УДК 004:519.2
DOI: 10.15587/1729-4061.2018.121620

Розробка адаптивних комбінованих моделей прогнозування часових рядів на основі ідентифікації подібностей

О. Ю. Кучанський, А. О. Білощицький, Ю. В. Андрашко, С. В. Білощицька, Є. Є. Шабала, О. В. Миронов

Пропонуються адаптивні комбіновані моделі гібридного та селективного типів для прогнозування часових рядів на основі програмного набору з адаптивних поліноміальних моделей різних порядків. Пропонуються адаптивні комбіновані моделі прогнозування часових рядів з врахуванням результатів ідентифікації подібностей в ретроспекції цих часових рядів. Оцінена ефективність прогнозування різних комбінованих моделей залежно від рівня персистентності часових рядів. Розроблені моделі дозволяють підвищити точність у випадку середньострокового прогнозування нестаціонарних часових рядів, зокрема фінансових показників

Ключові слова: прогнозування часових рядів, пошук подібностей, адаптивна комбінована модель, показник Герста

1. Вступ

Відомо, що переважна більшість фінансових, технічних, фізичних процесів, для яких виникає задача прогнозування, характеризуються нелінійністю і нестійкістю відносно середнього рівня. Застосування класичних економетричних прогнозних моделей і відповідних методів прогнозування таких часових рядів, які відображають дані процеси, доволі обмежене. Це пов'язано з малою ефективністю даних моделей в цих умовах. Перспективними напрямками в розробці інформаційних систем аналізу і прогнозування часових рядів та відповідних систем підтримки прийняття рішень є побудова комбінованих багаторівневих моделей та використання методів інтелектуального аналізу даних. Крім того, актуальним є передпрогнозне оцінювання вхідних даних, в тому числі фінансових показників, що може бути реалізоване на базі фрактального R/S-аналізу.

Розв'язання протиріч між вимогами щодо планування стратегій діяльності фінансових установ і завданнями науково обґрунтованої підтримки прийняття рішень можливо за рахунок використання математичних моделей прогнозування фінансових часових рядів. Відомі математичні моделі не дозволяють отримати необхідну точність прогнозування часових рядів для розв'язання задач управління фінансовими ризиками в умовах невизначеності. У зв'язку із цим, актуальними слід вважати дослідження, які полягають у розв'язанні зазначеного протиріччя шляхом розроблення адаптивних комбінованих моделей з використанням результатів інтелектуального аналізу часових рядів. Зокрема ідентифікації подібностей методами найближчого сусіда. Також використання апарату фрактального аналізу даних для передпрогнозного аналізу часових рядів.

2. Аналіз літературних даних і постановка проблеми

Задача ідентифікації подібностей в динаміці часових рядів є важливою складовою в дослідженні зміни цінових значень на біржі, знаходженні товарів зі схожими особливостями в термінах та умовах реалізації. Також задача ідентифікації подібностей використовується для аналізу на подібність звуку, текстової, числової або графічної інформації, детекції аномалей або подій в сигналах різної природи тощо.

В роботі [1] розроблена концепція "approximately similar timeseries data", яка дає розуміння схожості між часовими рядами і яка є в основі задачі ідентифікації подібностей. Проте ця концепція не включає метод розрахунку подібності між часовими рядами шляхом звичайного розрахунку метричних відстаней між ними. Класичні основи задачі ідентифікації подібностей та класифікацію було описано в роботах [2, 3]. В роботі [4] описано основні положення задачі пошуку подібностей шляхом зіставлення зі зразком, розглянуті методи розрахунку метричних відстаней, задачі сегментації та апроксимації часових рядів.

Можна виділити ряд робіт [5–26], які описують застосування ідентифікації подібностей в різних прикладних задачах. Зокрема, в роботі [5] досліджено прогностичну потужність методу кластеризації з використанням подібної поведінки часових рядів фондового ринку, а також використання даного методу для ефективного прогнозування цін на акції. В роботі [6] запропоновано метод моделювання зразків для задачі короткострокового прогнозування часових рядів. В роботі [7] розглянуто метод селективного зіставлення зі зразком. Даний метод використовується для задачі побудови комбінованих моделей прогнозування знаків приростів часових рядів з нестабільним характером коливань з врахуванням ідентифікації подібностей або індексації. В роботі [8] описано метод прогнозування знаків приростів часових рядів в умовах невизначеності, що використовує трендові моделі плинних середніх. В роботах [9, 10] для цієї задачі розглядаються не зіставлення зі зразком, а метод найближчого сусіда. В роботі [11] описано метод індексації часових рядів на основі ідентифікації подібностей між ними. Даний метод ефективний для часових рядів різної довжини. Метод кластеризації для пошуку подібностей в багатовимірних просторах, які представляються багатовимірними векторами або часовими рядами фіксованої довжини описано в роботі [12].

Використання методологій ідентифікації подібностей використовується також в освіті для прогнозування потенціалу розвитку наукових напрямів. Метод, який дозволяє розрахувати прогноз потенціалу наукових напрямів описано в роботі [13]. Значення часових рядів в цьому випадку представляють собою відношення оцінок результатів науково-дослідної діяльності науковців за різні періоди часу. Метод може бути використано для виявлення перспективних напрямків досліджень, які формуються в науковому середовищі. Метод класифікації науковців за напрямами досліджень, який базується на ідентифікації подібностей описано в роботі [14]. Метод побудови оцінок результатів науково-дослідної діяльності науковців на основі аналізу цитувань публікацій описано в роботі [15]. В роботі [16] описано параметричну модель оцінювання та прогнозування якості закладів освіти, що використовує підходи порівняння оцінок. Метод, який дозволяє оцінити не тільки заклади освіти, але і структурні підрозділи вищих навчальних закладів, використовуючи перетворення комплексу розроблених оцінок з якісних у кількісні описано в роботі [17]. Задача ідентифікації подібностей в цьому випадку може бути корисною для побудови актуальних оцінок, враховуючи інші заклади освіти або структурні підрозділи, які мають аналогічну вагу.

Задача ідентифікації подібностей також використовується для детекції неповних дублікатів в текстовій інформації з ціллю боротьби з плагіатом та незаконного поширення інформації. Зокрема в роботі [18] розроблено метод детекції неповних дублікатів у таблицях, який базується на основі методів найближчого сусіда та локально-чутливого хешування. В роботі [19] описана концептуальна модель системи знаходження неповних дублікатів з використанням ідентифікації подібностей в електронних документах. Модель дозволяє ідентифікувати неповні дублікати в документах, що містять дані різних типів: текстову інформацію, числові дані, таблиці, математичні формули, схеми, діаграми та інші графічні зображення.

Для прогнозування часових рядів, окрім ідентифікації подібностей, можуть використовуватися і традиційні моделі прогнозування. Проте слід враховувати, що традиційні економетричні моделі часто показують результати зі значною похибкою, особливо у випадку застосування для фінансових часових рядів. Це пов’язано з тим, що в умовах кризових явищ, фінансові часові ряди часто є слабко персистентними або навіть близькими до випадкових, тобто втрачають пам'ять про свої початкові умови. Тим не менше, у випадку, якщо вхідні часові ряди трендостійкі, застосування традиційних моделей прогнозування має сенс. Основними етапами при побудові прогнозу в цьому випадку є [7]: прогнозна ретроспекція, побудова моделі прогнозування і формалізація моделі, побудова проспекції, оцінка прогнозу і його верифікація.

В роботі [20] описано підходи інтелектуальних методів прийняття рішень в бізнесі, що використовують класичні моделі прогнозування часових рядів. Адаптивні моделі короткострокового прогнозування часових рядів, способи побудови комбінованих методів описані в роботі [21]. В роботі [22] розглянуто моделі і методи прогнозування часових рядів, які використовують інтелектуальний аналіз даних: моделювання нейромереж, генетичні алгоритми, нечіткий аналіз тощо. Всі ці методи можуть використовувати концепцію ідентифікації подібностей і, як наслідок, збільшувати свою ефективність.

