УДК 330.46; 519.65

ПОПЕРЕДНЯ ОБРОБКА ВХІДНИХ ДАНИХ ДЛЯ СИСТЕМИ МОНІТОРИНГУ ФОНДОВИХ РИНКІВ

Олександр Варфоломійович ПІСКУН
к.т.н., доцент кафедри вищої математики та інформаційних технологій Черкаського інституту банківської справи УБС НБУ (м. Київ)
E-mail: piskun@ukr.net

Анотація. У статті досліджено непараметричні методи згладжування часових рядів. Визначені оптимальні методи попередньої обробки рядів даних для системи моніторингу фондових ринків.

Аннотация. В статье рассмотрены непараметрические методы сглаживания временных рядов. Определены оптимальные методы предварительной обработки рядов данных для системы мониторинга фондовых рынков.

Ключові слова: *фондовий ринок, моніторинг, непараметричні методи згладжування.*
Ключевые слова: *фондовый рынок, мониторинг, непараметрические методы сглаживания.*

Постановка проблеми. Сучасні фінансові ринки характеризується значною складністю процесів, що протікають на них. Відбувається глобалізація міжнародних ринків, збільшується волатильність валют, процентних ставок, курсів цінних паперів, цін на сировину і, як підсумок, фінансові ринки стали більш нестабільними, складними, ризикованими. Для ефективного слідкування за станом ринку потрібен інструмент його моніторингу. Одним із новітніх методів дослідження часових рядів є рекурентний кількісний аналіз. Проведені дослідження показали здатність міри ламінарність (LAM) рекурентного кількісного аналізу виявляти різні періоди функціонування фінансових ринків та аналізувати протікання кризових явищ на них [1, 2].

Фінансовим ринкам властива висока турбулентність, а тому й у динаміці відповідних рекурентних мір буде присутня стохастична складова. При побудові системи моніторингу необхідно провести згладжування міри LAM для автоматизації визначення точок переходів між періодами функціонування ринку.

Аналіз останніх досліджень та публікацій. Методи обробки та аналізу часових рядів досліджено в роботах відомих вітчизняних та зарубіжних науковців і спеціалістів, зокрема: Дж. Юла, М. Кендалла, А. Сьюарта, В. Хартле, Дж. Полларда, С. Айвазяна, Г. Кільдішева та інших. Проте вибір стратегії та методів попередньої обробки та аналізу рядів динаміки залежить від кінцевої мети дослідника.

Метою статті є огляд, аналіз та визначення оптимального методу попередньої обробки (згладжування) рядів даних для системи моніторингу фондових ринків.

Обґрунтування отриманих наукових результатів. Для довгих рядів, як правило, неможливо вказати відповідну параметричну криву для згладжування ряду на всій його довжині. У цьому випадку використовують різні непараметричні методи аналізу часових рядів, такі як згладжування ковзними середніми, частотну фільтрацію тощо [3].

Окреслимо найбільш поширені непараметричні методи згладжування часових рядів.

Метод ковзних середніх. При згладжуванні цим методом фактичні значення ряду динаміки замінюються середніми значеннями, які характеризують серединну точку періоду ковзання [4].

Просте згладжування ґрунтується на складанні нового ряду з простих середніх арифметичних, розрахованих для проміжків часу довжиною k:

$$\bar{y}(t) = \frac{1}{k} \sum_{t=i}^{i+k-1} y(t),$$

де k – довжина періоду згладжування, залежить від характеру часового ряду, а також від мети згладжування і вибирається дослідником,

i – порядковий номер середньої,

n – довжина ряду.

Зважене згладжування полягає у визначенні середніх, зважених для різних точок ряду динаміки. У основі методу лежить ідея локального наближення тренду поліномом невисокого ступеня. Значення оцінки тренду в точці t_j апроксимуються по рівнях ряду з часового інтервалу $[t_j - k, t_j + k]$ поліномом заданого порядку l:

$$\bar{y}(t) = \sum_{i=0}^{l} a_i t_j^i,$$

параметри якого оцінюються за методом найменших квадратів за допомогою рівнянь типу:

$$a_0 \sum_{i=-k}^{k} t_j^i + a_1 \sum_{i=-k}^{k} t_j^{i+1} + \cdots + a_l \sum_{i=-k}^{k} t_j^{i+k} = \sum_{i=-k}^{k} y_i t_j^i.$$

Розв'язуючи отримані рівняння щодо a_i, отримаємо послідовність ваг, що залежать тільки від ширини інтервалу $(2k+1)$ та порядку полінома l, а розрахунок значень оцінок тренда в точці t еквівалентний побудови зваженої суми значень ряду в інтервалі $[t_j - k, t_j + k]$.

