УДК 681.51

Демківський Є.О.

Побудова моделі часового ряду з детермінованим трендом

Наведено два підходи до моделювання та прогнозування нестаціонарних відносно тренду часових рядів. Запропоновано загальну схему моделювання та прогнозування нестаціонарних часових рядів, яка відрізняється тим, що одночасно аналізується можливість математичного описання тренду і його вилучення за допомогою різниць. Розглянуто можливість використання коефіцієнат Тейла для визначення придатності отриманих моделей для прогнозування.

Two approaches are considered to modeling and forecasting of trend stationary time series. A general modeling and forecasting schema is proposed for nonstationary time series that is distingnished with a simultaneous analysis of possibilities for mathematical description of trend as well as its elimination ming differences. A possibility of their coefficient use is considered to determine applicability of models to forecasting.

Вступ

Практично будь-який реальний динамічний процес довільної природи можна описати статистичною моделлю при умові наявності відповідних експериментальних даних. Застосування законів статистики й теорії імовірності до опису різних процесів на підприємстві - перспективний напрям у сучасному керуванні якістю, прийнятті бізнесових рішень та керуванні технічними процесами. За допомогою отриманих моделей описуються явища, у яких наявні статистичні фактори, що не дозволяють пояснити явище в чисто детерміністських термінах. Типові приклади такого роду процесів представляють часові ряди, що мають тренд-циклічний компонент і випадкову складову, викликану впливом сукупності різних факторів [1-3].

Нехай спостережувана величина $y(t)$ представляється у вигляді суми тренду $f(t)$ і випадкового процесу $u(t)$, тобто $y(t)=f(t)+u(t)$. Випадковий процес $u(t)$ некорельований і має нульове математичне сподівання й скінченну постійну дисперсію σ^2.

Якщо функція $f(t)$ є поліномом ступеня $q: f(t)=a_0+a_1t+\cdots+a_qt^q$, то говорять про поліноміальний тренд. Для визначення коефіцієнтів полінома $a_0, a_1, \ldots, a_q$ використовується метод найменших квадратів (МНК). Поліноміальний тренд є зручним способом опису часового ряду. У багатьох випадках поліном заміняє іншу, набагато складнішу, але невідому функцію часу. Підібраний поліном зазвичай може використовуватися для інтерполяції, але не для екстраполяції значень.

Постановка задачі

Розробити технологію побудови математичних моделей прогнозів з детермінованим і стохастичним трендом, представлених часовим рядом

$\{y(k)\}, k = 1,2,\ldots,N$ при $E[y(k)] = Const$. Тобто побудувати лінійну модель вигляду $y(k) = f(y(k-i), \varepsilon(л)), i = 1,2,\ldots,p$, де $\varepsilon(k)$ – випадів процес з нульовим середнім і скінченною дисперсією.

Для роботи з рядами, що містять тренд, необхідно визначитись:

– чи можливе моделювання та прогнозування ряду безпосередньо;

– чи потрібно попередньо описати та вилучити тренд і моделювати процес як дві окремі складові (тренд та коливальна складова);

– які критерії потрібно застосовувати при визначенні можливості застосування того чи іншого сценарію.

Розв'язок задачі моделювання та прогнозування

При моделюванні і прогнозуванні рядів з трендом пропонується алгоритм, наведений на рис. 1. В загальному випадку існують два способи обробки рядів з трендом: (1) моделювання тренду за допомогою вибраної функції; (2) вилучення тренду за допомогою перших різниць вищих порядків.

При моделюванні та прогнозуванні тренду часових рядів пропонується застосовувати алгоритм, наведений на рис.1.

Вибір методу визначається, по-перше, кінцевою метою моделювання, тобто потрібно прогнозувати тренд чи ні. Крім того, вибір методу визначається характером тренду – стохастичні або детерміновані випадки.

