УДК 519.688:336.76.066

О.П. Луценко, О.Г. Байбуз

Дніпропетровський національний університет імені Олеся Гончара

ВІДНОВЛЕННЯ ФОРМИ РОЗПОДІЛУ ТОЧОК РОЗЛАДНАНЬ НА ЧАСОВОМУ РЯДІ ВАЛЮТНИХ КОТИРУВАНЬ

Розглянуто форму розподілу, утвореного точками зміни статистичних характеристик ряду курсів валют. Проаналізовано можливості його відновлення з використанням параметричного та непараметричного підходів.

Ключові слова: *статистика, визначення розладнань, біржова торгівля, котирування, часовий ряд, нестаціонарні процеси, відновлення розподілів.*

Рассмотрена форма распределения, образованного точками изменения вероятностных характеристик временного ряда валютных котировок. Проанализированы возможности его восстановления с использованием параметрического и непараметрического подходов.

Ключевые слова: *статистика, определение разладок, биржевая торговля, котировки, временной ряд, нестационарные процессы, восстановление распределений.*

The distribution formed by change-points of financial quotes time series has been reviewed. Possibilities of its estimation using parametric and non-parametric approach have been analyzed.

Keywords: *statistics, change-point detection, stock market, market quotations, time series, non-stationary processб probability distribution estimation.*

Постановка проблеми. Висока прибутковість торгівлі на Forex поєднується з ризиками, для мінімізації яких застосовується технічний аналіз ринку. Технічний аналіз керується твердженням «курсураховує все» і ґрунтується тільки на поточній поведінці курсу валют, без урахування причин зміни.

Основою технічного аналізу є твердження, що ціни на ринку рухаються направлено, підкоряючись певній тенденції (тренду). Виділяють висхідну (бичачу), спадну (ведмежу) й бічну (флет) тенденції. Завданням технічного аналізу є визначення характеру діючої тенденції та виявлення моментів її зміни шляхом аналізу історії зміни цін на ринку.

Серед методів, найбільш часто вживаних у технічному аналізі, можна виділити такі: допоміжні побудови на графіках курсів валют (лінії опору і підтримки, лінії тренда тощо), індикатори та осцилятори. Існує велика кількість різновидів і модифікацій цих методів, але для основної маси характерні загальні недоліки: суб'єктивність допоміжних графічних побудов, значна затримка сигналу про розворот тренда у часі, велика кількість хибних сигналів в умовах невираженого тренда.

У зв'язку з перерахованими недоліками, а також зі зростанням рівня обчислювальних можливостей персональних комп'ютерів відносно недавно сформувалася тенденція до переходу від індикаторів та осциляторів технічного аналізу до більш точних і складних методів. Додатковий вплив на розвиток ринку програмного забезпечення Forex має збільшення доступності торгівлі на цьому ринку (збільшення кількості компаній-посередників, поширення торгівлі через інтернет, реклама у ЗМІ), яке приводить до подальшого збільшення кількості учасників торгівлі й зростання попиту на технології оцінки та мінімізації ризиків. У зв’язку з цим розробка нових методів та уточнюючих моделей, що можуть бути використані у технічному аналізі, є актуальною задачею.

Відмінними рисами методу оцінки ризиків, пропонованого авторами даної статті, є:

1) імовірнісний підхід. Оскільки зміни валютних котирувань є значною мірою випадковим процесом, його доцільно розглядати з позицій імовірності різних гіпотез прогнозу. Детерміністичне прогнозування не дає уявлення про співвідношення таких імовірностей, показуючи лише найбільш імовірний варіант розвитку подій, який часто буває обраний із мінімальною перевагою щодо інших варіантів. Перехід до ймовірнісної моделі прогнозування є перспективним з огляду на прозорість даних про достовірність прогнозу;

2) розгляд часового ряду валютних котирувань не як безлічі окремих точок, а як сукупності відрізків стаціонарності, на які він розбивається точками розладнання. Знаходження точок розладнання ряду дозволяє розділити ряд на ділянки з подібними статистичними властивостями, що відповідає одній з основних парадигм технічного аналізу – поділу часового ряду на відрізки з постійною тенденцією.

