УДК 311.76(477)(048)(533)

АНАЛІЗ ПОКАЗНИКІВ РИНКА ЦІННИХ ПАПЕРІВ НА ОСНОВІ СПЛАЙН-МОДЕЛЕЙ

КОНОНЕНКО В. В.
кандидат технічних наук

ПИРХ Д. О.
КРИВИЙ РІГ

В аналізі цінних паперів, як і в багатьох інших економічних задачах, не рідко потрібно визначити форму закону розподілу ймовірностей випадкової змінної. На практиці для цих цілей використовується побудова на вибіркових даних достатнього об'єму гістограма, що показує емпіричний закон розподілу.

Так, якщо процес, що представляється змінної, стаціонарне, той такий розв'язок задачі слід визнати цілком задовільним. Однак набагато частіше стохастичний процес, що породжує змінну, є нестаціонарним – змінюються його середній рівень, розмах відхилень від середнього, а також форма кривої розподілу амплітуд коливань. Якщо ці зміни носять еволюційний характер, то, як показав Р. Г. Браун [1], може бути побудована гістограма, що ґрунтується на експонентнім згладжуванні нуля або одиниці. Оновлені значення частин на момент t для різних діапазонів значень досліджуваної змінної й визначають гістограму, що приблизно відбиває закон розподілу.

У загальному випадку розглядається повна система n неспільних подій, певних на числовій осі за допомогою n_1 границі

$$X_0 < X_1 < X_2 < \cdots < X_n \qquad (1)$$

Події, пов'язаному зі спостереженням x_t у момент t відповідає номер інтервалу k – такий, що $X_{k-1} < x_t < X_k$, у який попадає спостережене значення. Припускаємо, що моменти спостереження розділені рівними проміжками часу й $t = 1, 2, \ldots, T$. Оцінку на момент t імовірності настання події k будемо позначати як

$$p(t) = P(X_{k-1} < x_t < X_k). \qquad (2)$$

Зробимо розбивку осі x на інтервали, розділивши весь діапазон $(X_{max} \div X_{min})$ між верхнім і нижнім значенням x на n рівних частин. У загальному випадку границі можуть бути задані довільно залежно від характеру задачі й цілей дослідження (рис. 1).

Початкові оцінки $\hat{p}_k(0), p = 1,2,\dots,n$ можна одержати, наприклад, використовуючи минулі дані або початкову частину вибірки об'ємом T_t T відповідно:

$$k(0) = \frac{\omega_k}{T_1} \qquad (3)$$

де k – число точок із вибірки $\{x_t\}, t = 1,2,\dots,T_1$, попавших в интервал $X_{k-1} < x_t < X_k$.

На кожному з інтервалів (x_{n-1}, x_n) задана висота стовпця гістограми u_n (рис. 2).

Побудуємо природний сплайн $S(x)$ порядку p з вузлами в x_n:

$$\int_{x_{n-1}}^{x_n} S(x)dx = u_n, \qquad (4)$$

точно передаючи сенс гістограми як інтегрального середнього по інтервалу.

Умова (4) дає N рівнянь. Сплайн порядку p на сітці має $N+p$ ступенів свободи. Невистачаючих параметрів p визначимо з умов природності:

$$\sum_{n=1}^{N-1} r_n^2 - min, \qquad (5)$$

де $r_n = S^{(p)}(x_n + 0) - S^{(p)}(x_n - 0)$ – розрив старшої (неіснуючою) похідної сплайна у вузлі x_n.

Запишемо сплайн у вигляді розкладання по базисних [2, 3]:

$$S(x) = \sum_{n=-p}^{N-1} c_n B_n(x), \qquad (6)$$

де базисний сплайн є:

$$B_n(x) = \sum_{k=0}^{p+1} b_{nk}(x - x_{n+k})_+^p \equiv (-1)^p \sum_{k=0}^{p+1} b_{nk}(x - x_{n+k})_+^p \qquad (7)$$

Таким чином, сплайн згладжування вектора ймовірностей дає оновлений вектор ймовірностей.