Рис. 1. Схема методу моделювання та прогнозування тренду часових рядів

Для описання часових рядів розглянемо можливість застосування рівняння авторегресії з ковзним середнім (АРКС). Як критерії адекватності моделі та можливості її застосування для прогнозування будемо використовувати:

– коефіцієнт детермінації - відношення дисперсії тієї частини часового ряду основної змінної, що описується отриманим рівнянням, до вибіркової дисперсії цієї змінної ($R^2 \to 1$);
– сума квадратів похибок моделі ($\sum e^2(k) \to \min$);
– статистика Дарбіна-Уотсона, що обчислюється за формулою: $DW = 2 - 2\rho \to 2$, де $\rho -$ коефіцієнт кореляції між значеннями випадкової змінної $\varepsilon(k) \approx e(k)$, тобто wей параметр дозволяє визначити ступінь корельованості похибок моделі (DW);
– середньоквадратична похибка (*СеКП*);
– середня абсолютна похибка (*САП*);
– середня абсолютна похибка в процентах (*САПП*);
– коефіцієнт Тейла, який свідчить про загальну придатність моделі для прогнозування (ідеальне значення – нуль) та має такий вигляд:

$$U = \frac{\sqrt{\frac{1}{N}\sum_{k=1}^{N}[y(k) - \hat{y}(k)]^2}}{\sqrt{\frac{1}{N}\sum_{k=1}^{N} y^2(k)} + \sqrt{\frac{1}{N}\sum_{k=1}^{N} \hat{y}^2(k)}}.$$

Розглянемо часовий ряд, що характеризує зміни ціни акцій компанії УКРНАФТА, динаміка якого наведена на рис. 2.

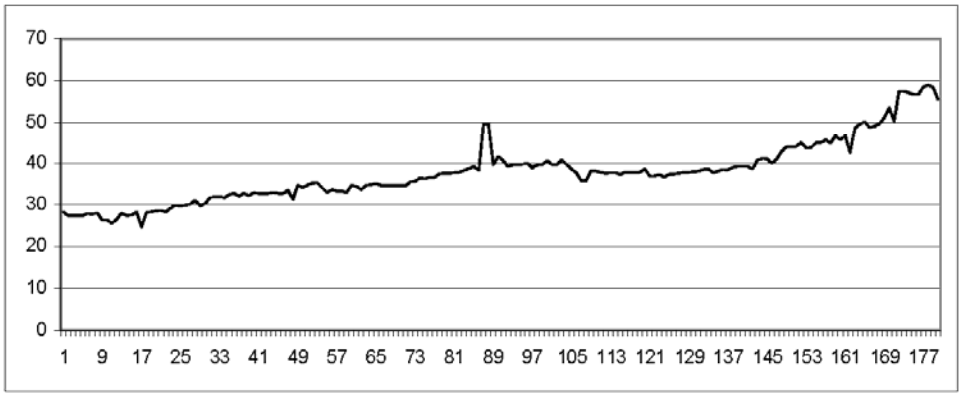

Рис. 2. Динаміка ціни акцій компанії УКРНАФТА

Опишемо даний процес за допомогою моделі АР(5) (кількість та номери запізнень вибираємо на основі значень парціального коефіцієнту кореляції). Модель має такий вигляд:

$$y(k+1) = -0{,}5628 + 0{,}6555\,y(k) + 0{,}04\,y(k-1) + 0{,}2321\,y(k-2) + 0{,}0962\,y(k-4) + \varepsilon(k)$$

Значення критеріїв якості моделі та придатності її для прогнозування наведено в таблиці 1.

Наведені критерії свідчать про те, що дана модель може бути застосована для короткострокового прогнозу. Результат прогнозування на п’ять кроків та порівняння його з істинними значеннями вхідного ряду наведено на рис. 3.

Дані, наведені в таблиці 2, свідчать про те, що наведений ряд може бути спрогнозовано безпосередньо, хоча з рис. 1 очевидно, що ряд містить тренд.