Постановка цілей статті. Основу моделі становить припущення, що точки розладнання статистичних характеристик ряду котирувань (зміни напряму тренда) є випадковою величиною, функція розподілу якої повільно змінюється з часом. Відновивши щільність розподілу, отримаємо можливість на кожному кроці спостережень судити про ймовірність зміни напряму ринкової тенденції.

Мета даної статті – опис форми розподілу, утвореного точками розладнань ряду валютних котирувань, а також аналіз ефективності його відновлення за допомогою параметричного (ЕМ-алгоритм) та непараметричного (сплайн-аппроксимація) підходів.

Основний матеріал. Перед тим як перейти до розгляду розподілу, опишемо методи, за допомогою яких було отримано дані.

Для вирішення завдання розбиття ряду на ділянки стаціонарності використовуються два методи: графічний метод пошуку перегинів графіка, призначення якого – знаходити точки розвороту апостериорно, і алгоритм кумулятивних сум (АКС), що належать до методів послідовного виявлення розладнань.

Графічний метод має високу точність і призначений для побудови основної вибірки ділянок стаціонарності, на основі якої відновлюється початкове значення щільності розподілу розладнань. Цей метод працює за таким принципом.

Зафіксуємо початковий і кінцевий моменти в процесі, всередині якого будемо шукати точку розладки: m – початковий момент часу, n – кінцевий момент часу. На цьому проміжку шукатимемо лише одну точку розладки.

Для цього проміжку побудуємо масив

$$\{d_i, i = \overline{n,m}\},$$

$$d_i = \frac{|kx_i - y_i + b|}{\sqrt{k^2+1}},$$

де k, b – знайдені коефіцієнти рівняння прямої ($y=kx+b$), що проходить через точки (m, x_m) та (n, x_n).

Точку розладки знаходимо в такий момент часу r, що

$$x_r = \max_{i=\overline{m,n}} d_i \ , r \in (m,n), \ x_r > bar,$$

де bar – бар'єр чутливості методу.

Якщо точка розладки була знайдена, то момент часу r розбиває проміжок [m, n] на два: [m, r] та [r, n], які розглядаємо аналогічним способом, переходячи до пункту 1.

Процес закінчується, коли жодне значення d_i на проміжку не перевищує встановленого бар'єра.

У процесі використання моделі часовий ряд котирувань поповнюється новими даними і виникає необхідність знаходити розладнання в найновіших отриманих даних. Для цієї мети використовуватимемо алгоритм кумулятивних сум (АКС). АКС полягає в аналізі поведінки величини [1]:

$$S_t = S_{t-1} + \ln(\omega(x_t/\theta_2)\omega(x_t/\theta_1)).$$

На кожному кроці сума порівнюється із заданим порогом h. Якщо на кроці t S>h, подається сигнал про розладнання, накопичення суми починається з нуля. Відбиваючий екран встановлено в точці 0.

$$S_t = \max(0, S_{t-1} + \Delta S),$$

$$\Delta S = \ln(\omega(x_t/\theta_2)/\omega(x_t/\theta_2)).$$

Знаючи тип розподілу та визначаючи одну з імовірнісних характеристик розподілу як параметр θ, можна отримати рекурентні формули накопичення кумулятивної суми. Так, для випадку виявлення зміни середнього значення нормального розподілу формула для прирощення матиме вигляд

$$g_t = \frac{m_2 - m_1}{\sigma^2}\left(x - \frac{m_1 + m_2}{2}\right),$$

де m_1 – математичне очікування величини до розладнання;

m_2 – передбачуване математичне очікування величини після розладнання;

σ – середнє квадратичне відхилення;

x – значення спостереження у момент часу t.

У даному випадку зручно застосовувати різновид АКС, у якому на кожному кроці із заданим порогом порівнюється різниця

$$g_t = S_t - \min_{k<t} S_k. \qquad (1)$$

Якщо необхідно виявляти зміни θ у бік зменшення, різниця приймає вигляд

$$g_t = \max_{k<t} S_k - S_t \qquad (2)$$

Сигнал про розладку подається в момент часу

$$\tau = \inf(t \geq 1 : g_t > h).$$

Розділення умов подання сигналу про розладнання «вгору» і «вниз», що забезпечується рівняннями (1) і (2), є важливим для даної задачі, оскільки зміни характеристик у бік посилення чинного тренда не призводять до розвороту тренда і не повинні враховуватися.