Як класичні моделі прогнозування, так і моделі прогнозування з використанням ідентифікації подібностей, часто використовують як складові складних систем прогнозування, моделювання та прийняття рішень. В роботі [23] проведено аналіз особливостей проектування інформаційно-аналітичних систем прогнозування часових рядів з використанням експертного оцінювання. В роботі [24] зокрема, розглядається метод нечіткої кластеризації, який використовує задачу ідентифікації подібностей. Застосування експертного оцінювання для прогнозування часових рядів в інформаційних системах також описано в роботі [25]. В роботі [26] розглядаються особливості управління конфігурацією проектів при розробці розподілених систем, яка може використовувати ідентифікацію подібностей між проектами за набором показників, які дані проекти характеризують.

Перспективною є спроба поєднання традиційних моделей прогнозування та методу ідентифікації подібностей в ретроспекції часових рядів та побудови комбінованих моделей прогнозування. Ці моделі повинні бути наділені адаптивними властивостями: щоб динамічно підлаштовувалися до механізмів, що генерують часові ряди. Передбачається, що такі моделі будуть ефективними для прогнозування часових рядів в умовах невизначеності, зокрема даних фондових ринків, сировинних часових рядів тощо.

3. Ціль і задачі дослідження

Ціллю дослідження є створення комбінованих моделей прогнозування часових рядів в умовах невизначеності з адаптивними характеристиками та максимальною ефективністю. Під максимальною ефективністю моделі розуміється досягнення цією моделлю мінімальних середніх похибок прогнозування.

Для досягнення мети були поставлені такі завдання:

– побудувати адаптивні комбіновані моделі гібридного та селективного типів для прогнозування часових рядів;

– побудувати адаптивні комбіновані моделі прогнозування часових рядів гібридного та селективного типів з врахуванням результатів ідентифікації подібностей в ретроспекції цих часових рядів.

4. Формальна постановка задачі прогнозування часових рядів та оцінювання результатів прогнозування

Нехай $\{z_i\}_{i=0}^{n}$ – деякий дискретний часовий ряд без пропусків довжини n:

$$\begin{aligned}\{z_i\}_{i=0}^{n} &= \{z_0, z_1, \ldots, z_n\} = \\ &= \{z(t_0), z(t_1), \ldots, z(t_n)\},\end{aligned} \tag{1}$$

де $t_i \in S$ – дискретні моменти часу, в які фіксуються значення часового ряду, $i = \overline{0,n},$ S – дискретна множина, t_0 – початковий момент часу [7].

На основі ретроспективних значень z_n, $z_{n+1}, \ldots,$ z_{n-m+1} часового ряду $\{z_i\}_{i=0}^{n}$, $n \geq m$, найбільш точно оцінити поведінку цього часового ряду в майбутньому в моменти часу t_{n+1}, $t_{n+2}, \ldots,$ $t_{n+\theta}$, тобто побудувати послідовність прогнозних значень:

$$\{\hat{z}_i\}_{i=n+1}^{n+\theta} = \{\hat{z}_{n+1}, \hat{z}_{n+2}, \ldots, \hat{z}_{n+\theta}\}, \tag{2}$$

де θ – горизонт прогнозування, а m – об'єм ретроспективної вибірки.

Нехай $\hat{z}_\tau(n)$ – прогноз, який розраховується в момент t_n (в точці n) на τ точок вперед, $\tau=\overline{1,\theta}$. Функціональна залежність, яка дозволяє описати поведінку часового ряду, називається моделлю прогнозування. Позначимо через F модель, яка описує поведінку часового ряду $\{z_i\}_{i=0}^{n}$. Прогноз даного часового ряду на одну точку вперед (τ=1) на основі моделі F можна записати у вигляді:

$$\hat{z}_{n+1}=\hat{z}_1(n)=F\left(z_{n-m+1},z_{n-m},\ldots,z_n\right), \tag{3}$$

де $\hat{z}_1(n)$ – прогноз часового ряду, що розраховується в точці n на одну точку вперед.
У випадку прогнозування з горизонтом τ>1, можна записати:

$$\begin{aligned}
&\hat{z}_{n+1}=\hat{z}_1(n)=F\left(z_{n-m+1},z_{n-m+2},\ldots,z_n\right),\\
&\hat{z}_{n+2}=\hat{z}_2(n)=F\left(z_{n-m+2},z_{n-m+3},\ldots,z_n,\hat{z}_{n+1}\right),\\
&\vdots\\
&\hat{z}_{n+\theta}=\hat{z}_\theta(n)=F\left(z_{n-m+\theta},\ldots,\hat{z}_{n+\theta-2},\hat{z}_{n+\theta-1}\right),
\end{aligned} \tag{4}$$

де m – об’єм ретроспективної вибірки, $n \geq m$.

Як правило, моделі прогнозування мають ряд параметрів, які необхідно оцінити перед розрахунком прогнозу. Ці параметри можуть мати адаптивний характер, тобто змінюватися динамічно, враховуючи похибки прогнозування на попередніх кроках. Найбільш точною буде вважатися та модель, яка задовольняє відповідний критерій оцінки якості прогнозування. Такими критеріями можуть бути середня квадратична похибка, середнє абсолютне відхилення, стандартне відхилення, відносна похибка і т. д.

Більшість моделей прогнозування для розрахунку прогнозу потребують певного об’єму ретроспективної інформації. Якщо наявно достатньо історичних даних спостережень часового ряду, то перед розрахунком прогнозу доцільно оцінити якість моделі прогнозування на даному часовому ряді. Оцінки можуть бути використані для побудови довірчих інтервалів прогнозів або для уточнення параметрів моделі прогнозування.

Для оцінки якості моделі прогнозування F розглянемо ретроспективний часовий ряд $Z'=\{z_i\}_{i=n-m}^{n}$ довжини m+1. Побудуємо прогнозні значення ретроспективного часового ряду на основі ряду $\{z_{n-m-h+1},z_{n-m-h+2},\ldots,z_{n-m-1}\}$ довжини h–1, який позначимо так: $Z''=\{z_i\}_{i=n-m-h+1}^{n-m-1}$. Для оцінки якості прогнозування на 1, 2,…, θ точок вперед ітераційно застосуємо модель F за схемою (4).

Для τ=1 формально це можна записати так:

$$\begin{aligned}
&\hat{z}_1(n-m-1)=F\left(z_{n-m-h+1},z_{n-m-h+2},\ldots,z_{n-m-1}\right),\\
&\hat{z}_1(n-m)=F\left(z_{n-m+h+2},z_{n-m+h+3},\ldots,z_{n-m}\right),\\
&\vdots
\end{aligned}$$

$$\hat{z}_1(n-1)=F(z_{n-m-1},\ldots,z_{n-2},z_{n-1}),\quad m+l\le n, \tag{5}$$

де $\hat{z}_1(i-1)$ – прогноз, який виконується в момент i–1 на 1 точку вперед, тобто прогноз i-го елемента ретроспективного ряду Z', $i=\overline{n-m,n}$. Послідовність таких прогнозів позначимо через $\hat{Z}'_1=\{\hat{z}_1(i-1)\}_{i=n-m}^{n}$.

Для τ=2 процес знаходження оцінок прогнозів буде мати вигляд:

$$\hat{z}_1(n-m-1)=F(z_{n-m-h+1},z_{n-m-h+2},\ldots,z_{n-m-1}),$$

$$\hat{z}_2(n-m-1)=F(z_{n-m+h+2},z_{n-m+h+3},\ldots,\hat{z}_1(n-m-1)),$$

$$\hat{z}_1(n-m)=F(z_{n-m-h+2},z_{n-m-h+3},\ldots,z_{n-m}),$$

$$\hat{z}_2(n-m)=F(z_{n-m+h+3},z_{n-m+h+4},\ldots,\hat{z}_1(n-m)),$$

$$\vdots$$

$$\hat{z}_1(n-2)=F(z_{n-m-2},z_{n-m-1},\ldots,z_{n-2}),$$

$$\hat{z}_2(n-2)=F(z_{n-m-1},z_{n-m},\ldots,\hat{z}_1(n-2)),\quad m+l\le n, \tag{6}$$

де $\hat{z}_2(i-2)$ – прогноз, який виконується в момент i–2 на 2 точки вперед, тобто прогноз і-го елемента ретроспективного ряду Z', $i=\overline{n-m+1,n}$. Послідовність з таких елементів позначимо через $\hat{Z}'_2=\{\hat{z}_2(i-2)\}_{i=n-m+1}^{n}$. За наведеною схемою можна побудувати й інші послідовності прогнозів $\hat{Z}'_3$, $\hat{Z}'_4$ і т. д.