Спробуємо отримати прогноз цього ж ряду за допомогою наступних трьох кроків: (1) опишемо й вилучимо тренд та отримаємо його прогноз; (2) змоделюємо та отримаємо прогноз залишків ряду після вилучення тренду; (3) на основі прогнозів, отриманих на кроках (1) і (2), спрогнозуємо поведінку вхідного часового ряду та

порівняємо його з прогнозом, оцінки якого наведені в таблиці 2.

Таблиця 1

Модель	Характеристики моделі			Характеристики прогнозу			
	R^2	$\sum e^2(k)$	DW	*САП*	*САП*	*САПП*	Коеф-т Тейла
AR(5)	0,9434	402,9998	2.0180	3,8352	3,2285	8,4128	0,0503

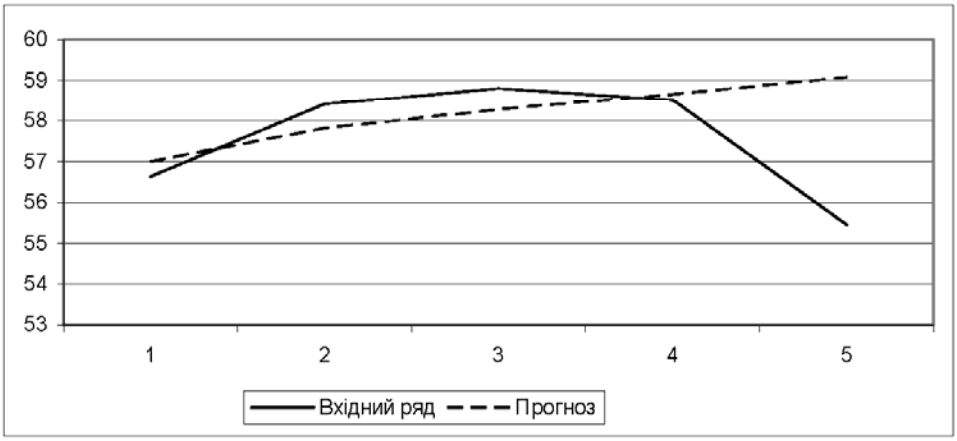

Рис. 3. Прогноз ряду ціни акцій компанії УКРНАФТА

Результат оцінки якості прогнозу на п'ять кроків наведено в таблиці 2.

Таблиця 2

Метод прогнозування	Максимальне відхилення		Мінімальне відхилення		Сума квадратів похибок
	Абсолютне	%	Абсолютне	%	
За моделлю	0,65	6,57	0,15	0,26	14,08

Тренд вхідного ряду, описаний як поліном шостого ступеню, має вигляд наведений на рис. 4.

Поліном, що описує даний тренд:

$$y = -3{,}5167 \cdot 10^{-11} \cdot x^6 + 1{,}9179 \cdot 10^{-8} \cdot x^5 - 3{,}6372 \cdot 10^{-6} \cdot x^4 + 0{,}0003 \cdot x^3 - 0{,}0092 \cdot x^2 + 0{,}2272 \cdot x + 26{,}3694$$

Залишки, отримані після вилучення тренду, мають вигляд наведений на рис.5.

Згідно з аналізом автокореляційної функції побудована модель АР(3), яка описує ряд залишків, має такий вигляд:

$$e(k+1) = -0{,}0101 + 0{,}5025 \cdot e(k) - 0{,}0183 \cdot e(k-1) - 0{,}1552 \cdot e(k-2) + \varepsilon(k)$$

Значення критеріїв якості моделі та придатності її для прогнозування наведені в таблиці 3.

Як видно з таблиці 3, значення коефіцієнта Тейла значно відрізняється від нуля. Але, як виявлено експериментальним шляхом, коефіцієнт Тейла для знако-перемінних рядів не може бути застосований. Для підтвердження цього факту наведемо приклад. Змоделюємо ряд авторегресії другого порядку ($y(k+1) = a_0 + 0{,}5y(k) - 0{,}4y(k-1) + \theta$) з початковими значеннями $y(0) = -1, y(1) = 1$ та білим шумом θ, рівномірно розподіленим на інтервалі [-0,1;0,1]. Для даного модельного ряду візьмемо три різних значення зсуву a_0 :

(а) – -5; (б) – 0; (в) – +5. Результати моделювання наведено на рис.6.