Оскільки стабільний тренд сам по собі характеризується зміною середнього значення спостережень, у даному випадку доцільно застосовувати АКС не до початкової кривої спостережень, а до її першої похідної за часом. Фізичний сенс такої похідної – швидкість зміни значень функції. Різке змінення значення першої похідної означатиме розладнання процесу . У разі, коли потрібно визначити зміну діючої тенденції, точка розвороту графіка буде характеризуватися тим, що значення першої похідної перетне нульову позначку. Відсіваючи розладнання, зміна спостережень у яких направлена у протилежний від нульової позначки бік (посилення тренда), а також розладнання, що лежать надто далеко від нульової позначки (уповільнення тренда, яке ще далеко від розвороту), отримуємо точки, що розділяють області дії трьох основних ринкових трендів.

Нехай після розбиття графіка маємо n відрізків стаціонарності. Кожен із відрізків характеризується вектором параметрів v, що включає в себе його тривалість у часі (довжину відрізка стаціонарності по осі x – x_S) і різницю між цінами на початку і в кінці відрізка (різниця по осі y – y_S). Таким чином, процес описується вектором випадкових величин (x_S, y_S), типова гістограма розподілу якого показана на рис. 1.

Обчислення критеріїв χ^2 і точного критерію Фішера вказують на наявність імовірнісної залежності між функціями розподілу x_S та y_S; отже, спрощення форми розподілу вектора до добутку двох одновимірних неможливе.

Отримана функція розподілу має складну форму з двома піками і не може бути апроксимована з прийнятим ступенем точності жодним зі стандартних двовимірних розподілів.

Наблизимо отриману функцію розподілу сумішшю двомірних нормальних розподілів за допомогою ЕМ-алгоритму.

ЕМ-алгоритм припускає, що дані можуть бути кластеризовані і підпорядковуються лінійній комбінації (суміші) нормальних (гауссових) розподілів:

$$p(x) = \sum_{j=1}^{k} w_j p_j$$

$$\sum_{j=1}^{k} w_j = 1,$$

$$w_j \geq 0,$$

де $p_j(x)$ – функція правдоподібності j-ї компоненти суміші, w_j – її апріорна ймовірність.

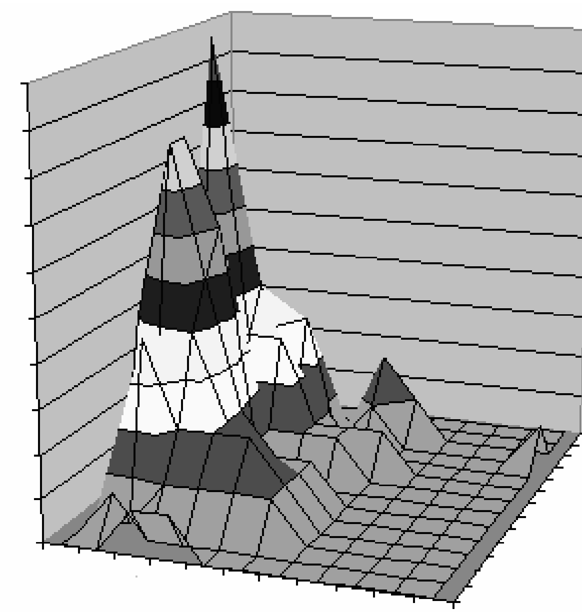

Рис. 1. Гістограма розподілу вектора (x_S, y_S)

Нехай функція правдоподібності належить до параметричного сімейства $\varphi(x;\theta)$ і відрізняється лише значеннями параметра θ.

Задача розділення суміші розподілів полягає в тому, щоб, маючи вибірку X^m випадкових спостережень із суміші, знаючи число k та функцію $\varphi(x;\theta)$, оцінити вектор параметрів $\Theta = (\omega_1,...\omega_k,\theta_1,...\theta_k)$.

Щільність імовірності нормального розподілу має вигляд

$$\varphi(x;\theta) = p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\},$$

де μ – математичне очікування, σ^2 – дисперсія.

