

УДК 336.761:338.124.4

Пластун О. Л.

Українська академія банківської справи Національного банку України

## АНАЛІЗ ДОЦІЛЬНОСТІ ПОПЕРЕДНЬОЇ ФІЛЬТРАЦІЇ ДАНИХ В R/S АНАЛІЗІ ФІНАНСОВИХ РИНКІВ

Досліджено доцільність модифікації методології R/S аналізу шляхом попередньої фільтрації даних. В якості технік фільтрації використовувались ковзні середні (прості та зважені) різних періодів, а також метод Ірвіну. Як показав проведений аналіз, фільтрація даних призводить до штучного завищення значень показника Херста, тому модифікація методології R/S аналізу шляхом попередньої фільтрації даних є недоцільною.

**Ключові слова:** персистентність, R/S аналіз, експонента Херста, фільтрація даних, ковзні середні.

**Постановка проблеми.** Принциповим моментом в дослідженні і прогнозуванні фінансових ринків є визначення специфіки цінових коливань: рух цін має випадковий характер чи ринок в даний момент володіє пам'яттю. Поняттям, що відповідає за цю характеристику фінансового ринку є персистентність – здатність стану існувати довше, ніж процес, що створив його. Концепція, що пояснює поведінку фінансових ринків з позиції наявності/відсутності на них пам'яті є фрактальна гіпотеза ринку.

Значний рівень персистентності не лише свідчить про присутність автокореляції в коливаннях цін на фінансові активи, але і обумовлює вибір інструментарію для прогнозування на фінансових ринках.

Існує принципова можливість визначення рівня персистентності ринку за допомогою так званої експоненти Херста. Втім методичні підходи до її розрахунку різняться. З позиції аналізу фінансових даних найбільшим прийнятним є так званий R/S аналіз.

**Аналіз останніх досліджень і публікацій.** Підтвердженням фрактальної природи фінансових ринків було здійснено у працях науковців, Е. Федера, Е. Петерса, В. Ларіна, М. Дубовика, О. Сохацької та багатьох інших. Зокрема, Б. Мандельброт довів, що поведінка фінансових інструментів може бути описана за допомогою фрактальних моделей, що свідчить про, те що в основі існування ринку лежать фрактальні випадкові процеси, повноцінне застосування яких на практиці потребує суттєвих напрацювань у математичному апараті [5]. Е. Петерсом були досліджені основні фрактальні властивості фінансового ринку [8]. О. Сохацька доводить обґрунтованість застосування синергетичного підходу із застосуванням фрактальної геометрії та теорії детермінованого хаосу до аналізу ф'ючерсних ринків і визначає його фрактальну природу [3].

**Виділення не вирішених раніше частин загальної проблеми.** Незважаючи на значну кількість досліджень, присвячених прогнозуванню на фінансових ринках на базі оцінки персистентності ринку, на сьогодні ще залишається цілий ряд аспектів, що потребують розвитку та уточнення. Одним із таких є методологія R/S аналізу.

**Мета статті.** Метою даного дослідження є аналіз доцільності попередньої фільтрації даних при проведенні R/S аналізу фінансових даних. Гіпотеза, що перевіряється в даному випадку: фільтрація дозволяє знизити «рівень шуму» даних, як наслідок результати R/S аналізу стануть більш адекватними.

**Виклад основного матеріалу.** Метод R/S аналізу, вперше застосовано Херстом (1951) у гідро-

логічних дослідженнях [4] і удосконалено Б. Мандельбротом (1972) [5], Е. Петерсом (1991, 1994) [7, 8] та ін. дослідниками фрактальної природи явищ та фінансових ринків. У порівнянні з іншими підходами, метод R/S аналізу є доволі простим, придатним для програмування та графічної інтерпретації.

