



НОВІ ЗАСОБИ КІБЕРНЕТИКИ, ІНФОРМАТИКИ, ОБЧИСЛЮВАЛЬНОЇ ТЕХНІКИ ТА СИСТЕМНОГО АНАЛІЗУ

В.С. САЖЕНЮК, Г.О. ЧОРНОУС, Ю.А. ЯРМОЛЕНКО

УДК 330. 4:519. 622. 1:51-77

ІНФОРМАЦІЙНА МОДЕЛЬ ЦІНОУТВОРЕННЯ НА ЕЛЕКТРОННИХ РИНКАХ

Анотація. Розроблено підхід до моделювання та прогнозування ціни ринкового блага на основі змін інформаційного фактору з використанням апарату диференціальних рівнянь першого порядку. Побудовано інформаційну модель ціноутворення шляхом розв'язання задачі Коші. Показано, що запропонована модель може бути корисною як для прогнозування цін активів на фондовій та валютній біржі, так і на товарних електронних ринках. Представлено процес підготовки даних для реалізації моделі.

Ключові слова: інформація, задача Коші, метод штрафу, різницева схема, інформаційний фактор, модель ціноутворення, прогнозування цін.

ВСТУП

У більшості моделей для прогнозування цін фінансових інструментів використовують історичні дані для виявлення закономірностей часових рядів, виділення так званих технічних фігур на графіках, та низку інших підходів, що надають змогу на основі ретроспективного аналізу визначити подальші коливання цін. Зазвичай такі моделі демонструють низьку ефективність, коли на ринку з'являється несподівана інформація, що може сколихнути «передбачуваний» сценарій руху вартості інструменту. Проте, як показує практика, економічні гравці, а відповідно й ціни, є чутливими до важливих новин (зміни інформаційного фактору). Слід також зазначити, що, незалежно від того, достовірною є деяка новина чи ні, вона все одно має вплив на формування ціни, адже може бути помилково сприйнята економічними суб'єктами як правдива (особливо від впливових джерел).

Відповідно до гіпотези ефективного ринку, вартість активу формується на основі усієї наявної на ринку у певний момент часу інформації про нього. Виходячи з цього, спрогнозувати його ціну можна, якщо проаналізувати всю наявну про нього інформацію на ринку. Те саме можна зауважити стосовно будь-якого економічного блага (товару, послуги) на електронних ринках.

Отже, актуальною є потреба в розробленні і практичній реалізації прогностичних моделей ціноутворення економічних благ, що використовували б не лише логіку минулих подій, яка міститься в історичних даних, а й найактуальнішу інформацію про зміни інформаційного фону. Нинішня ситуація у сфері інформаційних технологій та аналізу даних сприяє цьому. По-перше, глобальна інформатизація суспільства призвела до того, що постійно генерується величезна кількість відкритих даних, які можна і необхідно аналізувати

© В.С. Саженьюк, Г.О. Черноус, Ю.А. Ярмоленко, 2020

з метою виявлення оцінок економічної реальності. По-друге, найсучасніші інструменти аналізу різних типів даних надають змогу в автоматичному режимі перетворити характеристики відкритих даних у кількісні показники та активно використовувати математичне моделювання для прийняття більш раціональних та дієвих рішень.

Багато уваги в цьому напрямі вже приділено у дослідженнях щодо ціноутворення на фондовому і валютних ринках [1–4]. Запропоновано до використання низку різновидів моделей ціноутворення на ринку нафти, природного газу та ін. [5–11]. Проте існує певний дефіцит підходів, що можуть бути реалізовані на довільних електронних ринках.

Останні дослідження демонструють, що різні статистичні, математичні, економетричні підходи є досить ефективними під час моделювання та прогнозування цін на основі змін інформаційного фактору: стохастичні диференціальні рівняння [12], метод випадкового пошуку [13], ARIMA моделі [14], методи оброблення сигналів [15] тощо. До того ж, моделі можна значно покращити, якщо застосувати як один із зовнішніх факторів ціноутворення дані Google Trends [1, 14, 16, 17] та Twitter [2, 17].

Метою цього дослідження є розроблення альтернативного підходу до моделювання та прогнозування ціни ринкового блага на основі змін інформаційного фактору з використанням апарату диференціальних рівнянь та визначення практичної цінності цього підходу в разі його застосування на довільних електронних ринках.

МАТЕМАТИЧНА МОДЕЛЬ ЦІНОУТВОРЕННЯ НА ОСНОВІ ЗМІН ІНФОРМАЦІЙНОГО ФАКТОРУ

Позначимо $u = u(t)$ ціну активу залежно від часу t , $t \in [t_0, T]$. $I = I(t)$ — інформація про актив, що має вплив на ціну, $\Omega_t = [t_0, T]$.

Будемо вважати, що попит та пропозиція збалансовані і не мають впливу на ціну, а ціна залежить від наявності інформації.