Значення ваг для різних k і l визначені та представлені у відповідних таблицях [4, 5]. Для поліномів нульового та першого порядків ваги a_i рівні між собою, що зводить цей метод до простого згладжування.

При викладеній організації локального згладжування незгладженими виявляються перші й останні k елементів вихідного часового ряду. Для згладжування кінцевих відрізків ряду розроблено відповідні формули [4].

*Метод локальної р*егресії (узагальнене представлення методів зваженого ковзного середнього). Згладжене значення $\bar{y}_j$, що відповідає середині відрізку локального згладжування $[t_j - k, t_j + k]$ обчислюють за допомогою вектора

$$\hat{A} = (K^T K)^{-1} K^T Y_{t_j - k}^{t_j + k},$$

де K – матриця розміром $(2k + 1)(l + 1)$, елементи якої обчислюються за формулою

$$k_{ij} = (i - k - 1)^{j-1}, i = 1,2,\dots,2k-1;\ \ j = 1,2,\dots,l+1,$$

В якості значення оцінки тренду в точці t_j береться значення a_0. Інші компоненти вектору коефіцієнтів полінома застосовуються при розрахунку крайніх точок тренду, оскільки алгоритм у представленій вище формі не дозволяє отримати оцінки тренду перших k і останніх k точок часового ряду [6]. Для обчислень крайніх значень використовуються вектора коефіцієнтів полінома, вичислені, відповідно, для точок з номерами $k + 1$ і $n - k$ за формулами:

$$\bar{y}_j = \sum_{i}^{l} t_{i+1}^{(k+1)} (j - k - 1)^i, j \le k,$$

$$\bar{y}_j = \sum_{i}^{l} t_{i+1}^{(n-k)} (j + k - n)^i, j > n - k.$$

Крім того, для згладжування кінцевих відрізків ряду можна застосувати алгоритми типу рекурентного методу найменших квадратів.

Метод локально зваженої регресії (LOWESS або LOESS) передбачає, що кожен відлік згладженого сигналу шукається як значення функції локально зваженої регресії, що оцінюється в околі даного відліку [7]. Локальні ваги обчислюються за формулою

$$w_i = \left(1 - \left|\frac{t_k - t_i}{d(t_k)}\right|^3\right)^3,$$

де t_k – точка, в якій оцінюється функція регресії,

t_i – точка, для якої обчислюється вагова функція,

$d(t)$ – половина ширини інтервалу, на якому оцінюється регресія.

Найчастіше застосовують лінійну (Lowess) або квадратичну (Loess) функцію регресії, значно рідше застосовують поліноміальні функції вищих порядків.

Параметром окреслених вище методів згладжування, що визначає якість і характеристики згладженого сигналу, є ширина вікна, тобто кількість відліків сигналу в околі деякого відліку, які використовуються для обчислення його згладженого значення. В загальному випадку, що більша ширина вікна, то більше високочастотних складових і особливостей сигналу усувається в результаті згладжування.

Метод згладжування сплайнами. Нехай відрізок $[a, b]$ розділений на $(k+1)$ частини точками $a_1 <\dots< a_k$. Сплайном або кусочно-поліноміальної функцією ступеня m з k сполученнями в точках $a_1, \dots, a_k$ називається функція $S(t)$, яка на кожному інтервалі (a_j, a_{j+1}), $j = 0, \dots, k$ описується алгебраїчним поліномом $P_j(t)$ ступеня m. Коефіцієнти поліномів $P_j(t)$ узгоджені між собою так, щоб виконувалися умови безперервності функції $S(t)$ та її $(m$-1) похідних у вузлах сполучень [8].

Найкращий спосіб наближення сплайнами – апроксимація в рівновіддалених вузлах, причому найбільш застосовні сплайни третього ступеня (кубічні сплайни) [9].