Для побудови критерію оцінки якості моделі прогнозування необхідно визначити, які з отриманих прогнозних значень кожної з послідовностей $\hat{Z}'_\tau$, $\tau=\overline{1,\theta}$ слід врахувати у цільовій функції критерію. Підпослідовності послідовностей $\hat{Z}'_\tau$, $\tau=\overline{1,\theta}$, які використовуються для оцінювання точності прогнозування назвемо оцінювальними і позначимо через $\hat{Z}'^{*}_\tau=\{\hat{z}_\tau(i-\tau)\}_{i\in J_\tau}$, де J_τ – множини індексів елементів послідовностей $\hat{Z}'_\tau$, які відібрані для побудови оцінювальної послідовності $\hat{Z}'^{*}_\tau$, $\tau=\overline{1,\theta}$, $n-m+\tau-1\le i\le n$, card(J_τ)≤m, card(J_τ) – кількість елементів послідовностей $\hat{Z}'^{*}_\tau$. Послідовність $\hat{Z}'^{*}_\tau$, $\tau=\overline{1,\theta}$ може бути побудована з найбільш значимих прогнозів за допомогою експертного оцінювання. Крім того, може бути вказана кількісна оцінка значимості кожного окремого прогнозу за допомогою вагових коефіцієнтів. Тобто в результуючу цільову функцію будуть включені всі прогнозні значення, $\hat{Z}'^{*}_\tau=\hat{Z}'_\tau$, $\tau=\overline{1,\theta}$, але з ваговими коефіцієнтами, які можуть приймати нульові значення.

Після побудови оцінювального ряду розраховується критерій якості моделі прогнозування, який дозволяє встановити міру задоволення поставленим цілям. В даному випадку ціллю є досягнення максимальної точності прогнозів.

Розглянемо основні критерії оцінки якості:

$$E_\tau^0 = E^0\left(Z',\hat{Z}'^*_\tau\right) = \frac{1}{v_\tau}\sum_{\substack{i=\overline{1,v_\tau}\\ j\in J_\tau}} \omega_i \left|\hat{z}_j^\tau - z'_j\right| \tag{7}$$

– прогноз на основі середнього абсолютного відхилення з врахуванням значимості прогнозів за допомогою нормованих ваг ω_i,

$$i=\overline{1,v_\tau},\quad \sum_{i=1}^{v_\tau}\omega_i = 1,$$

тут і далі $n-m+\tau-1 \le j \le n$, $\nu_\tau = \mathrm{card}(J_\tau)$, $\nu_\tau \le m$, $\tau=\overline{1,\theta}.$ Позначення $E_\tau^0 = E^0\left(Z',\hat{Z}'^*_\tau\right)$ означає, що для оцінки використовується критерій $E^0,$ за яким члени ретроспективного ряду Z' порівнюються з відповідними членами оцінювального ряду $\hat{Z}'^*_\tau,$ причому прогноз розраховується τ кроків вперед $\tau=\overline{1,\theta}.$ Також для оцінювання може бути використана середньоквадратична похибка (8), стандартне відхилення (9), середня відносна похибка (10) та максимальна абсолютна похибка (11) відповідно, які розраховуються за формулами для $\tau=\overline{1,\theta},\ \sum_{i=1}^{v_\tau}\omega_i = 1$:

$$E_\tau^1 = E^1\left(Z',\hat{Z}'^*_\tau\right) = \frac{1}{v_\tau}\sum_{\substack{i=\overline{1,v_\tau}\\ j\in J_\tau}} \omega_i \left(\hat{z}_j^{\,\tau} - z'_j\right)^2, \tag{8}$$

$$E_\tau^2 = E^2\left(Z',\hat{Z}'^*_\tau\right) = \sqrt{\frac{1}{v_\tau}\sum_{\substack{i=\overline{1,v_\tau}\\ j\in J_\tau}} \omega_i \left(\hat{z}_j^\tau - z'_j\right)^2}, \tag{9}$$

$$E_\tau^3 = E^3\left(Z',\hat{Z}'^*_\tau\right) = \frac{1}{v_\tau}\sum_{\substack{i=\overline{1,v_\tau}\\ j\in J_\tau}} \omega_i \frac{\left|\hat{z}_j^\tau - z'_j\right|}{z'_j}\times 100\,\%, \tag{10}$$

$$E_\tau^4 = E^4\left(Z',\hat{Z}'^*_\tau\right) = \max_{\substack{i=\overline{1,v_\tau}\\ j\in J_\tau}} \omega_i \left|\hat{z}_j^\tau - z'_j\right|. \tag{11}$$

Нехай потрібно оцінити якість *L* моделей прогнозування типу *F*, які були протестовані на ретроспективному ряді. Позначимо через $\hat{Z}'^{\,*}_{\tau,p}$ – оцінювальний ряд, який був отриманий за *i*-ю моделлю при прогнозуванні ретроспективного ряду Z' на τ кроків вперед при $p=\overline{1,L}.$ Тоді оптимальною для кожного $\tau=\overline{1,\theta}$ буде вважатися та модель, якій відповідає мінімальне відхиленням прогнозних значень від реальних $E^j\left(Z',\hat{Z}'^{\,*}_{\tau,p}\right)\to \min,\quad j=\overline{0,4},\quad p=\overline{1,L}.$ Одним з підходів, на основі якого можна врахувати переваги моделей з програмної множини на різних ділянках часового ряду, є побудова комбінованих моделей прогнозування селективного і гібридного типів.

5. Адаптивні комбіновані моделі прогнозування з врахуванням подібностей в ретроспекції часових рядів

Нехай задана програмна множина $\Im_{PS}$ моделей прогнозування $f_1, f_2, \ldots, f_L$, на основі яких для ряду $\{z_i\}_{i=0}^{n}$ в точці n можуть бути побудовані оцінки майбутніх елементів часового ряду

$$\left\{\hat{z}_{n+1}^{p}, \hat{z}_{n+2}^{p}, \ldots, \hat{z}_{n+\theta}^{p}\right\}, \tag{12}$$

де $\hat{z}_{n+\tau}^{p}$ – прогноз, який розраховується в точці n на τ точок вперед за p-ою моделлю, $\tau = \overline{1,\theta}$, $p = \overline{1,L}$.

Необхідно на основі множини $\Im_{PS}$ та ретроспективних значень часового ряду $\{z_i\}_{i=0}^{n}$ побудувати найбільш точну послідовність прогнозних значень

$$\hat{Z} = \left\{\hat{z}_i\right\}_{i=n+1}^{n+\theta} = \left\{\hat{z}_{n+1}, \hat{z}_{n+2}, \ldots, \hat{z}_{n+\theta}\right\}.$$

Нехай для часового ряду $\{z_i\}_{i=0}^{n}$ побудована множина з неопорних історій однакової довжини m. Позначимо цю множину через $\Re$. Неопорними (N, m)-історіями називаються такі ділянки часового ряду $\{z_i\}_{i=0}^{n}$ фіксованої довжини, які будують за методом плинного вікна згідно правила:

$$z_{(m)}^{N} = \left\{z_N, z_{N+1}, \ldots, z_{N+m-1}\right\} = \left\{z_{N+j}\right\}_{j=0}^{m-1},$$

$$N = \overline{0, n-m}, \tag{13}$$

де в позначенні $z_{(m)}^{N}$ верхній індекс визначає точку, з якої починається побудова історії, а нижній індекс в дужках – довжину історії або довжину ділянки часового ряду. Іншими словами, під історією часового ряду розуміються його підпослідовності ретроспективних значень або ретроспекції фіксованої довжини. Детальніше про побудову неопорних і опорних історій описано в роботі [7].