Рис.1. Вхідні данні котирувань акцій на прикладі Arcelor Mital Stell

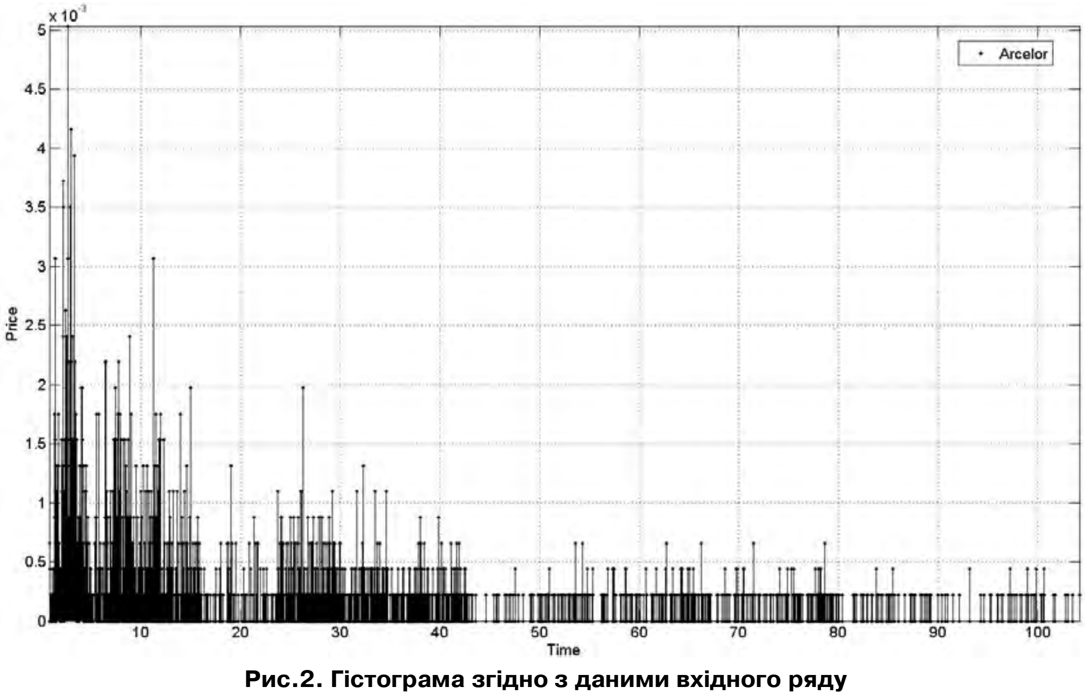

Рис.2. Гістограма згідно з даними вхідного ряду

Розглянемо подію t. Якщо закон розподілу спостережених значень x_t, не змінюється, то математичне очікування значення компоненти i вектора $u(t)$, який необхідно згладжувати, точно дорівнює дійсної ймовірності p_t настання події i і математичне очікування оцінки дорівнює дійсної ймовірності:

$$M[\hat{p}_i(t)] = \sum_{j=t}^{\infty} a(1-a)^{j-t} M[u_i(t)] = p_i \sum_{j=t}^{\infty} a(1-a)^{j-t} = p_i. \quad (8)$$

Ймовірність того, що прийде згладжувати одиницю, рівна p_i, а того, що згладжувати нуль, становить $1-p_i$. Легко підрахувати, що компонента i вектора $u(t)$ характеризується дисперсією

$$D[u_i(t)] = (1-p_i)^2 p_i + (0-p_i)^2(1-p_i) = p_i(1-p_i). \quad (9)$$

Відповідно до експоненційного згладжування до компонента u_i застосовуємо процедуру, та одержуємо дисперсію на виході яка виражається через дисперсію вхідного потоку за формулою:

$$\sigma_i^2 = \frac{a}{2-a} D[u_i(t)], \quad (10)$$

де a – постійна згладжування.

Після підстановки (9) одержимо дисперсію оцінки ймовірності настання події t.

З (11) Р. Г. Браун робить висновок, що є два найбільш кращі способи конструювання системи границь $X_0, X_1, X_2, \ldots, X_n$ при якій дисперсія оцінки порівняно невелика. Він пропонує встановлювати такі границі подій, щоб p_i була або дуже великою (близької до одиниці), або дуже маленької (майже рівної нулю). Це забезпечить мале значення добутку $p_i(1-p_i)$ і, отже, низьку дисперсію оцінок компонент вектора ймовірностей. Максимум дисперсії досягається при $p_i = 0{,}5$.