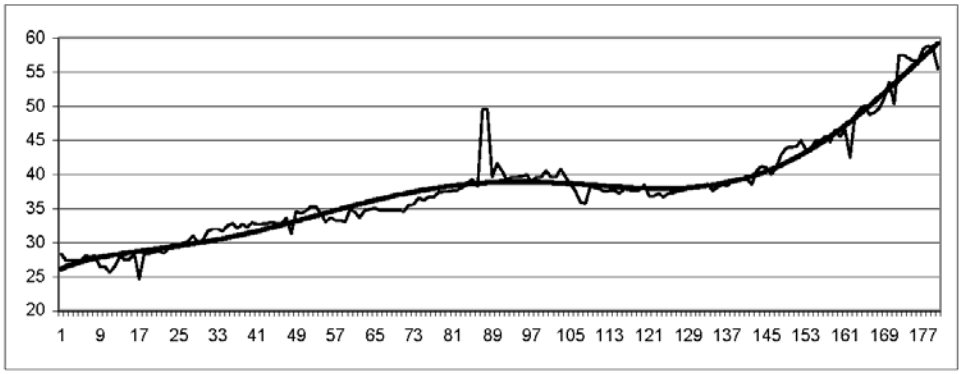

Рис. 4. Тренд ряду ціни акцій компанії УКРНАФТА

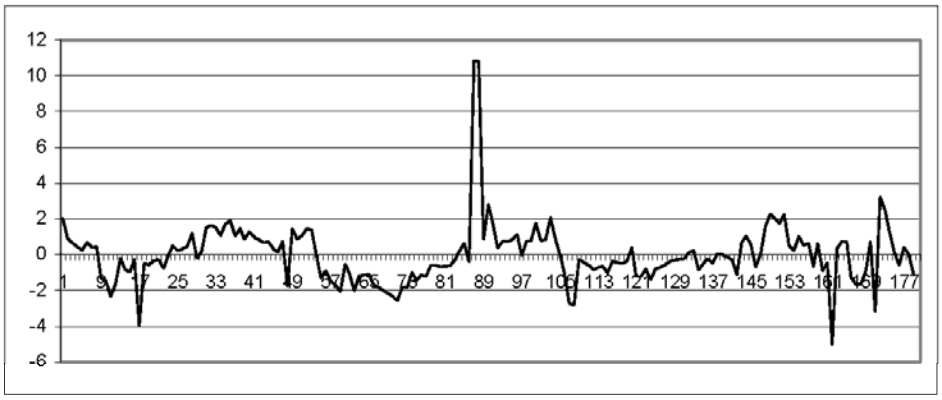

Рис. 5. Залишки ряду ціни акцій компанії УКРНАФТА після вилучення тренду

Таблиця 3

Модель	Характеристики моделі			Характеристики прогнозу			
	R^2	$\sum e^2(k)$	DW	*САП*	*САП*	*САПП*	Коеф-т Тейла
АР(3)	0,3177	348,193	1,9996	1,7235	1,1163	99,145	0,963

Параметри моделей (а), (б) і (в) наведено в таблиці 4.

Прогноз, отриманий за трьохкроковою процедурою, має вигляд, наведений на рис.7.

Результат оцінки якості прогнозу на п’ять кроків наведено в таблиці 5.

Якщо порівняти значення, наведені в таблиці 2 і таблиці 5, то можна зробити висновок, що наведена технологія роздільного прогнозування (окремо тренду та коливальної складової, може бути успішно застосована на практиці з метою отримання якісного багатокрокового прогнозу.

Для прийняття рішення щодо визначення виду та способу прогнозування часових рядів на основі проведених досліджень пропонується алгоритм, наведений на рис. 8.