Прогнози, які розраховуються на основі кожної моделі f_p з програмної множини $\Im_{PS}$, $p = \overline{1,L}$ в точках m, m+1,…, n–1, m<n ряду $\{z_i\}_{i=0}^{n}$, які відповідають останнім елементам неопорних історій $z_{(m)}^{N}$ (13), $N = \overline{0, n-m}$, довжини m на $\tau = \overline{1,\theta}$ точок вперед позначимо через:

$$\hat{z}_1^{p}(m+N-1) = f_p\left(\left\{z_{N+j}\right\}_{j=0}^{m-1}\right),$$

$$\hat{z}_2^{p}(m+N-1) = f_p\left(\left\{z_{N+j}\right\}_{j=0}^{m-1}, \hat{z}_1^{p}(m+N-1)\right),$$

$$\vdots$$

$$\hat{z}_{\theta}^{p}(m+N-1)= \\ = f_p\left(\left\{z_{N+j}\right\}_{j=0}^{m-1}, \hat{z}_1^p(m+N-1), \hat{z}_2^p(m+N-1), \ldots, \hat{z}_{\theta-1}^p(m+N-1)\right), \tag{14}$$

де $\hat{z}_{\tau}^{p}(m+N-1)$ – прогнозне значення точки $z_{m+N-1+\tau}$, що слідує після (N, m)-історії, отримане на основі p-ої прогнозної моделі. Запис $f_p\left(\left\{z_{N+j}\right\}_{j=0}^{m-1}\right)$ означає, що для побудови прогнозу за моделлю f_p було використано відповідну неопорну (N, m)-історію, що належить множині неопорних історій, $z_{(m)}^{N} \in \Re$.

Побудуємо опорну (N, m)-історію для вхідного часового ряду $\{z_i\}_{i=0}^{n}$:

$$z_{(m)}^{n-m+1} = \left\{z_{n-m+1}, z_{n-m+2}, \ldots, z_n\right\} = \left\{z_j\right\}_{j=n-m+1}^{n} \tag{15}$$

і на основі деякої міри близькості, наприклад, за допомогою відстані Евкліда за методом найближчого сусіда знайдемо з множини неопорних історій $\Re$ таку, яка є найбільш подібною до опорної.

Неопорна історія $z_{(m)}^{x} \in \Re$ називається найбільш подібною до опорної $z_{(m)}^{n-m+1}$, якщо не існує інших неопорних історій $z_{(m)}^{k}$, $k \in [1, n-m-1]$, $k \neq x$, для яких

$$d\left(z_{(m)}^{n-m+1}, z_{(m)}^{k}\right) < d\left(z_{(m)}^{n-m+1}, z_{(m)}^{x}\right),$$

де d – деяка метрична відстань.

Відповідно до принципів побудови комбінованих моделей є два підходи до розрахунку прогнозів: селективний і гібридний. Розглянемо спочатку звичайні комбіновані моделі селективного і гібридного типів, а потім ці ж моделі, тільки з використанням ідентифікації подібностей.

Адаптивна комбінована селективна за B-критерієм селекції ACSM-B (Adaptive combined selective model)

Нехай $\Im_{PS}$ – програмна множина моделей прогнозування. Селективний підхід полягає у відборі для кожного значення τ з програмної множини моделей $\Im_{PS}$, єдиної моделі, яка забезпечує високу точність прогнозування за певним критерієм селекції: B-критерій, K-критерій [21] тощо. Параметри критеріїв селекції, як правило, мають адаптивний характер. Крім того, часто для підвищення точності прогнозування критерії відбору застосовують не до програмної, а до, так званої, основної множини, яку позначимо через $\Im_{BS}^{\tau}$. Ця множина формується в кожній точці часового ряду в процесі прогнозування і складається з таких моделей, які дають найбільш точні прогнози на поточній ділянці часового ряду, $\Im_{BS}^{\tau} \subseteq \Im_{PS}$. Відбір моделей до основної множини може здійснюватись, наприклад, на основі D-критерію [21]:

$$D_p(\tau) \leq \lambda D_{\min}(\tau), \tag{16}$$

де c – період передісторії, λ – параметр критерію,

$$D_{\min}(\tau) = \min_{p=\overline{1,L}} D_p(\tau),$$

$$D_p(\tau) = \frac{1}{c}\sum_{j=0}^{c}\left(\hat{z}_\tau^p\left(m-\tau-j\right) - z_{m-j}\right)^2.$$

Розрахуємо значення B-критерію для кожної моделі для моменту часу n за формулою:

$$B_{j,\tau}^p = (1-\alpha_B)\cdot B_{j-1,\tau}^p + \alpha_B \cdot \left|\hat{z}_\tau^p\left(j-\tau\right) - z_j\right|, \tag{17}$$

де $B_{j,\tau}^p$ – значення B-критерію, яке розраховане в момент $j=\overline{n-c-1,n}$, c – період передісторії, для кожної моделі f_p, $p=\overline{1,L}$ з множини $\Im_{PS}$ та для кожного значення періоду прогнозу $\tau=\overline{1,\theta}$, α_B – параметр згладжування. Задача визначення оптимального параметру α_B в B-критерії не відрізняється від оцінки параметру згладжування експоненціальної моделі [20].

Порівняємо отримані кінцеві значення $B_{n,\tau}^p$ між собою. Для кожного $\tau=\overline{1,\theta}$ знайдемо мінімальне значення B-критерію $\min\limits_{p=\overline{1,L}} B_{n,\tau}^p$ для кожної з моделей програмної множини. Позначимо через $\tilde{f}^\tau$ – моделі, яким відповідають мінімальні значення B-критеріїв для кожного $\tau=\overline{1,\theta}$, $\tilde{f}^\tau \in \Im_{PS}$. Тоді прогноз в точці n за моделлю ACSM-B розраховується за формулою:

$$\hat{z}_\tau\left(n\right) = \hat{z}_\tau^*\left(n\right), \tag{18}$$

де $\hat{z}_\tau^*(n)$ – прогнози за моделями $\tilde{f}^\tau$, які розраховуються в точці n на $\tau=\overline{1,\theta}$ кроків вперед.

Адаптивна комбінована селективна модель за R-критерієм селекції ACSM-R

Для кожної моделі з програмної множини $f_p \in \Im_{PS}$, $p=\overline{1,L}$ введемо коефіцієнт λ_p. Нехай z_n – остання точка часового ряду Z, в якій виконується селекція моделей. Розглянемо інтервал $[n–c, n]$, де c – період передісторії. В кожній точці цього відрізку відбувається відбір найбільш точних моделей за D-критерієм (16). Якщо модель f_p задовольняє D-критерій, то коефіцієнт моделі f_p – λ_p збільшується на одиницю. Модель, коефіцієнт λ_p якої в момент n максимальний, обирається для розрахунку прогнозу за адаптивною комбінованою моделлю. Якщо існують декілька моделей з однаковими максимальними коефіцієнтами, то обирається та з них, яка отримує приріст коефіцієнта пізніше інших. Після надходження нової точки z_{n+1} і переходу до наступного інтервалу $[n+1–c, n+1]$ всі коефіцієнти λ_p, $p=\overline{1,L}$ обнуляються і розрахунок починається спочатку.

Нехай період прогнозу $\tau=1$, а момент, для якого розраховуються значення D-критерію t_{n-p}. Задано послідовність коефіцієнтів λ_p для кожної моделі f_p з програмної множини, $p=\overline{1,L}$. Будемо вважати, що перед виконанням алгоритму,

λ_p=0. Загальна формула для розрахунку значення зваженого D-критерію в довільний момент t для періоду прогнозу τ – $D_p^t(\tau)$ має вигляд:

$$D_p^t(\tau)=\left(\sum_{k=1}^{c}k\right)^{-1}\cdot\sum_{j=0}^{c-1}(c-j)\cdot\left(\hat{z}_\tau^p(t-\tau-j)-z_{t-j}\right)^2, \tag{19}$$

де c – період передісторії.