Для кубічного сплайну функція $S(t)$ на кожному відрізку $[t_{i-1}, t_i]$ є поліном третього ступеню $S_i(t)$, коефіцієнти якого треба визначити:

$$S_i = a_i + b_i(t - t_i) + c_i(t - t_i)^2 + d_i(t - t_i)^3$$

Умови безперервності всіх похідних до другого порядку включно записуються у такому вигляді:

$$S_i(t_{i-1}) = S_i(t_{i-1})$$
$$S_i'(t_{i-1}) = S_{i-1}'(t_{i-1})$$
$$S_i''(t_{i-1}) = S_{i-1}''(t_{i-1})$$

Згладжуючий кубічний сплайн визначається як сплайн, який мінімізує наступний функціонал, що залежить так само і від деякого параметра p:

$$I(s,p) = p\sum_{k=1}^{n} w_k\,(y_k - s(t_k))^2 + (1-p)\int_{t_1}^{t_n}\left(\frac{d^2 s(t)}{dt^2}\right)^2 dt,$$

де (t_k, y_k), $k=1,2,\ldots, n$ – дані, що наближуються;

w_k – ваги даних;

p – згладжуючий параметр, що змінюється від 0 до 1, який визначає кривизну сплайну.

Ваги w_k, зазвичай, вибираються як ваги узагальненої квадратурної формули трапецій. У разі рівновіддалених точок з кроком h:

$$w_1 = w_n = \frac{h}{2}, \;\; w_k = h, \;\; k = 2,3,\ldots,n-1.$$

Якщо задавати значення згладжуючого параметра p близькі до нуля, то згладжуючий сплайн буде схожий на пряму, що наближає дані за методом найменших квадратів, оскільки основним у функціоналі, що мінімізується, стане другий доданок

$$(1-p)\int_{t_1}^{t_n}\left(\frac{d^2 s(t)}{dt^2}\right)^2 dt,$$

який якраз і відповідає за гладкість, його мінімізація відповідає побудові сплайна з найменшим значенням другої похідної (нуль, для полінома першого порядку). Навпаки, якщо значення згладжуючого параметра близьке до одиниці, то основним у функціоналі, що мінімізується, стане перший доданок

$$p\sum_{k=1}^{n} w_k\,(y_k - s(t_k))^2,$$

який відповідає за проходження сплайна через задані точки. При p=1 згладжуючий сплайн перетворюється на звичайний кубічний сплайн.

Ми оглянули методи локальної апроксимації. Існує й інший підхід, який умовно можна називати нелокальними методами згладжування. У цьому випадку скоригований ряд визначається всіма значеннями вихідного ряду. Якщо значення змінюються або додаються нові точки, то складові також змінюються. Одним із таких методів є вейвлет-перетворення.

Вейвлети – це особливі функції у вигляді коротких хвиль (сплесків) з нульовим інтегральним значенням і з локалізацією по осі незалежної змінної, здатних до зсуву по цій осі і масштабуванню (розтягуванню/стисненню). Будь-який з найчастіше використовуваних типів вейвлетів породжує повну ортогональну систему функцій. У разі вейвлет-аналізу (декомпозиції) процесу (сигналу) у зв'язку зі зміною масштабу вейвлети здатні виявити відмінність у характеристиках процесу на різних шкалах, а за допомогою зсуву можна проаналізувати властивості процесу в різних точках на всьому досліджуваному інтервалі. Саме завдяки властивості повноти цієї системи можна здійснити відновлення (реконструкцію або синтез) процесу за допомогою зворотного вейвлет-перетворення [10, 11].

Вейвлет-перетворення одномірного сигналу, яким є фінансовий часовий ряд, полягає у його розкладі за системою базисних функцій

$$\psi_{ab}(t) = \frac{1}{\sqrt{a}}\,\psi\!\left(\frac{t-b}{a}\right),$$

сконструйованих з материнського (вихідного) вейвлета $\psi(t)$, що має певні властивості, за рахунок операцій зсуву в часі (b) і зміни часового масштабу (a). Для заданих значень параметрів a и b функція $\psi_{ab}(t)$ і є вейвлет, що породжується материнським вейвлетом $\psi(t)$.

Базисна функція $\psi(t)$ має бути локалізована в часовій та частотній областях і мати такі властивості: нульове середнє значення, обмеженість та автомодельність (останнє означає, що масштабні перетворення не змінюють кількість осциляцій). Вибір $\psi(t)$ визначається метою дослідження.