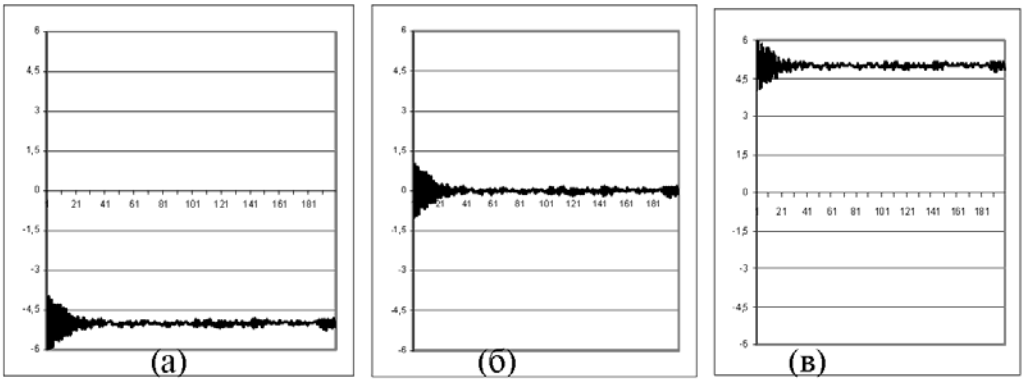

Рис. 6. Модельні ряди зі значеннями a_0 : (а) – -5; (б) – 0; (в) – +5

Таблиця 4

Модель	Характеристики моделі			Характеристики прогнозу			
	R^2	$\sum e^2(k)$	*DW*	*САП*	*САП*	*САПП*	Коеф-т Тейла
(а)	0,9173	0,677	1,9132	0,083	0,0667	1,3348	0,0083
(б)	0,9173	0,677	1,9132	0,083	0,0667	126,7693	0,2205
(в)	0,9173	0,677	1,9132	0,083	0,0667	1,3404	0,0083

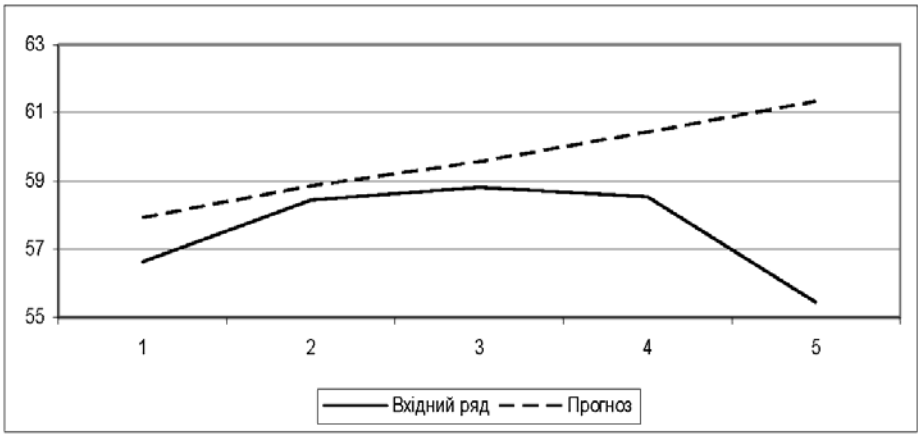

Рис. 7. Прогноз ряду ціни акцій компанії УКРНАФТА

Таблиця 5

Метод прогнозування	Максимальне відхилення		Мінімальне відхилення		Сума квадратів похибок
	Абсолютне	%	Абсолютне	%	
За моделлю	5,87	10,59	0,42	0,42	40,47

Висновок і перспективи досліджень

Таким чином у роботі запропоновано підхід до побудови математичних моделей процесів з детермінованим і стохастичним трендом. Показано, що вилучення тренду може призводити до утворення знакоперемінного ряду і, відповідно, погіршення якості моделі залишків. Модель завжди має кращі характеристики у випадку, коли значення

ряду мають один знак.

У подальших дослідженнях планується дослідити якість моделей в умовах двох видів нестаціонарності – наявність тренду і гетероскедастичність. Це більш складний випадок, оскільки необхідно описувати окремо рівень змінної, тренд і дисперсію.