Першим кроком алгоритму є розрахунок значення $D_p^t(\tau)$ для кожної моделі $p=\overline{1,L}.$ Відбір з моделей f_p тих, для яких виконується умова

$$D_p^t(\tau)\le\left(1.2+(\tau-1)\cdot h\right)\cdot D_{\min}^t(\tau), \tag{20}$$

де $D_{\min}^t(\tau)=\min\limits_{p}D_p^t(\tau),$ поріг

$$h=\frac{1.9-1.2}{c-1}.$$

Коефіцієнти моделей f_p, для яких виконується умова (20), збільшуються на одиницю: $\lambda_p^t=\lambda_p^t+1,$ λ_p^t – коефіцієнт моделі f_p в момент t. Якщо $r\neq0$, то поставимо $r=r-1$ і перейдемо до першого кроку алгоритму, тобто продовжимо розрахунок D-критерію для моменту $t=n-r-1$. Якщо r=0, то алгоритм закінчено і можна розрахувати прогнозне значення для τ=1. Позначимо через $\tilde{f}^\tau$ – моделі, якій відповідає максимальне значення серед коефіцієнтів $\lambda_p^n,$ для заданого τ, $\tilde{f}^\tau\in\Im_{PS},$ а через $\hat{z}_\tau^*(n)$ – прогноз за цією моделлю, який розраховується в точці n. Прогноз, як і в попередній моделі, визначається за формулою (18).

Якщо відібрано декілька моделей з однаковими максимальними коефіцієнтами, тоді порівнюються значення коефіцієнтів для точок, які передують точці n: n–1, n–2,… Це необхідно з метою знаходження тієї моделі, коефіцієнт якої отримав приріст в момент, найближчий до n.

Обнуляємо коефіцієнти λ_p=0, збільшуємо значення τ, τ=2 і переходимо до першого кроку. Продовжуємо виконання алгоритму до тих пір, поки в момент n не будуть розраховані прогнози для кожного заданого $\tau=\overline{1,\theta}.$

Слід зазначити, що параметр r у даній моделі має бути невеликим (r=3), проте може бути скоригований перед реалізацією прогнозування. Використання даного критерію селекції може збільшувати точність прогнозування в порівнянні з іншими підходами.

Адаптивна комбінована селективна модель за P-критерієм селекції ACSM-P

Розглянемо P-критерій селекції. Нехай z_n – остання точка часового ряду. Як і в R-критерії, розглянемо відрізок [n–c, n] і для $\forall j=\overline{n-c,n}$ розрахуємо величину

$$S_j^p=sign\left(\hat{z}_1^p(j)-z_j\right), \tag{21}$$

де $\hat{z}_1^p(j)$ – прогноз, який розраховується в точці j на одну точку вперед за моделлю f_p, $p=\overline{1,L}$. Тоді за даним критерієм в момент n відбирається та модель, яка приймає максимальне значення η_p:

$$\eta_p \xrightarrow[p=\overline{1,L}]{} \max, \tag{22}$$

де

$$\eta_p = \prod_{j=n-c}^{n} \left(\frac{z_{j+1}}{z_j} \right)^{S_j^p}. \tag{23}$$

За моделлю ACSM-P прогноз розраховується як і в попередніх моделях за формулою (18), де $\hat{z}_\tau^*(n)$ – прогноз на τ кроків вперед в точці n за тією моделлю, для якої величина η_p максимальна.

Адаптивна комбінована гібридна модель ACHM (Adaptive combined hybrid model)

Нехай за формулою (16) було здійснено відбір найбільш точних моделей в точці n. Тобто для кожного τ побудована основна множина $\Im_{BS}^\tau$. Позначимо через $f_1^\tau, f_2^\tau, \ldots, f_{L_B^\tau}^\tau$ – моделі прогнозування, які включені до множини $\Im_{BS}^\tau$, L_B^τ – кількість моделей для кожного фіксованого τ, $L_B^\tau \le L$.

Нехай в точці n розраховані значення $B_{j,\tau}^p$ за формулою (17) для кожної з моделей f_p^τ, $p=\overline{1,L_B^\tau}$ з множини $\Im_{BS}^\tau$ та для кожного значення періоду прогнозу $\tau=\overline{1,\theta}$. Прогноз за моделлю ACHM буде визначатися за формулою:

$$\hat{z}_\tau(n) = \sum_{p=1}^{L_B^\tau} \omega_\tau^p \cdot \hat{z}_\tau^p(n), \tag{24}$$

де вагові коефіцієнти ω_τ^p визначаються на основі значень $B_{j,\tau}^p$, причому виконується умова

$$\sum_{i=1}^{L_B^\tau} \omega_\tau^i = 1.$$

Алгоритм вибору вагових коефіцієнтів в моделі ACHM:

1. Сформуємо матрицю $G=\left(g_{ij}\right)_{i,j=1}^{L-1}$, яка складається з номерів всіх моделей прогнозування програмної множини $\Im_{PS}$, L – кількість моделей. Наприклад, якщо $L=6$, то матриця G матиме вигляд:

$$\mathrm{G} = \begin{pmatrix} 2 & 3 & 4 & 5 & 6 \\ 1 & 3 & 4 & 5 & 6 \\ 1 & 2 & 4 & 5 & 6 \\ 1 & 2 & 3 & 5 & 6 \\ 1 & 2 & 3 & 4 & 6 \end{pmatrix}.$$

2. В точці *n* сформуємо основні множини моделей $\Im_{BS}^{\tau}$ для кожного значення τ. Всі моделі цих множини перенумеровуємо від 1 до L_B^{τ}, де $L_B^{\tau} = card\left(\Im_{BS}^{\tau}\right)$. Розрахунок вагових коефіцієнтів проводиться за формулами:

$$\phi_{j,\tau} = \prod_{j=1}^{L_B^{\tau}} \prod_{i=1}^{L_B^{\tau}-1} B_{n,\tau}^{g_{ij}}, \quad j = \overline{1, L_B^{\tau}}, \quad \tau = \overline{1, \theta}, \tag{25}$$

$$\omega_{\tau}^{p} = \frac{\phi_p}{\sum_{j=1}^{L_B^{\tau}} \phi_j}, \quad p = \overline{1, L_B^{\tau}}, \tag{26}$$

де g_{ij} – елементи матриці G.

Адаптивна комбінована модель селективного типу з ідентифікацією подібностей ACSMwSI (Adaptive combined selective model with similarity identification)

Нехай на основі деякої міри близькості за методом найближчого сусіда було визначено історію $z_{(m)}^{x-m+1} \in \Re, \quad x \in [m, n-\theta-1]$, найбільш подібну до опорної історії $z_{(m)}^{n-m+1}$. Останнім елементом історії $z_{(m)}^{x-m+1}$ буде елемент z_x. В даній точці побудуємо основні множини моделей прогнозування $\Im_{BS}^{\tau}$, використовуючи D-критерій. Тобто отримаємо моделі $f_1^{\tau}, f_2^{\tau}, \ldots, f_{L_B^{\tau}}^{\tau}, \quad L_B^{\tau} \le L$. Детально про вибір подібних історій в роботі [7].

Для кожної моделі з основних множин $\Im_{BS}^{\tau}$ розрахуємо значення B-критерію за формулою:

$$B_{x,\tau}^{q_{\tau}} = (1-\alpha_B) B_{x-1,\tau}^{q_{\tau}} + \alpha_B e_{\tau}^{q_{\tau}}(x-\tau), \tag{27}$$

де $0<\alpha_B\le 1$ – параметр згладжування,

$$e_{\tau}^{q_{\tau}}(x-\tau) = \left| \hat{z}_{\tau}^{q_{\tau}}(x-\tau) - z_x \right| \tag{28}$$

– абсолютна похибка прогнозу, який розраховується в момент $t_{x-\tau}$ на τ кроків уперед за моделями $f_{q_{\tau}}^{\tau}, \quad \tau = \overline{1,\theta}, \quad q_{\tau} = \overline{1, L_B^{\tau}}$.

Позначимо через f^{τ^*} – модель, яка відбирається з основної множини за деяким критерієм селекції для фіксованого значення τ (27), (28). Прогноз, який розраховуються на основі моделей f^{τ^*} в точці *n* на τ кроків вперед позначимо через $\hat{z}_{\tau}^{*}(n)$.

Якщо кожному τ відповідає модель f^{τ^*} з мінімальним значенням B-критерію $\min_{q_{\tau}=\overline{1,L_B^{\tau}}} \left\{ B_{x,\tau}^{q_{\tau}} \right\}$. Тоді прогноз за комбінованою моделлю селективного типу за B-критерієм селекції (27) з використанням ідентифікації подібностей за методом найближчого сусіда буде розраховуватись за формулою:

$$\hat{z}_{\tau}(n) = \alpha\, \hat{z}_{\tau}^{*}(n) + (1-\alpha)\, z_{x+\tau}, \tag{29}$$

де $z_{x+\tau}$ – значення часового ряду, яке слідує після історії $z_{(m)}^{x-m+1}$, найбільш подібної до опорної історії $z_{(m)}^{n-m+1}$, $\alpha \in [0,1]$ – деякий параметр.