Розглянемо таку задачу:

$$\frac{du(t)}{dt} = f(t, I(t)), \quad t \in (t_0, T], \quad (1)$$

$$u(t_0) = u_0, \quad (2)$$

$$u(t) > 0, \quad t \in (t_0, T]. \quad (3)$$

Модель ціноутворення активу (1)–(3), що ґрунтується на наявній інформації на ринку, математично являє собою задачу Коші для звичайного диференціального рівняння першого порядку, з додатковим обмеженням на шукану функцію в середині області. Отже, пошук розв'язку цієї задачі здійснюється на замкненій, опуклій множині $K = \{v \in V, v > 0\}$, де V — функціональний простір. Наявність та регулярність єдиного розв'язку задачі (1)–(3) залежить від правої частини співвідношення (1). Зокрема, якщо $f(t, I) \in V = L_2(\Omega_t)$, де $(L_2(\Omega_t))$ — функціональний простір функцій, інтегрованих з квадратом, то відомо [18], що існує єдиний розв'язок $u(t)$ задачі (1)–(3), який належить простору Соболева $W_2^1(\Omega_t)$.

Застосуємо до задачі (1)–(3) метод штрафу, а саме розглянемо таку задачу.

Знайти $u_\varepsilon(t)$ таке, що

$$\frac{du_\varepsilon(t)}{dt} - \frac{1}{\varepsilon} u_\varepsilon^- = f(t, I(t)), \quad t \in (t_0, T], \quad (4)$$

$$u_\varepsilon(t_0) = u_0, \quad (5)$$

де $u_\varepsilon^- = \frac{1}{2} u_\varepsilon \times (\text{sign}(u_\varepsilon) - 1)$, $\varepsilon > 0$.

Задача зі штрафом (4)–(5) має єдиний розв’язок [19, с. 27]. Крім того, розв’язок цієї задачі збігається для $\varepsilon \rightarrow 0$ до розв’язку задачі (1)–(3). При цьому має місце оцінка

$$\|u - u_\varepsilon\|_C \leq M \cdot \sqrt{\varepsilon}, \quad (6)$$

де константа $M > 0$ і не залежить від ε, τ . Доведемо нерівність (6).

Помножимо співвідношення (4) на функцію u_ε^- та проінтегруємо від t_0 до T :

$$-\int_{t_0}^T \frac{du_\varepsilon(t)}{dt} u_\varepsilon^-(t) dt + \frac{1}{\varepsilon} \int_{t_0}^T (u_\varepsilon^-(t))^2 dt = -\int_{t_0}^T f(t, I(t)) \cdot u_\varepsilon^-(t) dt.$$

Звідси, використовуючи нерівність Коші-Буняковського, після нескладних перетворень, отримаємо оцінку

$$\|u_\varepsilon^-\|_0 \leq \varepsilon \|f(t, I(t))\|_0. \quad (7)$$

Тут застосовано такі позначення:

$\|\cdot\|_C$ — норма у функціональному просторі $C[t_0, T]$ неперервних функцій;

$\|\cdot\|_0$ — норма у функціональному просторі $L_2[t_0, T]$ функцій, інтегрованих з квадратом.

Позначимо $w = u - u_\varepsilon$. Для похибки методу штрафу w маємо задачу

$$\frac{dw(t)}{dt} + \frac{1}{\varepsilon} u_\varepsilon^- = 0, \quad t \in \Omega_t, \quad (8)$$

$$w(t_0) = 0. \quad (9)$$

Оскільки $u_\varepsilon^- \geq 0$ і $u \geq 0$, то з (8) випливає нерівність

$$\frac{dw}{dt} \cdot u \leq 0.$$

Але $u = w + u_\varepsilon = w + u_\varepsilon^+ - u_\varepsilon^-$, $u_\varepsilon^+ = \frac{1}{2} u_\varepsilon \times (\text{sign}(u_\varepsilon) + 1)$. Це означає, що

$$\frac{dw}{dt} \cdot w \leq \frac{dw}{dt} \cdot u_\varepsilon^- - \frac{dw}{dt} \cdot u_\varepsilon^+. \quad (10)$$

Помножимо (8) на u_ε^+ . Враховуючи, що $u_\varepsilon^+ \cdot u_\varepsilon^- = 0$, отримаємо

$$\frac{dw}{dt} \cdot u_\varepsilon^+ = 0.$$

З нерівності (10) і нерівності $\frac{dw}{dt} \leq 0$ випливає нерівність

$$\frac{dw}{dt} \cdot w \leq u_\varepsilon^-.$$

Проінтегруємо останню нерівність від t_0 до $\xi \forall \xi \in (t_0, T)$, отримаємо

$$\int_{t_0}^{\xi} \frac{dw(t)}{dt} w(t) dt \leq \int_{t_0}^{\xi} u_\varepsilon^-(t) dt.$$

Але $w(t_0) = 0$ та

$$\int_{t_0}^{\xi} \frac{dw(t)}{dt} w(t) dt = \frac{1}{2} w^2(\xi) \quad \forall \xi \in (t_0, T).$$

Використовуючи нерівність Коші–Буняковського та оцінку (7), отримуємо

$$\begin{aligned} w^2(\xi) &\leq 2 \cdot \int_{t_0}^{\xi} u_{\varepsilon}^{-}(t) dt \leq 2 \cdot \int_{t_0}^T u_{\varepsilon}^{-}(t) dt \leq 2 \cdot \|u_{\varepsilon}^{-}\|_0 \cdot \sqrt{T-t_0} \leq \\ &\leq 2\varepsilon \|f(t, I(t))\|_0 \cdot \sqrt{T-t_0}. \end{aligned}$$

Звідси можна отримати оцінку

$$|w(\xi)| \leq \sqrt{2\varepsilon \|f\|_0} (T-t_0)^{\frac{1}{4}}.$$

Виберемо $\xi = \max w(\xi)$, $\xi \in (t_0, T)$, і позначимо $M = \sqrt{2\|f\|_0} (T-t_0)^{\frac{1}{4}}$.