Опишемо докладніше кожний крок алгоритму R/S-аналізу:

1. Почнемо з часового ряду довжини  $M$ . Перетворимо його у часовий ряд довжини  $N = M - 1$  з наступних логарифмічних відношень:

$$N_i = \log\left(\frac{M_{i+1}}{M_i}\right), i=1,2,3,\dots,(M-1) \quad (1)$$

2. Ділимо цей період часу на  $A$  суміжних підперіодів довжиною  $n$ , так, що  $An = N$ . Позначаємо кожен підперіод за  $Ia$ , з урахуванням того, що  $a = 1, 2, 3, \dots, A$ . Кожен елемент у  $Ia$  позначений  $Nk$ , при цьому  $k = 1, 2, 3, \dots, N$ .

Для кожного  $Ia$  довжини  $n$  середнє значення визначається як:

$$e_a = \frac{1}{n} \sum_{k=1}^n N_{k,a}, k=1,2,3,\dots,N, a=1,2,3,\dots,A \quad (2)$$

3. Часовий ряд накопичених відхилень  $Xk$ , а від середнього значення для кожного підперіоду  $Ia$  визначається як:

$$X_{k,a} = \sum_{i=1}^k (N_{i,a} - e_a) \quad (3)$$

Діапазон визначається як максимальне значення за вирахуванням мінімального значення  $Xk$ , а в межах кожного підперіоду та  $Ia$ :

$$R_{Ia} = \max(X_{k,a}) - \min(X_{k,a}), 1 \leq k \leq n$$

5. Стандартне відхилення, яка розраховується для кожного з півперіоду  $Ia$ :

$$S_{Ia} = \left( \frac{1}{n} \sum_{k=1}^n (N_{k,a} - e_a)^2 \right)^{0,5}$$

6. Кожен діапазон  $RIa$  тепер нормалізується шляхом ділення на відповідний  $SIa$ . Тому повторно нормований розмах протягом кожного  $Ia$  підперіоду дорівнює  $RIa/SIa$ . У кроці 2 вище ми отримали суміжні підперіоди довжини  $n$ . Отже, середнє значення R/S для довжини  $n$  визначається як:

$$(R/S)_n = (1/A) \sum_{i=1}^A (R_{Ia}/S_{Ia})$$

7. Довжина  $n$  збільшується до наступного вищого значення, а  $(M-1)/n$  і має бути цілочисловим значенням. При цьому використовуються значення  $n$ , що включають початкові і кінцеві точки часового ряду, і кроки 1 – 6 повторюються до  $n = (M-1)/2$ .

8. Тепер можна застосувати рівняння  $\log(R/S) = \log(c) + H \log(n)$ , виконуючи регресію методом

найменших квадратів на  $\log(n)$ , як незалежної змінної, і  $\log(R/S)$ , як залежно змінної. Відрізок, що відсікається на координатній осі, є оцінкою  $\log(c)$  константою. Нахил лінії рівняння регресії є оцінкою показника Херста  $H$ . [1]

Показник Херста може набувати значень в інтервалі  $[0;1]$ , а його розрахункові значення знаходяться в наступних межах (табл. 1):

Нами було зроблено спробу удосконалити методологію R/S аналізу шляхом попередньої фільтрації вхідних даних, а також врахувати необхідність перевірки адекватності отриманих результатів. Гіпотеза, що при цьому перевіряється: попереднє згладжування даних дозволяє знизити рівень «шуму» даних, знизити вплив аномальних цінових коливань, таким чином зробивши їх наближеними до відображення фактичного стану речей (ситуації на ринку). Важливим етапом R/S аналізу є перевірка отриманих результатів шляхом розрахунку експоненти Херста для випадково перемішаних даних, що виступали об'єктом аналізу. У теорії випадковим чином перемішані базові дані мають формувати випадковий ряд, для якого коефіцієнт Херста дорівнює 0,5.

З метою перевірки гіпотези нами було проаналізовано прості, втім найбільш живі методи згладжування:

1) згладжування за допомогою ковзних середніх (проста ковзна середня (SMA)), а також зважена ковзна середня (WMA) з періодом згладжування 2 та 5);

2) згладжування методом Ірвіна [2].