Для двовимірного випадку

$$p(x_1,x_2) = \frac{1}{2\pi\sigma_1\sigma_2\sqrt{1-\rho^2}} \times$$

$$\times \exp\left\{-\frac{1}{2(1-\rho^2)}\left[\frac{(x_1-\mu_1)^2}{\sigma_1^2} - \rho\frac{(x_1-\mu_1)(x_2-\mu_2)}{\sigma_1\sigma_2} + \frac{(x_2-\mu_2)^2}{\sigma_2^2}\right]\right\}.$$

ЕМ-алгоритм складається з ітераційного повторення двох кроків [2]:

1. На Е-кроці обчислюється значення вектора прихованих змінних G.

$$g_{ij} \equiv P(\theta_j \mid x_i),$$

$$\sum_{j=1}^{k} g_{ij} = 1,$$

$$g_{ij} = \frac{\omega_j p_j(x_1, x_2)}{\sum_{s=1}^{k} \omega_s p_j(x_1, x_2)},$$

де $p_j(x_1, x_2)$ – значення функції щільності двовимірного нормального розподілу;

g_{ij} – приховані змінні, фізичний зміст яких – апостеріорна ймовірність того, що об'єкт x_i належить до j-ї компоненти суміші.

На М-кроці вирішується задача максимізації логарифма правдоподібності і знаходиться наступне значення вектора Θза поточними значеннями вектора прихованих змінних.

$$Q(\Theta) = \ln \prod_{i=1}^{m} p(x_1, x_2) \to \max_{\Theta}.$$

Маємо

$$w_j = \frac{1}{m} \sum_{i=1}^{m} g_{ij}, \quad j = 1,\dots,k,$$

$$\theta_j = \arg\max_{\theta} \sum_{i=1}^{m} g_{ij} \ln p(x_1, x_2), \quad j = 1,..k.$$

Метод не є стійким до початкового наближення, і кінцевий результат залежить від даних, якими ініціалізується алгоритм. Під час рішення задачі ініціалізація проводилася двома типами даних: результатом кластерізації вибірки за методом k середніх, а також заданням наближення в ручному режимі. На жаль, в обох випадках алгоритм виявився нечутливим до спаду щільності розподілу біля нульових значень по y, яке ясно видно на гістограмі незалежно від вибору відрізку вхідних даних і розміру вибірки. Результат роботи наведений на рис. 2.

Як бачимо, спроби наблизити форму функцій сумою гауссіан за допомогою ЕМ-алгоритму дають досить низьку точність апроксимації. Набагато кращий результат був отриманий при використанні непараметричної апроксимації функції розподілу за допомого В-сплайна. Вибір В-сплайн-апроксимації серед інших методів пояснюється тим, що:

1) гладкість одержуваної на виході поверхні відповідає природним формам функцій розподілу;

2) В-сплайни швидкі у побудові, а їхня властивість локальності дозволяє з меншими витратами оновлювати розподіл при отриманні нових даних про розладнання.

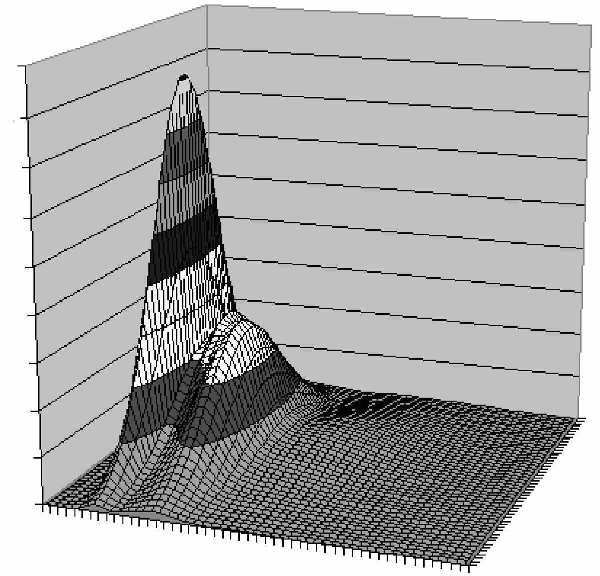

Рис. 2. Результат аппроксимації розподілу сумою гауссіан за допомогою ЕМ-алгоритму