Отримаємо

$$\|w(t)\|_C \leq M \cdot \sqrt{\varepsilon}.$$

Оцінку (6) доведено.

Далі позначимо

$$\begin{aligned} \omega_{\tau} &= \left\{ t = t_j = j\tau, j=1, N; \tau = \frac{T}{N} \right\}, \quad \|y\|_* = \left\{ \max_{t \in \omega_{\tau}} \|y(t)\|^2 + \tau \|y_t\|_{L_2(\omega_T)}^2 \right\}^{\frac{1}{2}}, \\ g_{\varepsilon}(y) &= -\frac{1}{\varepsilon} y_{\varepsilon}^{-}, \quad P_0 v(t) = \frac{1}{\tau} \int_{t-\tau}^t v(\xi) d\xi, \quad t = \tau, \dots, N\tau. \end{aligned}$$

Задачу (4)–(5) апроксимуємо такою різницевою схемою:

$$y_{\bar{t}} + g_{\varepsilon}(y) = P_0 f(\cdot, I(\cdot)), \quad t \in \omega_T, \quad (11)$$

$$y(t_0) = u_0. \quad (12)$$

Легко бачити, що функція $g_{\varepsilon}(y)$ монотонна та Ліпшиц-неперервна, тобто

$$\forall x, y \in V \Rightarrow 0 \leq (g_{\varepsilon}(x) - g_{\varepsilon}(y)) \cdot (x - y) \leq (x - y)^2. \quad (13)$$

У разі виконання умови (13) різницева схема (11)–(12) є стійкою та має єдиний розв'язок на множині сіткових функцій [20]. Отримаємо оцінку швидкості збіжності. Позначимо похибку $v = P_0 u_{\varepsilon}(\cdot) - y$. Застосуємо до рівняння (4) оператор P_0 . Отримаємо

$$(u_{\varepsilon})_{\bar{t}} + P_0 g_{\varepsilon}(u_{\varepsilon}) = P_0 f(\cdot, I(\cdot)), \quad t \in \omega_T. \quad (14)$$

Віднімемо від (14) рівність (11). Отримаємо для похибки задачу

$$\begin{aligned} (z)_{\bar{t}} + g_{\varepsilon}(P_0 u_{\varepsilon}) - g_{\varepsilon}(y) &= g_{\varepsilon}(P_0 u_{\varepsilon}) - P_0 g_{\varepsilon}(u_{\varepsilon}) - (P_0 u_{\varepsilon})_{\bar{t}} + (u_{\varepsilon})_{\bar{t}}, \quad t \in \omega_T, \\ z(t_0) &= 0. \end{aligned}$$

Позначимо $\eta = (g_{\varepsilon}(P_0 u_{\varepsilon}) - P_0 g_{\varepsilon}(u_{\varepsilon}))$, $\mu = -(P_0 u_{\varepsilon})_{\bar{t}} + (u_{\varepsilon})_{\bar{t}}$.

Задача для похибки різницевої схеми має вигляд

$$(z)_{\bar{t}} + g_{\varepsilon}(P_0 u_{\varepsilon}) - g_{\varepsilon}(y) = \eta + \mu, \quad t \in \omega_T, \quad (15)$$

$$z(t_0) = 0. \quad (16)$$

Помножимо (15) скалярно на $z(t)$, скориставшись нерівністю (13), тотожністю $z \cdot z_{\bar{i}} = \frac{1}{2} \cdot (z^2)_{\bar{i}} + \frac{\tau}{2} \cdot z_{\bar{i}}^2$ та оцінками [20] $\|u\|_{L_2(\omega_T)} \leq M \cdot \tau \|u_\varepsilon\|_{W_2^1(t_0, T)}$, $\|\eta\|_{L_2(\omega_T)} \leq M \cdot \frac{\tau}{\varepsilon} \|u_\varepsilon\|_{W_2^1(t_0, T)}$. Отримаємо оцінку

$$\|z\|_* \leq M \cdot \left(\sqrt{\tau} + \frac{\tau}{\sqrt{\varepsilon}} \right) \|u_\varepsilon\|_{W_2^1(t_0, T)}.$$

Враховуючи оцінку (6) та використовуючи останню оцінку, можна отримати оцінку швидкості збіжності розв'язку різницевої схеми (11)–(12) до розв'язку задачі (1)–(2):

$$\|\bar{u} - y\|_{W_2^1(\omega_T)} \leq M \cdot \left(\sqrt{\tau} + \frac{\tau}{\sqrt{\varepsilon}} + \sqrt{\varepsilon} \right) \|f(t)\|_{L_2(t_0, T)}, \text{ де } \bar{u} = P_0(u(\cdot)).$$

Максимальна швидкість збіжності має порядок $\sqrt{\tau}$ для $\varepsilon = \tau$.