Кожна функція $\psi(t)$ має свої особливості в часовій та частотній областях, тому за допомогою різних функцій можна краще виявити властивості досліджуваного процесу.

Вейвлет-перетворення сигналів поділяється на:

- неперервне (continuous wavelet transform, CWT) – параметри перетворення (a, b) приймають будь-які дійсні значення;

- дискретне (discrete wavelet transform, DWT) – параметри перетворення (a, b) приймають дискретні значення. При цьому, масштабні перетворення та здвиги базисного вейвлету відбуваються за цілими степенями двійки.

Оскільки множина коефіцієнтів, що отримується при використанні неперервного вейвлет-перетворення є надлишковою, то при дослідженні часових рядів використовується дискретне вейвлет-перетворення [12].

Для ортонормованих вейвлетів маємо дві функції: масштабуюча функція $\varphi(t)$ та вейвлет $\psi(t)$. При здійсненні вейвлет-перетворення сигналу материнський вейвлет $\psi(t)$ дозволяє виділити деталізовану (високочастотну компоненту) з сигналу, що аналізується, у той час, як масштабуюча функція $\varphi(t)$ – згладжену (низькочастотну) складову.

Таким чином, сигнал можна представити у вигляді сукупності послідовних наближень грубої (апроксимуючої) $Am(t)$ і уточненої (деталізуючою) $Dm(t)$ складових

$$y(t) = A_m(t) + \sum_{j}^{m} D_j(t),$$

з подальшим їх уточненням ітераційним методом. Кожен крок уточнення відповідає певному масштабу $a = 2^m$ (тобто рівню m) аналізу (декомпозиції) та синтезу (реконструкції) сигналу.

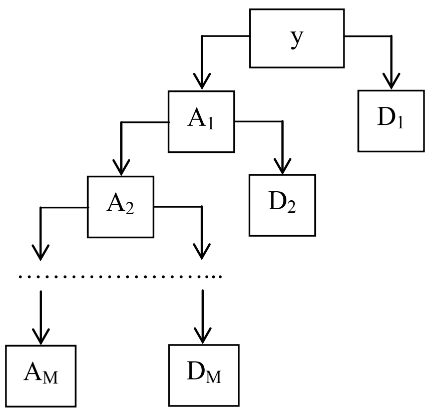
На першому кроці сигнал $y(t)$ розкладається на дві складові:

$$y(t) = A_1(t) + D_1(t) = \sum_k a_{1k}\,\varphi_{1k}(t) + \sum_k d_{1k}\,\psi_{1k}(t).$$

Далі процес декомпозиції може бути продовжений за $A_1(t)$, потім – за $A_2(t)$ і т.д. Сигнал $y(t)$ на рівні розкладання m буде представлений сукупністю коефіцієнтів a_{mk} и d_{mk} [11]:

$$y(t) = \sum_k a_{mk}\,\varphi_{mk}(t) + \sum_k d_{mk}\,\psi_{mk}(t) + \sum_k d_{m-1,k}\,\psi_{m-1,k}(t) + \cdots + \sum_k d_{1k}\,\psi_{1k}(t).$$

Вейвлет-розкладання сигналу $y(t)$, проведене до рівня M, може бути представлено графічно у вигляді дерева (рис. 1): декомпозиция сигналу – зверху – вниз і реконструкція – знизу-вгору.



$m = 0$ $y = \{y_i\}$

$m = 1$ $y = A_1 + D_1$

$m = 2$ $y = A_2 + D_2 + D_1$

$m = \mathrm{M}$ $y = A_M + D_M + D_{M-1} + \ldots + D_1$

Рис. 1. Дерево розкладання

При використанні цього алгоритму на кожному з рівнів низькочастотна складова, яка сформована на попередньому кроці, розкладається на високочастотну та низькочастотну наступного рівня.

Застосуємо розглянуті методи для згладжування рядів ламінарності та на основі отриманих результатів виберемо оптимальний метод попередньої обробки рядів даних для системи моніторингу фондових ринків.

Для отримання достовірних моделей суттєве значення має загальний ступінь згладженості вхідних даних. При моніторингу для аналізу динаміки ринку найбільш важливими точками будуть ті, які розташовані у правого краю ряду, – останні за часом спостереження рівнів показника. Саме вони визначають основні тенденції динаміки показника в сьогоденні та майбутньому.