Розділимо графік розподілу функції (x_S, y_S) на прямокутні ділянки розміром dx на dy . Підрахуємо кількість попадань у кожну ділянку і на підставі отриманих значень побудуємо гістограму. Точки гістограми будуть служити вузлами сплайна. Отримуємо двовимірний масив вузлових точок величиною $m \times n$. Двовимірну В-сплайн-функцію запишемо у вигляді тензорного добутку одновимірних В-сплайнів, побудованих на кожній із двох осей координат. У такому випадку значення В- сплайн-поверхні в довільній точці буде рівним [3]:

$$x(s,t) = \sum_{i=0}^{n} \sum_{j=0}^{m} P_{i,j} B_i(s) B_j(t),$$

$$B_{k,1} = \begin{cases} 1, & t_k < t < t_{k+1} \\ 0, & інакше \end{cases}$$

$$B_{k,d} = \left(\frac{t - t_k}{t_{k+d-1} - t_k} \right) B_{k,d-1}(t) + \left(\frac{t_{k+d} - t}{t_{k+d} - t_{k+1}} \right) B_{k+1,d-1}(t),$$

де *s, t* – координати точки,

$P_{i,j}$ – значення функції в точці *i,j*;

$B_{k,d}(t)$ – значення стикувальної функції порядку d в точці t, що належить до інтервалу (t_k; t_{k+1});

t_k; t_{k+1} – пари сусідніх вузлових точок сплайна.

При рішенні задачі використовувалися кубічні стикувальні функції і відкриті однорідні вузлові вектори

$$U=(-3,-2,-1,0,1,..,n+1),$$
$$V=(-3,-2,-1,0,1,..,m+1).$$

Результат відновлення щільності розподілу (x_S, y_S) наведений на рис. 3.

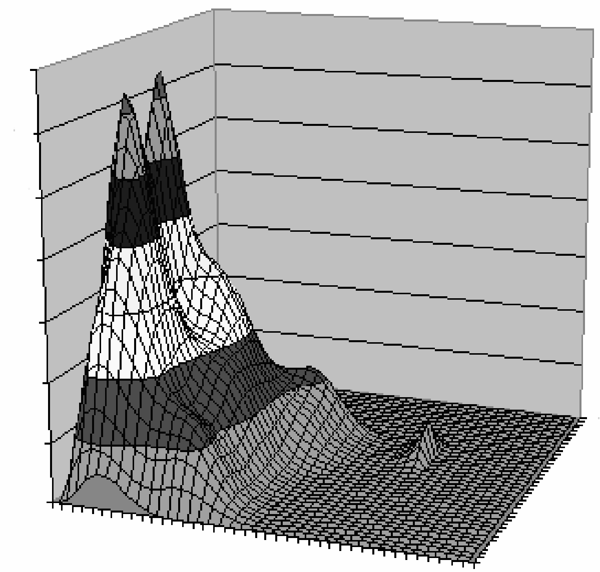

Рис. 3. Результат апроксимації щільності розподілу кубічним В-сплайном

Можемо бачити, що отримана поверхня є значно ближчою за формою до гістограми експериментального розподілу, за якою відновлювалася щільність. Крім того, на відміну від ЕМ-алгоритму, ступінню гладкості поверхні легко керувати, змінюючи розмір комірок гістограми та вводячи кратні вершини при побудові сплайна.

Висновки. Ураховуючи складність форми розподілу бачимо, що використання непараметричних методів відновлення щільності є більш доцільним, тому що на виході дає результат, значно ближчий до форми вихідної гістограми. Відновлена щільність може бути використана для отримання функції ризиків, яка виражає ймовірність зміни напряму діючої тенденції у певний відрізок часу в майбутньому. Така функція може використовуватись як самостійно, так і як уточнююча величина при побудові торговельних стратегій.

Перспективи подальших досліджень: побудова функцій ризику, встановлення оптимальних налаштувань алгоритмів залежно від масштабу та форми вхідних даних.

Бібліографічні посилання

1. **Brodsky B.E.** Nonparametric Methods in Change-Point Problems / B.E.Brodsky, B.S. Darkhovsky. – Dordrecht: Kluver Academic Publishings, 1993. – 210 p.
2. **Королёв В.Ю.** ЕМ-алгоритм, его модификации и их применение к задаче разделения смесей вероятностных распределений: теоретический обзор / В.Ю. Королев. – М.: ИПИРАН, 2007. – 94 с.
3. **Rogers D.F.** An Introduction to NURBS with Historical Perspective / D.F. Rogers. – San Francisco: Morgan Kaufmann Publishers, 2000. – 344 p.

Надійшла до редколегії 12.09.2013 р.