Скористаємося різницевою схемою (11)–(12) для прогнозування ціни. Позначимо $U(t)$ прогнозовану ціну. Після нескладних перетворень з (11) отримаємо для прогнозу ціни таку формулу:

$$U(t) = U(t - \tau) - \tau \cdot g_\varepsilon(t) + \tau \cdot P_0 f(\cdot, I(\cdot)),$$

або

$$U(t) = U(t - \tau) + \frac{\tau}{\varepsilon} U^-(t) + P_0 f(\cdot, I(\cdot)) \times \tau, \quad (17)$$

$$U(t_0) = u_0. \quad (18)$$

Покладемо $\varepsilon = \tau$. З практичного погляду це доцільно, наприклад, коли потрібно спрогнозувати рух ціни на фондовому або валютному ринках відразу після оголошення фундаментальних показників, що мають суттєве значення для ціни активу. Тоді зазвичай відбувається сильний рух ціни за малий проміжок часу. Співвідношення (17)–(18) набувають вигляду

$$U(t) = U(t - \tau) + U^-(t) + P_0 f(\cdot, I(\cdot)) \times \tau, \quad (19)$$

$$U(t_0) = u_0. \quad (20)$$

Будемо називати отриману модель (19)–(20) інформаційною моделлю ціноутворення (ІМЦ). Зауважимо, що формула (19) відносно $U(t)$ є нелінійною, тому потрібно використовувати ітераційні методи для розв'язання (19)–(20) та прогнозування. Певною альтернативою може стати така формула для прогнозу, яка випливає з (17):

$$U(t) = U^+(t - \tau) + P_0 f(\cdot, I(\cdot)) \times \tau. \quad (21)$$

АДАПТАЦІЯ МОДЕЛІ ПРОГНОЗУВАННЯ ЦІН АКТИВІВ ДЛЯ РІЗНИХ РИНКІВ

Формули (19) та (21) можуть виявитися корисними для трейдерів, які торгують на фондовій або валютній біржі за стратегією «на новинах», для визначення точок виходу з позиції. В цьому випадку праву частину співвідношення (1) моделі (1)–(3) можна вибрати у такий спосіб.

Позначимо $I_b(t)$ кількість інформації про можливий ріст ціни активу, а $I_s(t)$ — кількість інформації про можливе падіння ціни активу.

Покладемо в (1)

$$f(t, I) = \alpha \cdot (I_b(t) - I_s(t)),$$

де α — коефіцієнт чутливості, який характеризує чутливість активу до зміни інформаційного фактору (швидкість зміни ціни залежно від зміни кількості інформації). Коефіцієнт α залежить від конкретної задачі.

Якщо інформація являє собою оцінки експертів щодо росту або зниження ціни активу, тобто відомі оцінки ймовірностей $p_b(t)$ — зростання та $p_s(t)$ — зниження ціни, і $p_b + p_s = 1$, то кількість інформації $I_b = -\log_2 p_b$ і $I_s = -\log_2 p_s$ відповідно. Тоді

$$f(t, I) = \alpha \cdot (I_b(t) - I_s(t)) = \alpha \cdot \log_2 \frac{p_s}{p_b}.$$

Якщо до того ж здійснено оцінку ймовірності $p_0(t)$ події, яка полягає в тому, що ціна активу не буде змінюватися впродовж певного проміжку часу, то $I_0 = -\log_2 p_0$ — відповідна інформація, $p_b + p_s + p_0 = 1$.

Функцію $f(t, I)$ можна взяти у вигляді

$$f(t, I) = \begin{cases} \alpha (I_b - I_s), & I_0 < Z, \\ 0, & I_0 \geq Z, \end{cases}$$

де Z — деяке граничне значення.

Розглянемо ще один підхід до вибору правої частини співвідношення (1).

Покладемо

$$p_b = p_b(t, u(t-\tau)) = \frac{\max u(\xi) - u(t-\tau)}{\max u(\xi) - \min u(\xi)},$$

$$p_s = p_s(t, u(t-\tau)) = \frac{u(t-\tau) - \min u(\xi)}{\max u(\xi) - \min u(\xi)},$$

$$\xi \in [t-\tau, t].$$

Очевидно, що $p_b + p_s = 1$. У цьому випадку

$$\begin{aligned} f(t, I, u) &= \alpha \cdot (I_b(t, u) - I_s(t, u)) = \\ &= \alpha \cdot \log_2 \frac{p_s(u)}{p_b(u)} = \alpha \cdot \log_2 \frac{u(t-\tau) - \min u(\xi)}{\max u(\xi) - u(t-\tau)}. \end{aligned} \quad (22)$$

Зауважимо, що $p_b = p_s = \frac{1}{2}$ для

$$u(t) = u_*(t) = \frac{\min u(\xi) + \max u(\xi)}{2},$$

і $f(\xi, I, u_*) = 0$, $\forall \xi \in [t-\tau, t]$.

Розглянемо ще один приклад:

$$p_b = p_b(t, u(t-\tau)) = e^{-\frac{\max u(\xi) - u(t-\tau)}{\max u(\xi) - \min u(\xi)}},$$

$$p_s = p_s(t, u(t-\tau)) = e^{\frac{\min u(\xi) - u(t-\tau)}{\max u(\xi) - \min u(\xi)}}.$$

Тоді

$$f(t, I) = \alpha \cdot (I_b - I_s) = \log_2 e \cdot \alpha \cdot \left[\frac{\max u(\xi) - u(t-\tau)}{\max u(\xi) - \min u(\xi)} - \frac{u(t-\tau) - \min u(\xi)}{\max u(\xi) - \min u(\xi)} \right]. \quad (23)$$

Точка u_* також є «точкою інформаційної рівноваги»: $f(\xi, I, u_*) = 0$. Вираз у квадратних дужках являє собою популярний у технічному аналізі індикатор Чайкіна [21].