Оглянемо ряд котирувань фондового індексу DJI за період з 09.08.1999 р. по 11.07.2011 р. довжиною 3000 значень та відповідну міру LAM (рис. 2). Згідно з рис. 2 ми можемо чітко визначити період нормального функціонування ринку, кризу та період релаксації. Візуальний аналіз міри LAM дозволяє розпізнавати різні періоди функціонування ринку. Коли ринок нормально функціонує, траєкторія міри має горизонтальний тренд у деякому діапазоні. Падіння LAM свідчить про кризові явища, а підвищення – про релаксацію та відновлення.

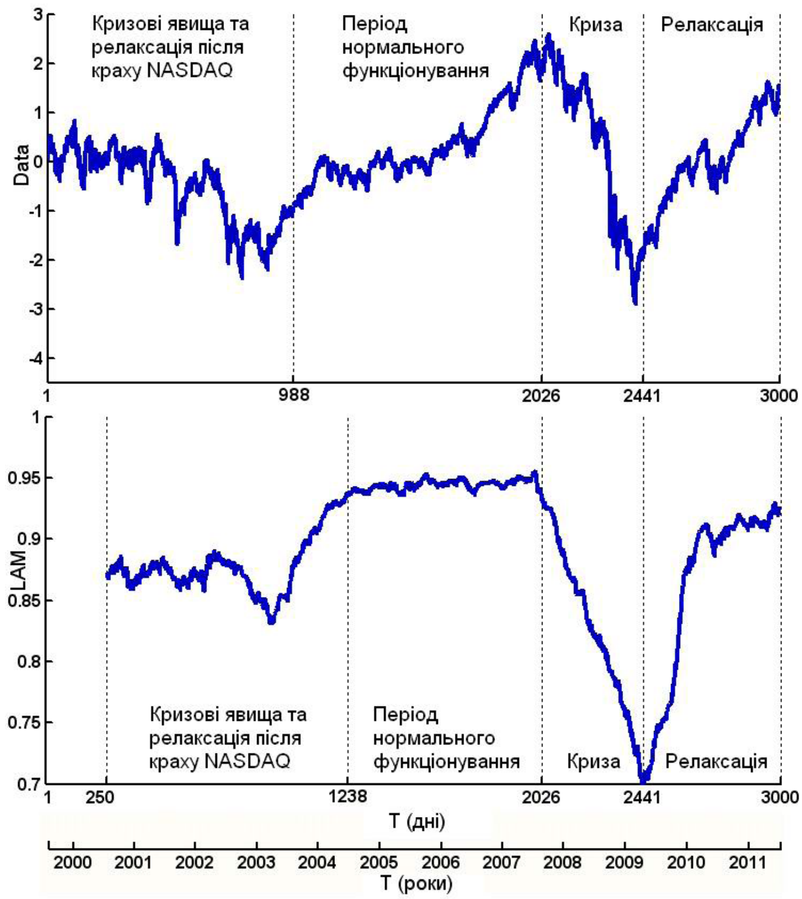


Рис. 2. Динаміка котирувань індексу DJI та відповідна міра LAM

Для дослідження методів згладжування пропонується симулювати процес моніторингу у режимі реального часу. Для цього виберемо довжину частини ряду, що слугуватиме початковим вікном, 1500 значень. Беремо цю частину ряду (з 1 по 1500 точки) та розраховуємо для неї міру LAM вже зі своїм вікном у 250 точок. Потім беремо наступну частину ряду (з 2 по 1501 точки), розраховуємо міру LAM і так до кінця ряду. У результаті маємо набір рядів LAM, що відповідає щоденному моніторингу в режимі реального часу. Далі, згладжуємо ряди LAM в околицях точок зміни режимів функціонування ринку (точка 1 (2026 на рис. 2) – період нормального функціонування ринку/криза; точка 2 (2441 на рис. 2) – криза/релаксація) різними методами та порівнюємо результати.

Для методів локальної апроксимації розрахунки виконувалися за допомогою Curve Fitting Toolbox для MATLAB R2011b та MS Excel. Результати дослідження показали, що методи зваженої ковзної середньої першого порядку (WMA), локальної регресії першого порядку (LR), Lowess та згладжування сплайнами забезпечують достатній загальний рівень згладженості рядів. Задовільне згладжування кінцевих відрізків рядів, що відповідає реальній динаміці процесу, показала тільки LR (рис. 3).