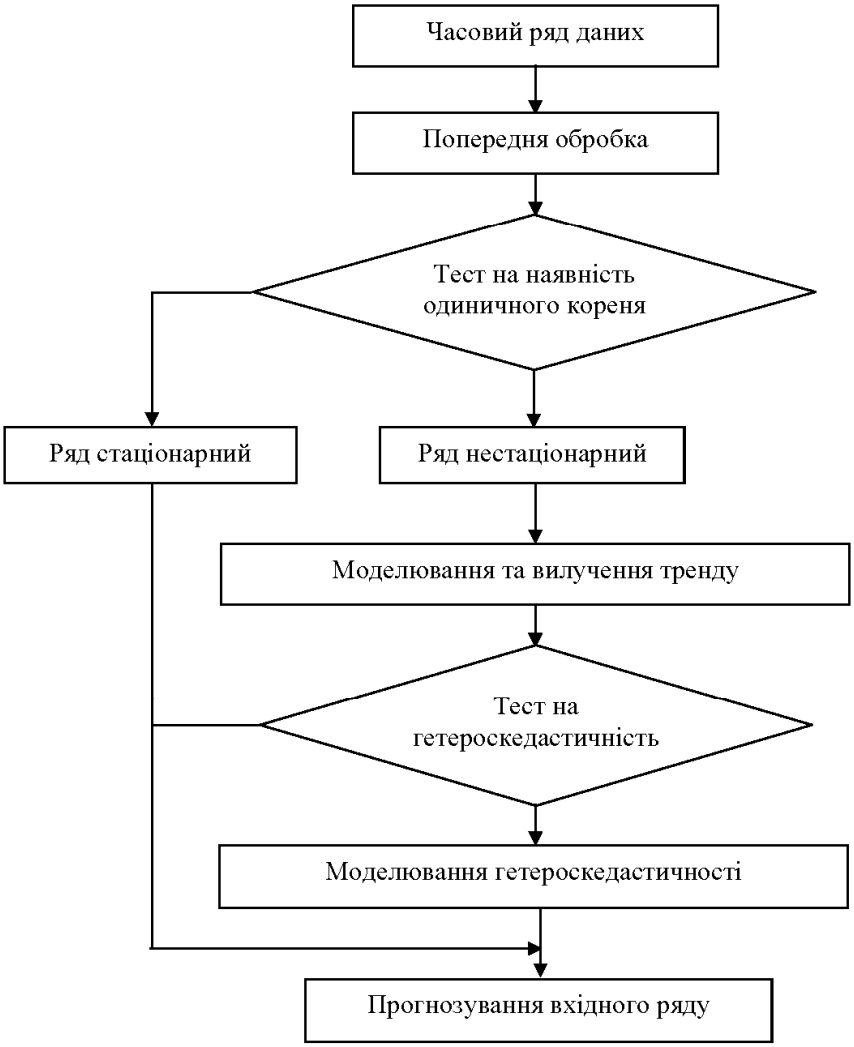

Рис. 8. Схема прийняття рішення, щодо визначення виду та способу прогнозування часових рядів

Література

1. Дж. Бокс, Г. Дженкинс. Анализ временных рядов. Прогноз и управление. Вып. 1.– М.: Мир, 1974. – 408 с.
2. Доугерти К. Введение в эконометрику. – М.: Инфра-М, 2001. – 402 с.
3. Бідюк П.І., Савенков О.І, Баклан І.В. Часові ряди: моделювання та прогнозування. –К: ”ЕКМО”, 2003 р., –144 с.
4. Зайченко Ю.П. Нечіткий метод індуктивного моделювання в задачах прогнозування макроекономічних показників.// Системні дослідження і інформаційні технології. – 2003. № 3. – с. 25-45.
5. Лукашин Ю.П. Прогнозирование временных рядов с помощью моделей авторегрессии–скользящего среднего первого и второго порядка. – М.: ИМЭМО, 1983. – с.107.