозу може бути вибрано просту екстраполяцію, сталий темп приросту тощо. Частковим випадком показників такого типу є коефіцієнт невідповідності (KH), для якого $P_t^* = 0$ для всіх значень t , тобто

$$KH = \frac{\sum_{t=1}^n (y_t^* - y_t)^2}{\sum_{t=1}^n y_t^2}.$$

$KH=0$ для випадку ідеального прогнозу і $KH=1$, коли прогнозає туж похибку, що і еталонний.

Розглядають дві модифікації KH . Для першої із них KH обчислюється як відношення середньоквадратичної похибки прогнозу до такої ж похибки, яка має місце, коли за прогнозний рівень приймається середній рівень за весь рік

$$KH_1 = \sqrt{\frac{\sum_{t=1}^n (y_t^* - y_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}}.$$

Якщо $KH_1 > 1$, то прогноз методом середнього рівня забезпечує кращі результати порівняльно з даним прогнозом.

Другою модифікацією KH є коефіцієнт розбіжності. Він обчислюється як відношення середньоквадратичної похибки до такої ж похибки, яка має місце, коли за прогнозний рівень для кожного року приймається значення аналітичного тренда y_{ta} , тобто

$$V = \sqrt{\frac{\sum_{t=1}^n (y_t^* - y_t)^2}{\sum_{t=1}^n (y_{ta} - y_t)^2}}.$$

До порівняльних показників відносять і коефіцієнт кореляції r між прогнозними та фактичними значеннями рівнів

$$r = \frac{\sum_{i=1}^n (y_i^* - \bar{y}^*) (y_i - \bar{y})}{\sqrt{\sum_{i=1}^n (y_i^* - \bar{y}^*)^2 \sum_{i=1}^n (y_i - \bar{y})^2}},$$

де $\bar{y}^* = \frac{\sum_{i=1}^n y_i^*}{n}$.

Можна розкласти похибку прогнозу на частки невідповідності. Всі розглянуті вище показники якості прогнозу в своїй основі мають середньоквадратичну похибку. Її квадрат можна подати у вигляді

$$\frac{\sum_{i=1}^n (y_i^* - y_i)^2}{n} = (\bar{y}^* - \bar{y})^2 + (S_{y^*} - S_y)^2 + 2(1-r)S_{y^*}S_y, \quad (5.2)$$

де S_{y^*}, S_y – середньоквадратичні значення прогнозних та фактичних рівнів відповідно, які відповідають формулам:

$$S_{y^*} = \sqrt{\frac{\sum_{i=1}^n (y_i^* - \bar{y}^*)^2}{n}},$$

$$S_y = \sqrt{\frac{\sum_{i=1}^n (y_i - \bar{y})^2}{n}}.$$

$$1 = V_\mu + V_S + V_I,$$

де V_μ, V_S, V_I – так звані коефіцієнти невідповідності: частка зміщення, частка дисперсії, частка коваріації відповідно, які знаходяться за виразами:

$$V_\mu = \frac{n(\bar{y}^* - \bar{y})^2}{\sum_{i=1}^n (y_i^* - y_i)^2}, \quad V_S = \frac{n(S_{y^*} - S_y)^2}{\sum_{i=1}^n (y_i^* - y_i)^2}, \quad V_I = 2 \frac{n(1-r)S_{y^*}S_y}{\sum_{i=1}^n (y_i^* - y_i)^2}.$$

Всі розглянуті вище показники точності прогнозу використовуються для перевірки точності прогнозу моментних рядів. Якщо ж прогнозується інтервальний ряд, то за міру точності прогнозування використовується величина η , яка відповідає відношенню

$$\eta = \frac{p}{p+q},$$

де p – кількість прогнозів, що підтверджується фактичними даними;

q – кількість прогнозів, що не підтверджується фактичними даними.

Якісні показники

До якісних показників відносять такі показники, які дають можливість виконати аналіз видів похибок прогнозів, розкласти їх на деякі складові. Такий аналіз є особливо важливим для рівнів, що змінюються циклічно, коли необхідно прогнозувати не тільки загальний розвиток об'єкту, але і зворотні точки циклу.

Одним із методів такої о аналізу є діаграма „прогноз-реалізація”. Суть методу полягає в побудові поля точкових прогнозів у координатах, де одна вісь відповідає реальним значенням змінної, а друга — її прогнозним значенням. Це дозволяє змістовно оцінити якість різних прогнозів, розрахувати деякі коефіцієнти для аналізу якості прогнозування зворотних точок, виділити найбільш типові похибки (недооцінку чи переоцінку змін).