Ще одним прикладом використання моделі є прогнозування цін активів на ринку нафти. Для цього насамперед потрібно виділити інформаційні потоки, які можуть мати вплив на ціну нафти. Приміром, це може бути інформація про зміну запасів нафти в США, зміну об'ємів видобутку нафти в державах ОПЕК, зміну санкцій відносно Ірану, технічні характеристики ціноутворення на ф'ючерсному ринку тощо. Тоді права частина співвідношення (1) матиме такий вигляд:

$$f(t, I(t)) = \sum_{k=1}^n f_k(t, I_k).$$

Функції $f_k(t, I_k)$ будуються окремо для кожного інформаційного потоку. Зокрема, для інформаційних потоків щодо зміни об'ємів видобутку нафти в державах ОПЕК, зміни санкцій відносно Ірану, можна використовувати експертні оцінки та відповідні формули, представлені вище; для інформації щодо технічних характеристик ціноутворення на ф'ючерсному ринку — формули (22), (23). Зауважимо, що задача побудови правої частини співвідношення (1) для відповідних інформаційних потоків у кожному конкретному випадку потребує індивідуального розв'язування.

ЕФЕКТИВНІСТЬ МОДЕЛІ ДЛЯ ПРОГНОЗУВАННЯ ЦІН НА ЕЛЕКТРОННИХ ТОВАРНІХ РИНКАХ

Для підтвердження теоретичної цінності та практичної спроможності запропонованої ІМЦ застосуємо результати дослідження до певного електронного товарного ринку, для якого інформація є суттєвим фактором становлення цін попиту та пропозиції. Для прикладу розглянемо ринок вторинного продажу брендових товарів — одягу та аксесуарів, а саме, жіночих брендових сумок. Припустимо, що цей ринок є чутливим до новин (змін інформаційного фактору), адже індустрія моди загалом є досить мінливою та схильною до впливу трендів. Для порівняння рівня ефективності запропонованої моделі відносно інших підходів застосуємо економетричний метод, що добре зарекомендував себе у низці досліджень [14, 22] щодо прогнозування цін на ринках електронної комерції, а саме модель ARIMAX.

Для практичної реалізації обох моделей необхідно здійснити низку важливих кроків, ключові моменти яких внаслідок їхньої практичної значущості описано далі більш детально. Ці кроки є такими:

1. Збір необхідних даних щодо рівня ринкових цін товарів.
2. Оброблення даних та зведення їх до часових рядів.
3. Збір даних щодо популярності пошукових запитів, що стосуються відповідних товарів і можуть бути використані як інформаційний фактор.
4. Побудова моделей ARIMAX та ІМЦ і порівняння їхньої прогностичної здатності.

Реалізація кроку 1. Процес збору та оброблення інформації для цього дослідження має декілька «вузьких місць». По-перше, потрібно створити інструмент для збору даних, що даватиме якнайменше спотворення у представленні кінцевих даних та буде правильно розпізнавати ознаки окремих товарів. По-друге, етап очищення даних має бути дуже точним аби зменшити варіацію спостережень та мінімізувати вплив викидів на результативні показники. Особливо важливим це є за умови невеликих вибірок.

З метою проведення дослідження щодо впливу інформації на формування цін на відкритому ринку було обрано сайт перепродажу брендових товарів

Tradesy.com [23]. Для дослідження було відібрано 5 моделей сумок всесвітньо відомих брендів: Chanel Classic Flap у розмірі Medium; Saint Laurent Sac de Jour у розмірі Small; Gucci Sylvie; Louis Vuitton Alma у розмірі Pm; Fendi Kan I.

Моделі були відібрані у такий спосіб: деякі з них є більш «класичними», тобто очікуються рівномірні розподіли пропозиції та попиту впродовж тривалого періоду (Chanel, Louis Vuitton), а деякі є такими, що тримаються на піку популярності у певні сезони.

Для отримання даних щодо середньої ціни товару певної моделі потрібно було зібрати усі актуальні пропозиції з продажу товару на різні моменти часу. Збір даних здійснювався упродовж лютого-квітня 2019 року за допомогою інструмента для парсингу веб-сайтів ParseHub [24]. Щодня в один і той самий час дані збирали з усіх сторінок за ключовими запитами щодо відповідних моделей. Для кожної активної позиції були зібрані такі характеристики: ідентифікатор товару (Id); назва моделі; колір; матеріал виробу; розміри; поточна ціна; початкова ціна; кількість збережень в «улюблених»; стан виробу; роздрібна ціна нового виробу.

Реалізація кроку 2. Для подальшого використання зібраних даних для моделювання ціноутворення необхідно реалізувати трансформацію і фільтрацію даних. Початкова трансформація потребує «видобування» кількісних та якісних характеристик товарів із текстових описів товарів. Оскільки під час збору даних випадкові продукти могли потрапити до списку за пошуковим запитом, необхідно було виконати розвідувальний аналіз та очистити дані від «викидів». Окрім видалення товарів, що випадковим чином потрапили до результатів пошукового запиту, потрібно було здійснити аналіз за параметрами, що можуть суттєво вплинути на кінцеву вартість окремої позиції. Наприклад, матеріал виробу є визначним фактором для ціноутворення. Щоб забезпечити «однорідність» товарів у вибірці, слід видалити позиції з лімітованих колекцій, для моделі Classic Flap — це вироби з тканини.