Однак подібні міркування занадто формальні, а отримані рекомендації не завжди реалізовані. Якщо, наприклад, покласти, що $p_i = 0{,}9$, те на всі інші події залишається лише 10% імовірності. Така розбивка на події, звичайно, можливо, якщо мова йде, наприклад, про поділ стандартних і нестандартних ситуацій. Але при аналізі розподілу ймовірностей буде потрібно побудувати границі, що розділяють числову вісь на 5–9 або більш подій. Тоді умова $p_i = 1$ стає неприйнятним по очевидних причинах. У цьому випадку доцільно зажадати, щоб усі оцінки ймовірностей мали рівні дисперсії $\sigma_i^2 = const$ або однакове відношення стандартної помилки до оцінки ймовірності $\frac{\sigma_i}{p_i} = const$. У першому випадку висунуті вимоги показують, що

$$\sigma_i^2 = \frac{a}{2-a} p_i(1-p_i) = const, i = 1,2,\ldots,n \quad (12)$$

або

$$p_i(1-p_i) = const, i = 1,2,\ldots,n. \quad (13)$$

Незважаючи на те, що формально можна одержати два розв'язки, годиться лише одне – те, яке менше 0,5. У другому випадку

$$\frac{\sigma_i^2 = \frac{a}{2-a} p_i(1-p_i)}{p_i^2} = const, i = 1,2,\ldots,n \quad (14)$$

або

$$\frac{p_i(1-p_i)}{p_i} = const, i = 1,2,\ldots,n, \quad (15)$$

звідки

$$p_i = \frac{1}{1+const} = const, i = 1,2,\ldots,n \quad (16)$$

Таким чином, в обох випадках приходимо до висновку, що для рівності дисперсій оцінок σ_i^2 або коефіцієнтів варіації $\frac{\sigma_i}{p_i}$ границі подій треба вибирати так, щоб імовірності настання цих події були однаковими.

Досить важливе питання, залишене Р.Г. Брауном без відповіді, стосується вибору оптимального значення параметра a. Формула (12) лише свідчить про те, що чим ближче a до нуля, тем менше дисперсія оцінки ймовірності. Але нас, в остаточному підсумку цікавить не стільки її коливання, скільки ступінь відповідності сформованого та скорегованого розподілу (точніше, гістограми) характеру реального розкиду значень досліджуваної величини, по припущенню, що змінюється в часі. Зовсім очевидно, що питання про оптимальне значення параметра a може бути вирішений тільки при наявності критерію якості розглянутої процедури.

За основу такого критерію в [4] пропонується обрати максимальний рівень значимості, при якім отриману в момент t фактичну точку можна віднести до побудованого на цей же момент розподілу. Сума (або нормована сума) таких максимальних рівнів значимості по моментах часу $t = 1,2,\ldots,T$ дасть значення підсумкового критерію Q, який буде функцією параметра адаптації a і початкових оцінок ймовірностей подій на момент $t = 0$. Максимальна величина Q визначає оптимальне значення a і початкових оцінок ймовірностей.

Сумування по всіх точках вибірки дозволяє сформувати критерій якості для процедури в цілому:

$$Q_2(x, a, T_1) = \frac{1}{T} \sum_{t=1}^{T} q_{\max\square}, t \quad (17)$$

Критерій Q_2 інтерпретується як усереднене по вибірці максимальне значення рівня значимості, при якій спостережені значення ряду можна вважати Приналежними розподілам, що представляються відповідними гістограмам.

У формулі (17) показане, що в обох випадках критерій якості процедури адаптації є функцією не тільки даних $\{x_t\}$ і параметра адаптації a але й об'єму вибіркових даних T_1 використовуваних для визначення початкових значень ймовірностей подій, необхідних для запуску рекурентної процедури.

Відмітимо також, що Q_2 сформовані як однобічні критерії, що припускають, що зіставлення реалізації процесу x_t с побудованим розподілом проводиться на одному його «хвості».