Адаптивна комбінована модель гібридного типу з ідентифікацією подібностей ACHMwSI (Adaptive combined hybrid model with similarity identification)

Застосуємо гібридний підхід до побудови прогнозу. Прогноз за гібридним підходом розраховується як зважена сума прогнозів за всіма моделями, які складають основну множину $\Im_{BS}^{\tau}$. Нехай після проведеної ідентифікації подібностей в точці z_x для кожного τ було сформовано основні множини $\Im_{BS}^{\tau}$ і розраховано значення B-критеріїв $B_{x,\tau}^{q_\tau}$, $q_\tau = \overline{1, L_B^{\tau}}$ (2.38). Позначимо через $\hat{z}_{\tau}^{q_\tau}(n)$ прогноз, який розраховується в точці n на τ точок вперед за моделями $f_{q_\tau}^{\tau}$ з основної множини $\Im_{BS}^{\tau}$, $q_\tau = \overline{1, L_B^{\tau}}$, $\tau = \overline{1, \theta}$. Тоді прогноз за комбінованою моделлю гібридного типу з ідентифікацією подібностей за методом найближчого сусіда визначається за формулою:

$$\hat{z}_{\tau}(n) = \alpha \sum_{q_\tau=1}^{L_\tau} \omega_{\tau}^{q_\tau} \, \hat{z}_{\tau}^{q_\tau}(n) + (1-\alpha)\, z_{x+\tau}, \tag{30}$$

де $\alpha \in [0,1]$, ваги $\omega_{\tau}^{q_\tau}$ можуть визначатися на основі B-критерію з врахуванням коефіцієнту пропорційності, який визначається з рівності суми ваг одиниці, $\sum_{q_\tau=1}^{L_B^{\tau}} \omega_{\tau}^{q_\tau} = 1$, тобто за формулами.

Простими словами, в описаних комбінованих моделях прогноз розраховується за відповідним гібридним або селективним принципом з врахуванням ретроспективних значень часового ряду, які слідували після відповідної найбільш подібної історії.

У випадку динамічного прогнозування після розрахунку прогнозу в точці z_n і надходження нової точки z_{n+1} будується нова опорна історія

$$z_{(m)}^{n-m+2} = \{ z_{n-m+2}, z_{n-m+3}, \ldots, z_{n-1}, z_n, z_{n+1} \},$$

а стара опорна історія $z_{(m)}^{n-m+1}$ стає неопорною, тобто включається до множини $\Re$, і процес розрахунку починається спочатку. Тобто знаходиться на основі певної міри близькості найбільш подібна до опорної історія, формуються множини $\Im_{BS}^{\tau}$ для кожного τ, розраховується значення B-критерію для кожної прогнозної моделі, яка використовує найбільш подібну неопорну історію в якості ретроспективної інформації. Далі будується комбінований прогноз.

6. Обговорення результатів дослідження з прогнозування часових рядів на основі адаптивних комбінованих моделей

Для проведення експерименту було обрано часові ряди цін на сировину (щоденні дані) за період з 2014 по 2017 рік (по 900 точок). Розроблені моделі було реалізовано та протестовано на цих часових рядах. В табл. 1–4 вказані результати тестування деяких з відібраних часових рядів: ціни на цинк, мідь, платину, нікель, срібло, алюміній. Дані часові ряди було завантажено та попередньо оброблено з використанням послідовного R/S-аналізу для визначення показника Герста. Детальніше про процедуру фрактального послідовного R/S-аналізу можна ознайомитися в роботі [27]. Причому теоретичний показник, який відповідає істинності основної гіпотези про випадковість даних часових рядів,

$$E\left(\frac{R_m}{S_m}\right) \approx 0.540$$

і розраховувався за формулою [28]:

$$E\left(\frac{R_m}{S_m}\right) = \frac{\Gamma\left(\frac{m-1}{2}\right)}{\sqrt{\pi}\,\Gamma\left(\frac{m}{2}\right)} \sum_{i=1}^{m-1} \sqrt{\frac{m-i}{i}}, \qquad (31)$$

де m – довжина часового ряду, R_m – розмах, а S_m – середньоквадратичне відхилення ділянки вхідного часового ряду Z довжини m. Результати прогнозування вказані в табл. 1–6 (жирним шрифтом з підкресленням виділені мінімальні середні відносні похибки прогнозування для кожного значення τ). Середні відносні похибки були розраховані за формулою (10). Позначення моделей в табл. 1–6: SESM-0, -1, -2 – адаптивні поліноміальні моделі Брауна 0,1 та 2 порядків (Simple exponential smoothing model), ACSM-B, -R, -P – адаптивні комбіновані моделі селективного типу з B, R, P критеріями селекції, ACHM – адаптивна комбінована гібридна модель, ACSMwSI-B, -R – адаптивна комбінована селективна модель з ідентифікацією подібностей в ретроспекції часового ряду та селекцією за B та R критеріями, ACHMwSI – адаптивна комбінована гібридна модель з ідентифікацією подібностей в ретроспекції часового ряду. Програмна множина комбінованих моделей була сформована на основі моделей SESM 0, 1 та 2 порядків з різними параметрами згладжування.

Таблиця 1

Середня відносна похибка прогнозування цін на мідь з горизонтом $\tau = \overline{1,10}$

Моделі \ τ	1	2	3	4	5	6	7	8	9	10
SESM-0	1.683	2.568	3.246	3.785	4.224	4.545	4.826	5.056	5.201	5.258
SESM-1	1.479	2.355	3.068	3.726	4.253	4.654	4.880	5.156	5.469	5.614
SESM-2	1.667	2.353	2.998	3.599	4.135	4.585	4.889	5.186	5.551	5.736
ACSM-B	1.667	2.579	3.330	3.703	4.173	4.615	4.729	5.008	5.390	5.526
ACSM-R	1.670	2.482	3.114	3.566	4.132	4.604	4.856	5.136	5.446	5.631

ACSM-P	1.655	2.349	2.977	3.540	4.075	4.525	4.827	5.124	5.538	5.754
ACHM	**1.351**	**2.270**	**2.966**	3.603	4.101	4.501	4.776	5.045	5.356	5.510
ACSMwSI-B	1.696	2.594	3.057	3.740	4.013	**4.167**	4.502	4.761	**4.941**	5.163
ACSMwSI-R	1.916	2.571	3.129	3.560	3.951	4.231	4.548	4.847	5.008	**4.996**
ACHMwSI	1.491	2.426	2.993	**3.502**	**3.822**	4.168	**4.448**	**4.751**	4.981	5.079

Таблиця 2

Середня відносна похибка прогнозування цін на платину з горизонтом $\tau=\overline{1,10}$

Моделі \ τ	1	2	3	4	5	6	7	8	9	10
SESM-0	1.203	1.783	2.251	2.457	2.715	3.061	3.353	3.570	3.701	3.806
SESM-1	1.190	1.887	2.349	2.545	2.703	3.005	3.400	3.665	3.838	3.935
SESM-2	1.434	2.165	2.578	2.819	2.913	3.066	3.473	3.790	3.975	4.044
ACSM-B	1.370	2.007	2.431	2.650	2.811	3.204	3.567	3.691	3.709	3.764
ACSM-R	1.373	1.870	2.398	2.578	2.920	3.132	3.467	3.674	3.849	3.887
ACSM-P	1.363	2.021	2.426	2.661	2.842	3.103	3.475	3.773	4.006	4.096
ACHM	**1.082**	**1.776**	**2.245**	2.506	2.651	2.987	3.360	3.596	3.766	3.842
ACSMwSI-B	1.840	2.271	2.418	**2.378**	2.559	2.842	**2.992**	**3.099**	**3.178**	3.312
ACSMwSI-R	1.918	2.248	2.605	2.659	2.605	2.812	3.129	3.175	**3.178**	**3.270**
ACHMwSI	1.470	2.117	2.345	2.431	**2.494**	**2.780**	3.022	3.186	3.227	3.306