Після етапу очищення даних було проведено агрегацію отриманих результатів. Як результативну величину обрано середню ціну пропозиції всіх товарів, виставлених на продаж у певний момент часу.

На рис. 1 продемонстровано середні ціни пропозиції моделі Classic Flap за період лютий-квітень 2019 року.

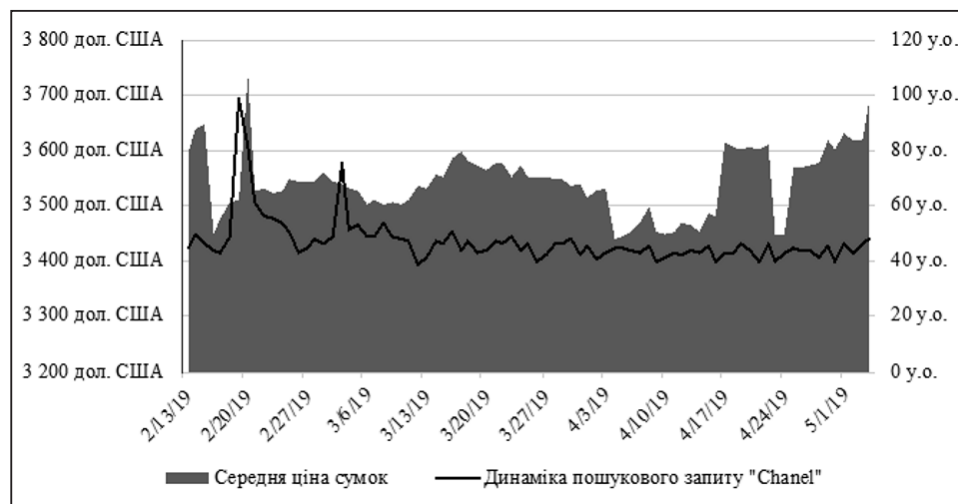


Рис. 1. Середні ціни на модель Chanel Classic Flap Medium (головна вісь) та динаміка пошукового запиту «Chanel» у системі Google Trends (другорядна вісь)

Реалізація кроку 3. Як інформаційний фактор досліджуваної системи розглянуто тренди пошукових запитів у системі Google за відповідні періоди часу. Часові ряди трендів мали явно виражену тижневу періодичність, яку було усунуто на початковому етапі.

Реалізація кроку 4. Для вибору найкращої функціональної форми моделей ARIMAX було використано метод перебору значень (p, d, q, l) у специфікації $AR(p)I(d)MA(q)X(l)$. Верхня межа значень p та q обмежена відповідними оптимальними значеннями P та Q , визначеними на основі візуального аналізу вибіркової автокореляційної (ACF) і часткової автокореляційної (PACF) функцій. Максимальна величина лагового значення фактору інформації, яке дорівнює одиниці, зафіксоване на рівні три (дні). Критерієм відбору найкращої моделі є мінімізація значення байєсівського інформаційного критерію (критерію Шварца) [25]:

$$BIC = \ln(\sigma^2) + \frac{(p+q+1)\ln(n)}{n},$$

де σ^2 — очікувана вибіркова дисперсія, n — величина вибірки. Доцільно використовувати цей критерій через те, що він надає змогу порівнювати між собою моделі з різною кількістю регресорів.

Під час побудови запропонованої у цьому дослідженні ІМЦ теоретичну формулу (19) було зведено до такого вигляду:

$$U(t) = U^+(t-\tau) + f(I(t), t) \times \tau,$$

$$f(I(t), t) = \alpha_0 \cdot \Delta I(t) + \alpha_1,$$

де $U(t)$ — ціна товару у момент часу t , $\tau \rightarrow 0$ — період, за який досліджується зміна ціни, $f(I(t), t)$ — інформаційна функція та $\Delta I(t)$ — зміна інформаційного фону.

Оскільки $\tau \rightarrow 0$, покладемо мінімальне доступне значення $\tau = 1$ (один день) та отримаємо остаточний варіант моделі:

$$U(t) = U^+(t-1) + \alpha_0 \cdot \Delta I(t) + \alpha_1.$$

Оскільки $U(t) > 0$, отримуємо

$$\Delta U(t) = \alpha_0 \cdot \Delta I(t) + \alpha_1.$$

Використовуючи метод найменших квадратів (МНК), знаходимо оцінку коефіцієнтів моделі α_0, α_1 .

Для порівняння прогностичної здатності моделі ARIMAX та запропонованої у цьому дослідженні ІМЦ розрахуємо відповідні значення BIC. Інформаційні критерії двох моделей є порівнюваними між собою, адже $U(t) = 1 \cdot U(t-1) + \alpha_0 \cdot \Delta I(t) + \alpha_1$, тобто здійснюється оптимізація рівняння із тими самими регресантом та фіксованим параметром для $U(t-1)$. У таблиці 1 наведено специфікації найкращих моделей ARIMAX з відповідним значенням критерію BIC_{\min} та найгіршим значенням критерію BIC_{\max} за різних інших специфікацій ARIMAX, відмінних від оптимальної, а також показник BIC для ІМЦ.