Очевидно, що Q_2 ухвалюють максимальне значення одночасно, тобто при тих самих a і T_1.

Слід відзначити, що при використанні побудованої гістограми можуть застосовуватися як однобічний, так і двосторонній критерії залежно від задач, і результати, звичайно, будуть різні.

Маючи критерій якості процедури, неважко визначити й оптимальне значення параметра a Для цього треба на інтервалі $0 < a < 1$, наприклад, методом проб знайти таке його значення, яке максимізує Q_2. Точно так само можна знайти й оптимальне значення T_1, – початкову частину вибірки, використовувану для обчислення грубих оцінок ймовірностей $\hat{p}_k(0)$, необхідних для першого застосування рекурентної формул експонентного згладжування.

Відповідно, оптимізація критерію Q по двом параметрам a і T_1, трохи ускладнює обчислювальний процес.

Необхідність аналізу нестаціонарних розподілів часто виникає в економічних дослідженнях.

Для розв'язку цієї задачі нами була створена програма, в основу якої покладений описаний вище алгоритм. Передбачена оптимізація процедури як по параметру a так і по початковому об'єму вибірки T_1 Результати рахунку представляються на рис. 3.

Імовірнісна модель може призначати для вивчення, наприклад, розподілу доходів у різних прошарках суспільства, попиту на окремі види товарів по вікових групах, еволюції розподілу сімейних бюджетів по статтях витрат, розподілу сум поточних внесків і інших фінансових, економічних і соціально-економічних процесів.

Особливе значення запропонований підхід має для аналізу ризиків, і, зокрема, в аналізі, що одержав назву Value at Risk (VAR). Такий аналіз дозволяє оцінити величину максимального знецінення активу при заданому рівні значимості (імовірності). ■

ЛІТЕРАТУРА

1. **Brown R. G.** Smoothing forecasting and prediction of discrete time series. – N.Y., 1963.
2. **Андерсон Г.** Статистический анализ временных рядов. – М.: Мир, 1976.
3. **Бартлетт М. С.** Введение в теорию случайных процессов. – М.: ИЛ, 1958.
4. **Лукашин Ю. П.** Анализ распределения кассовых остатков: адаптивная гистограмма, проблема оптимизации // Экономика и математические методы.–1997.– Т. 33.– Вып. 3. – С. 90–97.

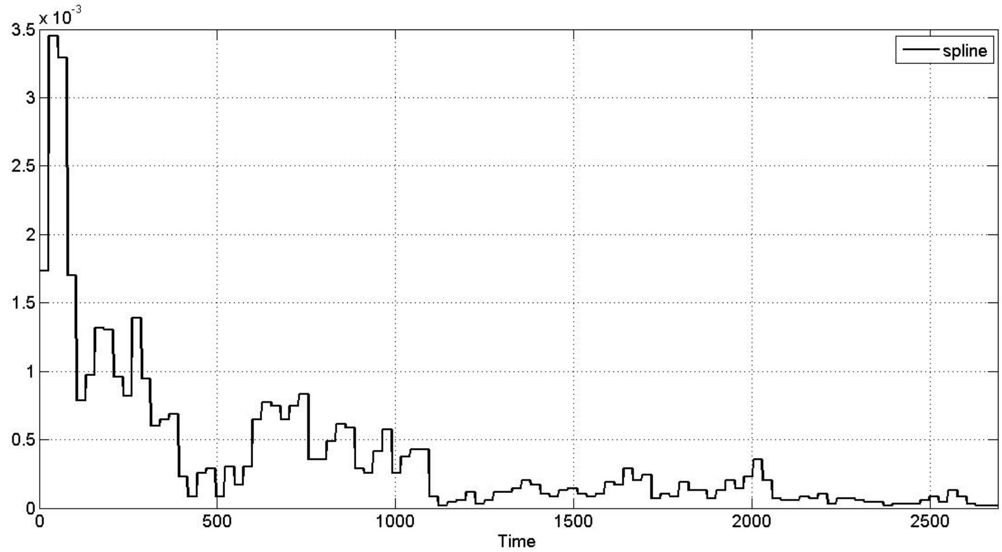

Рис.3. Згладжений сплайном вхідний ряд