Таблиця 3

Середня відносна похибка прогнозування цін на нікель з горизонтом $\tau=\overline{1,10}$

Моделі \ τ	1	2	3	4	5	6	7	8	9	10
SESM-0	2.507	3.711	4.366	4.827	5.077	5.360	5.540	5.643	5.813	6.002
SESM-1	2.353	3.707	4.550	5.034	5.411	5.737	5.944	6.160	6.259	6.425
SESM-2	2.692	4.098	4.974	5.503	5.833	6.016	6.194	6.398	6.523	6.752
ACSM-B	2.832	3.837	4.415	5.098	5.642	5.866	5.787	5.887	6.046	6.250
ACSM-R	2.621	3.604	4.525	4.968	5.575	5.639	5.742	6.087	6.232	6.567
ACSM-P	2.627	3.970	4.727	5.334	5.822	6.016	6.134	6.331	6.319	6.570
ACHM	**2.167**	**3.559**	4.332	4.890	5.301	5.599	5.733	5.934	5.995	6.188
ACSMwSI-B	3.025	3.730	**4.273**	4.773	5.102	5.320	5.514	5.620	5.751	5.919

ACSMwSI-R	3.003	3.859	4.569	4.776	**4.847**	**5.076**	**5.193**	**5.213**	5.911	6.106
ACHMwSI	2.337	3.769	4.343	**4.693**	4.980	5.277	5.388	5.551	**5.727**	**5.901**

Таблиця 4
Середня відносна похибка прогнозування цін на срібло з горизонтом $\tau=\overline{1,10}$

τ / Моделі	1	2	3	4	5	6	7	8	9	10
SESM-0	1.705	2.566	3.228	3.747	4.119	4.422	4.682	5.038	5.345	5.615
SESM-1	1.444	2.434	3.144	3.618	4.035	4.349	4.658	4.904	5.153	5.485
SESM-2	1.726	2.589	3.240	3.679	3.993	4.313	4.670	4.986	5.212	5.395
ACSM-B	1.723	2.666	3.093	3.651	4.033	4.276	4.608	5.065	5.271	5.287
ACSM-R	1.620	2.580	3.101	3.578	3.998	4.297	4.683	5.091	5.344	5.399
ACSM-P	1.626	2.500	3.160	3.635	3.922	4.262	4.650	4.952	5.170	5.420
ACHM	**1.311**	2.348	3.079	3.566	3.968	4.284	4.605	4.893	5.211	5.456
ACSMwSI-B	1.593	2.506	3.153	3.562	3.967	4.299	4.569	4.860	5.109	5.202
ACSMwSI-R	1.612	2.457	3.047	3.549	3.892	4.285	4.569	4.860	5.109	5.323
ACHMwSI	1.233	**2.345**	**3.005**	**3.438**	**3.827**	**4.167**	**4.457**	**4.705**	**4.979**	**5.170**

Таблиця 5
Середня відносна похибка прогнозування цін на цинк з горизонтом $\tau=\overline{1,10}$

τ / Моделі	1	2	3	4	5	6	7	8	9	10
SESM-0	2.507	3.711	4.366	4.827	5.077	5.360	5.540	5.643	5.813	6.002
SESM-1	2.353	3.707	4.550	5.034	5.411	5.737	5.944	6.160	6.259	6.425
SESM-2	2.692	4.098	4.974	5.503	5.833	6.016	6.194	6.398	6.523	6.752
ACSM-B	2.832	3.837	4.415	5.098	5.642	5.866	5.787	5.887	6.046	6.250
ACSM-R	2.621	3.604	4.525	4.968	5.575	5.639	5.742	6.087	6.232	6.567
ACSM-P	2.627	3.970	4.727	5.334	5.822	6.016	6.134	6.331	6.319	6.570
ACHM	**2.167**	**3.559**	4.332	4.890	5.301	5.599	5.733	5.934	5.995	6.188
ACSMwSI-B	3.025	3.730	**4.273**	4.773	5.102	5.320	5.514	5.320	5.751	5.919
ACSMwSI-R	3.003	3.859	4.869	5.262	**4.847**	**5.076**	**5.193**	**5.213**	5.911	6.106
ACHMwSI	2.337	3.769	4.343	**4.693**	4.980	5.277	5.388	5.551	**5.727**	**5.901**

Таблиця 6
Середня відносна похибка прогнозування цін на алюміній з горизонтом $\tau = \overline{1,10}$

Моделі \ τ	1	2	3	4	5	6	7	8	9	10
SESM-0	1.538	2.254	2.779	3.190	3.527	3.855	4.266	4.604	4.903	5.224
SESM-1	1.357	2.076	2.630	3.127	3.471	3.807	4.134	4.533	4.884	5.209
SESM-2	1.578	2.217	2.700	3.175	3.531	3.810	4.161	4.496	4.876	5.159
ACSM-B	1.512	2.028	2.556	**3.077**	3.616	4.060	4.450	4.675	4.824	**5.065**
ACSM-R	1.512	2.028	**2.547**	3.220	3.513	4.031	4.287	4.537	4.830	5.179
ACSM-P	1.541	2.238	2.704	3.200	3.546	3.794	4.144	**4.469**	4.852	5.143
ACHM	**1.293**	**2.017**	2.575	3.094	**3.441**	**3.775**	**4.120**	4.476	**4.815**	5.094
ACSMwSI-B	2.213	2.650	3.148	3.429	3.748	4.063	4.333	4.700	5.071	5.167
ACSMwSI-R	2.198	2.708	3.144	3.400	3.684	4.036	4.334	4.682	5.009	5.144
ACHMwSI	1.757	2.554	3.050	3.349	3.644	4.005	4.291	4.618	4.950	5.197

Можна зробити висновок, що у випадку короткострокового прогнозування часових рядів, максимальну точність має адаптивна комбінована гібридна модель з формуванням основної множини $\Im_{BS}$ за D-критерієм з адаптацією значення λ та комбіновані селективні моделі. У випадку прогнозування на середній період $\tau>3$ спостерігається підвищення точності комбінованих моделей з ідентифікацією подібностей. Крім того, чим більш персистентний вхідний часовий ряд, тим точніше модель ACHM. Для менш персистентних рядів, точність прогнозування за моделями ACSMwSI-B, ACSMwSI-R та ACHMwSI найбільша вже для $\tau>3$. Наприклад, показник Герста для часового ряду цін на алюміній складає H=0.818, тобто цей часовий ряд персистентний за результатами проведеного послідовного R/S-аналізу. Для інших часових рядів, результати прогнозування яких вказані в таблицях 1–5, показник Герста коливається в діапазоні значень $H\in[0.570,\ 0.670]$, що свідчить про слабку персистентність цих часових рядів. Причому найнижчий показник Герста спостерігається для часового ряду цін на цинк H=0.570 (табл. 5). Також в результаті дослідження виявлено, що у випадку достатньо персистентних часових рядів, при H>0.75 можна обмежитися звичайними адаптивними комбінованими моделями без ідентифікації подібностей в ретроспекції цих часових рядів.

Описані адаптивні комбіновані моделі з ідентифікацією подібностей можуть застосовуватись для прогнозування як стаціонарних, так і нестаціонарних рівновіддалених часових рядів без пропусків на τ кроків вперед. Виявлено, що більша точність даних моделей спостерігається для $\tau>3$, тобто у випадку середньострокового прогнозування. Для вказаних моделей основною вимогою є достатній об'єм ретроспективної інформації, тобто вхідні ряди мають мати не менше 700 точок.

7. Висновки

Для вирішення задачі прогнозування часових рядів було проведено формалізацію та запропоновані нові адаптивні комбіновані моделі прогнозування нестаціонарних часових рядів, а також розглянуті методи оцінювання точності прогнозування. Результатами дослідження є:

1. Побудовано адаптивні комбіновані селективні моделі прогнозування за B-, R-, P-критеріями селекції з автоматичним формуванням основної множини моделей на основі адаптивного D-критерію (моделі ACSM-B, ACSM-R, ACSM-P). Побудовано адаптивну комбіновану гібридну модель прогнозування часових рядів (модель ACHM).