Наведемо оцінки ІМЦ для різних моделей сумок:

$$\Delta U(\text{Chanel}) = 0.50 \cdot \Delta I(\text{Chanel}) + 2.71,$$

$$\Delta U(\text{Fendi}) = -0.85 \cdot \Delta I(\text{Fendi}) - 2.48,$$

$$\Delta U(\text{Gucci}) = 0.84 \cdot \Delta I(\text{Gucci}) - 0.16,$$

$$\Delta U(\text{Louis Vuitton}) = 1.05 \cdot \Delta I(\text{Louis Vuitton}) - 0.39,$$

$$\Delta U(\text{Saint Laurent}) = 0.49 \cdot \Delta I(\text{Saint Laurent}) - 1.51.$$

Таблиця 1. Порівняння прогностичної здатності запропонованої ІМЦ та ARIMAX

Назва товару	BIC_{min} ARIMAX	Оптимальний (p, d, q)	BIC_{max} ARIMAX	BIC ІМЦ
Chanel — Classic Flap	767.7	(1,0,0,2)	833.9	804.8
Fendi — Kan I	709.7	(1,0,0,1)	778.7	747.3
Gucci — Sylvie	858.7	(1,0,0,1)	898.8	866.9
Louis Vuitton — Alma Pm	575.8	(1,0,0,3)	607.0	610.4
Saint Laurent — Sac de Jour	718.0	(1,0,1,2)	834.3	766.9

ВИСНОВКИ

Як видно із результатів розрахунків, представлених у табл. 1, запропонована у цьому дослідженні ІМЦ у одному випадку (Gucci — Sylvie) дала результат, близький до оптимального, отриманого з використанням моделі ARIMAX, а у чотирьох інших випадках продемонструвала результат, який приблизно дорівнює результату, отриманому з використанням середньої за якістю моделі ARIMAX. Проте застосування моделі ІМЦ надає змогу значно зекономити час, який витрачається на підбір найкращої специфікації у випадку застосування моделі ARIMAX, та не потребує високої кваліфікації від аналітиків, яка є необхідною умовою для реалізації економетричних підходів.

До того ж, з результатів моделювання випливає, що зв'язок між динамікою кількості інформації та ціною блага може бути як прямим, так і оберненим. Це зумовлено тим, що насправді інформація окрім кількісних характеристик має ще й якісні ознаки, і в цьому випадку якісною ознакою інформації є контекст. Він може бути як позитивним, так і негативним. Цей висновок свідчить про те, що у подальших дослідженнях фактор інформації можна піддати більш глибокому аналізу під час моделювання для з'ясування того, у якому контексті інформація описує предмет дослідження.

Результати дослідження продемонстрували практичну цінність запропонованого підходу та його перспективність у використанні на сучасних електронних ринках. ІМЦ можна застосовувати для розв'язування широкого класу задач. Наведені приклади (22), (23), що вносять нелінійність у праву частину цієї моделі, є важливими з практичного погляду, проте поки що вони залишилися поза межами цього дослідження і потребують подальшого теоретичного і практичного розв'язання.

СПИСОК ЛІТЕРАТУРИ

1. Preis T., Moat H.S., Stanley H.E. Quantifying trading behavior in financial markets using Google trends. *Scientific Reports*. 2013. Vol. 3. P. 16–24. <http://dx.doi.org/10.1038/srep01684>.
2. Rao T., Srivastava S. Modeling movements in oil, gold, Forex and market indices using search volume index and Twitter sentiments. *Proc. 5th Annual ACM Web Science Conference (WebSci'13)*. (May, 2013, Paris, France). Paris, 2013. P. 336–345. <https://doi.acm.org/10.1145/2464464.2464521>.
3. Убиенных Г.Ф., Убиенных А.Г., Жеребцов Н.А. Моделирование ценообразования опционов на основе аппроксимации решения уравнения Блэка–Шоулза. *Актуальные направления научных исследований XXI века: теория и практика*. 2015. Т. 3. № 8–4 (19–4). С. 299–303.
4. Дунаев Б.Б., Любич А.А. Модель функционирования экономики при рыночном курсе валюты. *Кибернетика и системный анализ*. 2020. Т. 56, № 1. С. 147–162.