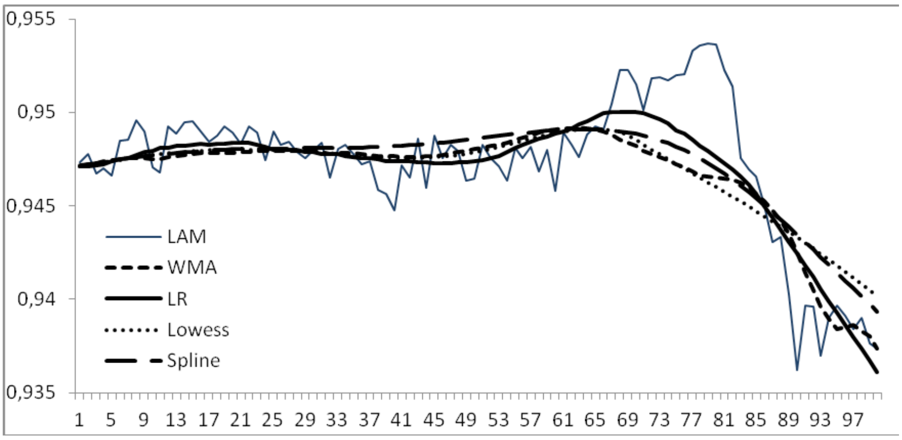


Рис. 3. Порівняння методів згладжування в околиці точки 1

Згладжування рядів за допомогою вейвлет-перетворення (використовувалися ортогональні вейвлети) було проведено за допомогою Wavelet Toolbox для MATLAB R2011b. Найкращі результати за критерієм мінімальної затримки при суттєвому згладжуванні ряду показав вейвлет Добеші 6-го порядку, 5-го рівня розкладу.

Таким чином, нами виділено два методи, які відповідають вимогам згладжування міри LAM. Порівняємо їх між собою. Для цього візьмемо ряди ламінарності з кроком у 5 значень в околиці точок 1 і 2, згладимо їх цими методами та розглянемо останні 50 точок (рис. 4, 5).

Ряд 1_12 закінчується за 15 точок до точки 1, ряд 1_15 включає в себе останнім значенням точку 1, ряди 1_16,1_18 закінчуються через 5 і 15 точок після точки 1 відповідно.

Ряд 2_12 включає в себе останнім значенням точку 2, ряди 1_15, 1_16,1_18 закінчуються через 15, 20 і 30 точок після точки 2 відповідно.

Порівняння отриманих результатів (рис. 4, 5) показало, що обидва методи забезпечують як достатній загальний рівень згладженості рядів, так і задовільне згладжування кінцевих відрізків рядів.