2. Побудовано адаптивні комбіновані селективні моделі прогнозування за R- та B-критеріями селекції з ідентифікацією подібностей в ретроспекції часових рядів за методом найближчого сусіда (моделі ACSMwSI-R, ACSMwSI-B). Побудовано адаптивну комбіновану гібридну модель прогнозування з ідентифікацією подібностей в ретроспекції часових рядів (модель ACHMwSI).

Адаптивні характеристики комбінованих моделей з ідентифікацією подібностей в ретроспекції часових рядів, в порівнянні зі звичайними комбінованими моделями, дозволяють підвищити точність у випадку середньострокового прогнозування нестаціонарних часових рядів, зокрема фінансових показників. В результаті проведеного експерименту встановлено, що у випадку короткострокового прогнозування для $\tau \le 2$ найбільшу точність має модель ACHM. Моделі ACSM з різними критеріями селекції є ефективними у випадку прогнозування персистентних часових рядів з показником Герста $H>0.75$ для $\tau>2$. У випадку прогнозування часових рядів з показником Герста $E\left(\frac{R_m}{S_m}\right) < H \le 0.75$ для $\tau>2$ більш точними є моделі ACSMwSI з різними критеріями селекції та ACHMwSI.

Література

1. Goldin D. Q., Kanellakis P. C. On similarity queries for time-series data: Constraint specification and implementation // Lecture Notes in Computer Science. 1995. P. 137–153. doi: 10.1007/3-540-60299-2_9

2. Agrawal R., Faloutsos C., Swami A. Efficient similarity search in sequence databases // Lecture Notes in Computer Science. 1993. P. 69–84. doi: 10.1007/3-540-57301-1_5

3. Fast similarity search in the presence of noise, scaling, and translation in time-series databases / Agrawal R., Lin K.-I., Sawhney H. S., Shim K. // VLDB. 1995. P. 490–501.

4. Similarity Measures and Dimensionality Reduction Techniques for Time Series Data Mining / Cassisi C., Montalto P., Aliotta M., Cannata A., Pulvirenti A. //

Advances in Data Mining Knowledge Discovery and Applications. 2012. doi: 10.5772/49941

5. Nayak, R., te Braak P. Temporal pattern matching for the prediction of stock prices // 2nd International Workshop on Integrating Artificial Intelligence and Data Mining (AIDM 2007). 2007. P. 95–103.

6. Singh S. Pattern modelling in time-series forecasting // Cybernetics and Systems. 2000. Vol. 31, Issue 1. P. 49–65. doi: 10.1080/019697200124919

7. Kuchansky A., Biloshchytskyi A. Selective pattern matching method for time-series forecasting // Eastern-European Journal of Enterprise Technologies. 2015. Vol. 6, Issue 4 (78). P. 13–18. doi: 10.15587/1729-4061.2015.54812

8. Berzlev A. A method of increment signs forecasting of time series // Eastern-European Journal of Enterprise Technologies. 2013. Vol. 2, Issue 4 (62). P. 8–11. URL: http://journals.uran.ua/eejet/article/view/12362/10250

9. Perlin M. S. Nearest neighbor method // Revista Eletrônica de Administração. 2007. Vol. 13, Issue 2.

10. Fernández Rodríguez F., Sosvilla Rivero S. J., Andrada Félix J. Nearest-Neighbour Predictions in Foreign Exchange Markets // SSRN Electronic Journal. 2002. doi: 10.2139/ssrn.300404

11. Kahveci T., Singh A. Variable length queries for time series data // Proceedings 17th International Conference on Data Engineering. 2001. doi: 10.1109/icde.2001.914838

12. Li C., Chang E., Garcia-Molina H., Wiederhold G. Clustering for approximate similarity search in high-dimensional spaces // IEEE Transactions on Knowledge and Data Engineering. 2002. Vol. 14, Issue 4. P. 792–808. doi: 10.1109/tkde.2002.1019214

13. Biloshchytskyi A., Kuchansky A., Andrashko Y., Biloshchytska S., Dubnytska A., Vatskel V. The method of the scientific directions potential forecasting in infocommunication systems of an assessment of the research activity results // 2017 4th International Scientific-Practical Conference Problems of Infocommunications. Science and Technology (PIC S&T). 2017. doi: 10.1109/infocommst.2017.8246352

14. Biloshchytskyi A., Kuchansky A., Andrashko Y., Biloshchytska S., Kuzka O., Shabala Y., Lyashchenko T. A method for the identification of scientists' research areas based on a cluster analysis of scientific publications // Eastern-European Journal of Enterprise Technologies. 2017. Vol. 5, Issue 2 (89). P. 4–11. doi: 10.15587/1729-4061.2017.112323

15. Biloshchytskyi A., Kuchansky A., Andrashko Y., Biloshchytska S., Kuzka O., Terentyev O. Evaluation methods of the results of scientific research activity of scientists based on the analysis of publication citations // Eastern-European Journal of Enterprise Technologies. 2017. Vol. 3, Issue 2 (87). P. 4–10. doi: 10.15587/1729-4061.2017.103651

16. Otradskaya T., Gogunskii V., Antoshchuk S., Kolesnikov O. Development of parametric model of prediction and evaluation of the quality level of educational institutions // Eastern-European Journal of Enterprise Technologies. 2016. Vol. 5, Issue 3 (83). P. 12–21. doi: 10.15587/1729-4061.2016.80790

17. Biloshchytskyi A., Myronov O., Reznik R., Kuchansky A., Andrashko Yu., Paliy S., Biloshchytska S. A method to evaluate the scientific activity quality of HEIs based on a scientometric subjects presentation model // Eastern-European Journal of Enterprise Technologies. 2017. Vol. 6, Issue 2 (90). P. 16–22. doi: 10.15587/1729-4061.2017.118377

18. Lizunov P., Biloshchytskyi A., Kuchansky A., Biloshchytska S., Chala L. Detection of near dublicates in tables based on the locality-sensitive hashing method and the nearest neighbor method // Eastern-European Journal of Enterprise Technologies. 2016. Vol. 6, Issue 4 (84). P. 4–10. doi: 10.15587/1729-4061.2016.86243

19. Biloshchytskyi A., Kuchansky A., Biloshchytska S., Dubnytska A. Conceptual model of automatic system of near duplicates detection in electronic documents // 2017 14th International Conference The Experience of Designing and Application of CAD Systems in Microelectronics (CADSM). 2017. doi: 10.1109/cadsm.2017.7916155

20. Vercellis C. Business intelligence: data mining and optimization for decision making. Cornwall: John Wiley & Sons, 2009. 417 p. doi: 10.1002/9780470753866

21. Lukashin Yu. P. Adaptive methods of near-term time series forecasting: manual. Moscow: Finance and Statistics, 2003. 416 p.

22. Snytyuk V. E. Forecasting. Models. Methods. Algorithms: manual. Kyiv: Maklaut, 2008. 364 p.

23. Mulesa O., Geche F., Batyuk A., Buchok V. Development of combined information technology for time series prediction // Advances in Intelligent Systems and Computing. 2017. Vol. 689. P. 361–373. doi: 10.1007/978-3-319-70581-1_26

24. Mulesa O., Geche F., Batyuk A. Information technology for determining structure of social group based on fuzzy c-means // 2015 Xth International Scientific and Technical Conference "Computer Sciences and Information Technologies" (CSIT). 2015. doi: 10.1109/stc-csit.2015.7325431

25. Mulesa O., Geche F. Designing fuzzy expert methods of numeric evaluation of an object for the problems of forecasting // Eastern-European Journal of Enterprise Technologies. 2016. Vol. 3, Issue 4 (81). P. 37–43. doi: 10.15587/1729-4061.2016.70515

26. Morozov V., Kalnichenko O., Liubyma I. Managing projects configuration in development distributed information systems // 2017 2nd International Conference on Advanced Information and Communication Technologies (AICT). 2017. doi: 10.1109/aiact.2017.8020088

27. Peters E. E. Fractal market analysis: applying chaos theory to investment and economics. John Wiley & Sons Inc., 1994. 336 p.

28. Anis A., Lloyd E. The expected value of the adjusted rescaled Hurst Range of independent normal summands. Biometrika. 1976. Vol. 63, Issue 1. P. 111–116. doi: 10.2307/2335090