5. Гурвич Е.Т., Прилепский И.В. Анализ экспертных и официальных прогнозов цен на нефть. *Вопросы экономики*. 2018. № 4. С. 26-48. <https://doi.org/10.32609/0042-8736-2018-4-26-48>.
6. Huntington H., Al-Fattah S.M., Huang Z., Gucwa M., Nouri A. Oil markets and price movements: a survey of models. 2013. *USAAE Working Paper*. N 13–129. <http://dx.doi.org/10.2139/ssrn.2277330>.
7. Muradov A.J., Oglu H.Y.H., Hajiyev N. World market price of oil impacting factors and forecasting. Cham: Springer International Publishing, 2019. 184 p. <https://doi.org/10.1007/978-3-030-11494-7>.
8. Бандура О.В. Прогнозування тренду світових цін на нафту на базі СМІ-моделі економічних циклів. *Економіка і прогнозування*. 2018. № 2. С. 91-110.
9. Махорт А.Ф. О влиянии зависимости структуры потребления товаров от цены на равновесие в экономической системе. *Кибернетика и системный анализ*. 2015. Т. 51, № 2. С. 52–61.
10. Григоркив В.С. Некоторые подходы к моделированию цен в эколого-экономической системе. *Кибернетика и системный анализ*. 2009. Т. 45, № 1. С. 3–9.
11. Горбачук В.М., Морозов О.О., Неботов П.Г. До моделювання ринку Європи на природний газ. *Інфраструктура ринку*. 2017. № 3. С. 147–152.
12. Liu W.W., Liu Y., Chan N.H. Modeling eBay price using stochastic differential equations. *Journal of Forecasting*, 2018. Vol. 38, Iss. 1. P. 63–72. <https://doi.org/10.1002/for.2551>.
13. Tyralis H., Papacharalampous G. Variable selection in time series forecasting using random forests. *Algorithms*. 2017. Vol. 10(4), N 114. <https://doi.org/10.3390/a10040114>.
14. Carta S., Medda A., Pili A., Recupero R., Saia R. Forecasting e-commerce products prices by combining an Autoregressive Integrated Moving Average (ARIMA) model and Google trends data. *Future Internet*. 2018. Vol. 11, Iss. 1. P. 5. <https://doi.org/10.3390/fi11010005>.
15. Tseng K.-K., Lin R. F.-Y., Zhou H., Kurniajaya K.J., Li Q. Price prediction of e-commerce products through Internet sentiment analysis. *Electronic Commerce Research*. 2018. Vol. 18, Iss. 1. P. 65-88. <https://doi.org/10.1007/s10660-017-9272-9>.
16. Yu L., Zhao Y., Tang L., Yang Z. Online big data-driven oil consumption forecasting with Google trends. *International Journal of Forecasting*. 2019. Vol. 35, Iss. 1. P. 213-223. <https://doi.org/10.1016/j.ijforecast.2017.11.005>.
17. Szczech M., Turetken O. The competitive landscape of mobile communications industry in Canada: predictive analytic modeling with Google trends and Twitter. In: *Analytics and Data Science*. Cham: Springer, 2017. P. 143-162. https://doi.org/10.1007/978-3-319-58097-5_11.
18. Гаевский Х., Грегер К., Захариас К. Нелинейные операторные уравнения и операторные дифференциальные уравнения. Москва: Мир, 1978. 336 с.
19. Киндерлерер Д., Стампакья Г. Введение в вариационные неравенства и их приложения. Москва: Мир, 1983. 256 с.
20. Саженок В.С. Алгоритм чисельного розв'язування одного класу варіаційних параболічних нерівностей. *Математические машины и системы*. 2007. № 2. С. 19–25.
21. Di Lorenzo R. The Chaikin volatility. In: *How to Make Money by Fast Trading. Perspectives in Business Culture*. Milano: Springer, 2012. P. 175–176. https://doi.org/10.1007/978-88-470-2534-9_36.
22. Tuladhar J.G., Gupta A., Shrestha S., Bania U. M., Bhargavi K. Predictive analysis of e-commerce products. In: *Intelligent Computing and Information and Communication*; Singapore: Springer, 2018. P. 279-289. https://doi.org/10.1007/978-981-10-7245-1_29.
23. Tradesy — Buy & Sell Designer BagsShoes & Clothes. URL: <https://www.tradesy.com/>.
24. ParseHub features, tools and settings. Documentation. URL: <https://help.parsehub.com/hc/en-s/categories/203678627>.
25. Бокс Дж., Дженкинс Г. Анализ временных рядов: прогноз и управление. Москва: Мир, 1974. 406 с.

Надійшла до редакції 25.10.2019

В.С. Саженьюк, Г.А. Черноус, Ю.А. Ярмоленко

ИНФОРМАЦИОННАЯ МОДЕЛЬ ЦЕНООБРАЗОВАНИЯ НА ЭЛЕКТРОННЫХ РЫНКАХ

Аннотация. Разработан подход к моделированию и прогнозированию цены рыночного блага на основе изменений информационного фактора с использованием аппарата дифференциальных уравнений первого порядка. Построена информационная модель ценообразования путем решения задачи Коши. Показано, что предложенная модель может быть полезной как для прогнозирования цен активов на фондовой и валютной бирже, так и на товарных электронных рынках. Представлен процесс подготовки данных для реализации модели.

Ключевые слова: информация, задача Коши, метод штрафа, разностная схема, информационный фактор, модель ценообразования, прогнозирование цен.

V.S. Sazheniuk, G.O. Chornous, Iu.A. Iarmolenko

INFORMATION MODEL FOR PRICING ON ELECTRONIC MARKETS

Abstract. The research is focused on the development of an approach to modeling and forecasting a market good prices based on information changes. The process is described by first order differential equations, and the solution of a corresponding Cauchy problem made it possible to develop an information pricing model. The proposed model is shown to be useful both for predicting asset prices on the stock and currency exchanges, and on commodity electronic markets. Data preparation stage before modeling is presented too.

Keywords: information, Cauchy problem, penalization method, difference scheme, information factor, pricing model, price forecasting.

Саженьюк Володимир Степанович,

кандидат фіз.-мат. наук, доцент кафедри Київського національного університету імені Тараса Шевченка, e-mail: vssag@ukr.net.

Чорноус Галина Олександрівна,

доктор екон. наук, доцент кафедри Київського національного університету імені Тараса Шевченка, e-mail: gach2012@gmail.com.

Ярмоленко Юлія Артурівна,

студентка Київського національного університету імені Тараса Шевченка, e-mail: yu.yarmolenko@gmail.com.