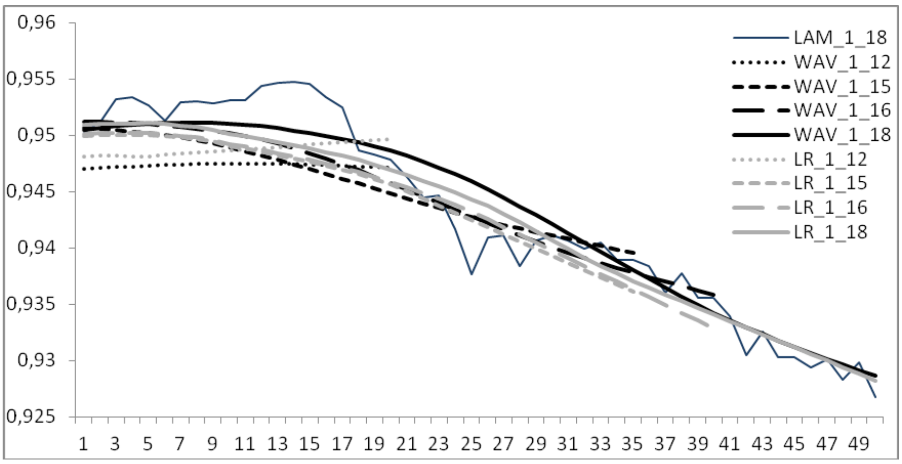


Рис. 4. Згладжені ряди LAM в околиці точки 1

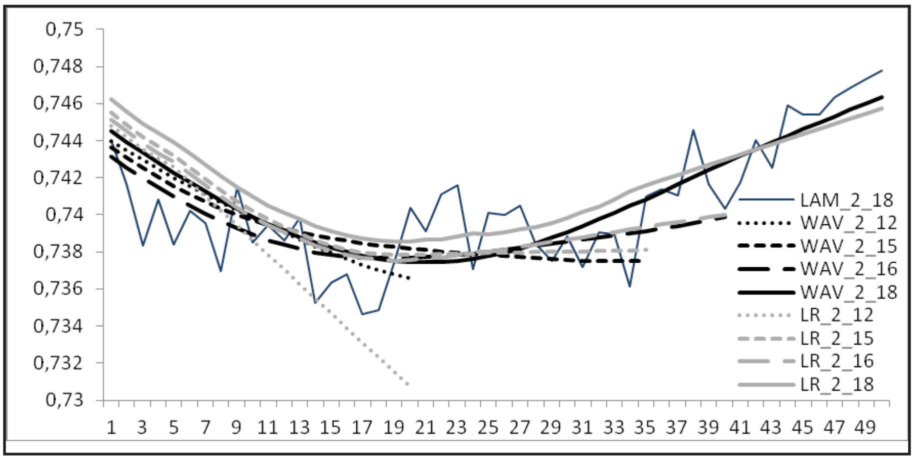


Рис. 5. Згладжені ряди LAM в околиці точки 2

Висновки. Проведений аналіз непараметричних методів згладжування дозволив виявити два методи (локальної регресії першого порядку та вейвлет-перетворення), які відповідають усім вимогам як інструменти попередньої обробки даних для автоматизації виявлення точок переходів між періодами функціонування ринку. Вибір методу залежить від практичної реалізації системи моніторингу фондових ринків.

Список використаних джерел

1. Піскун О. В. Особливості застосування рекурентних діаграм і рекурентного кількісного аналізу для дослідження фінансових часових рядів / О. В. Піскун // Фінансовий простір. — 2011. — № 3 (3). — С. 111–118.

2. Piskun O. Recurrence Quantification Analysis of Financial Market Crises and Crashes [Електронний ресурс] / O. Piskun, S. Piskun. — Режим доступу : http://arxiv.org/pdf/1107.5420.pdf

3. Розенберг Г. С. Экологическое прогнозирование (Функциональные предикторы временных рядов) / Г. С. Розенберг, В. К. Шитиков, П. М. Брусиловский. — Тольятти : ИЭВБ РАН, 1994. — 182 с.

4. Кендалл М. Временные ряды / М. Кендалл / Пер. с англ. и предисл. Ю. П. Лукашина. — М.: Финансы и статистика, 1981. — 199 с.

5. Поллард Дж. Справочник по вычислительным методам статистики / Дж. Поллард / Пер. с англ. В. С. Занадворова; Под ред. и с предисл. Е. М. Четыркина. — М. : Финансы и статистика, 1982. — 344 с.

6. Чураков Е. П. Математические методы обработки экспериментальных данных в экономике: учеб. пособ. / Е. П. Чураков. — М. : Финансы и статистика, 2004. — 240 с.

7. Cleveland W. S. Locally Weighted Regression: An Approach to Regression Analysis by Local Fitting / William S. Cleveland and Susan J. Devlin // Jornal of American Statistical Association. — 1988. — Vol. 83. — № 403. — P. 829–836.

8. Лившиц К. И. Сглаживание экспериментальных данных сплайнами / К. И. Лифшиц. — Томск : Изд-во ТГУ, 1991. — 181 с.

9. Pollock D.S.G. A Handbook of Time-Series Analysis, Signal Processing and Dynamics / D.S.G. Pollock. — London : ACADEMIC PRESS, 1999. — 807 p.

10. Астафьева Н. М. Вейвлет-анализ: основы теории и примеры применения / Н. М. Астафьева // Успехи физических наук. — 1996. — Т. 166. — № 11. — С. 1145–1170.

11. Яковлев А. Н. Введение в вейвлет-преобразования: учеб. пособ. / А. Н. Яковлев. — Новосибирск : Изд-во НГТУ, 2003. — 104 с.

12. Crowley P. M. An intuitive guide to wavelets for economists / P. M. Crowley // Journal of Economic Surveys. — 2007. — Vol. 21. — № 2. — P. 207–267.