З метою перевірки адекватності застосовуваної методології, щоб продемонструвати, що класичний варіант R/S аналізу дає результати близькі до теоретичних, нами було згенеровано декілька масивів даних, що відповідають певним характеристикам рядів за такими методиками:

1) генерація випадкових даних;

2) генерація штучного тренду (персистентного ряду);

3) генерація штучного антиперсистентного ряду;

4) генерація змішаних даних.

Суть методик генерації та їх послідовність наведено у табл. 2.

Впевнившись у адекватності обраної методології розрахунку показника Херста, проведемо ряд перевірок, використовуючи різні методи фільтрації.

Базою для аналізу слугували дані українського індексу – UX за період з 2008 по 2013 роки. Всього часовий ряд включав в себе 1300 значень.

Таблиця 1

Характеристика інтервальних значень показника Херста

Інтервал	Гіпотеза	Розподіл	«Пам'ять» ряду	Тип процесу	Торгові стратегії
$0 \leq H < 0,5$	ряд є фрактальним, виконується гіпотеза фрактальності ринку	«важкі хвости» розподілу	антиперсистентність ряду, з від'ємною кореляцією зміни вартості інструментів	рожевий шум з частою зміною напрямку руху цін	торгівля на ринку є більш ризиковою для індивідуального учасника
$H = 0,5$	справджується гіпотеза ефективності ринку	рух цін на активи є прикладом випадкового броунівського руху (вінерівський процес), часовий ряд розподілений за нормальним законом,	відсутність кореляції зміни вартості активів (пам'яті ряду)	білий шум незалежного випадкового процесу	«переграти» ринок з використанням тієї чи іншої торгової стратегії неможливо
$0,5 < H \leq 1$	ряд є фрактальним, виконується гіпотеза фрактальності ринку	«важкі хвости» розподілу	персистентність ряду, з додатною кореляцією зміни вартості активів	чорний шум	котирування ринку на підпорядковані тренду, схильність до виникнення трендів і криз,

Таблиця 2

Пропоновані методики генерації штучних даних для перевірки розрахунку показника Херста

Назва	Суть	Послідовність
генерація випадкових величин	Створення свого роду віртуального аналогу ряду з відсутністю кореляції між змінами в значеннях ряду	1. Генеруємо випадкову величину ( $p_i$ ) в діапазоні від 0 до 1 2. Розраховується величина зміни штучного індексу ( $\Delta$ ). $\begin{cases} p_i > 0,5, & \Delta_i = 100 \times p_i \\ p_i \leq 0,5, & \Delta_i = -100 \times (1 - p_i) \end{cases}$ 3. Обираємо базову константу (в нашому випадку це було 10000), до якої додаємо отриману величину зміни індексу ( $\Delta$ ). До отриманого значення додається наступна величина зміни індексу і т.д. $\begin{cases} y_i = 10000 + \Delta_i, & i \in [1, n] \\ y_i = y_{i-1} + \Delta_i \end{cases}$
генерація персистентного ряду	Створення віртуального ряду, кореляція зміни значень якого є додатною	1. Задаємо довільну лінійну зростаючу функцію виду $y = a \times x + b$ ; 2. Змінюючи по мірі зростання порядкового номеру елемента в ряді величину $x$ отримуємо трендовий ряд.
генерація антиперсистентного ряду	Створення віртуального ряду, кореляція зміни значень якого є від'ємною	1. Генеруємо випадкову величину в діапазоні від 0 до 1 2. Обираємо базову константу (в нашому випадку це було 10), до якої по черзі додаємо, а потім віднімаємо згенеровану випадкову величину ( $p_i$ ). В результаті чого отримуємо величину часового ряду ( $y_i$ ) 3. для непарних $i$ $y_i = 10 + p_i$ 4. для парних $i$ $y_i = 10 - p_i$ , $i \in [1, n]$
генерація змішаного ряду		1. Генеруємо випадкову величину ( $p_i$ ) в діапазоні від 0 до 1 в інтервалі $i \in [1, n]$ (стовпчик 1). 2. Додаємо стовбець 2 з даними ( $y_i$ ), що необхідно перемішати. 3. Додаємо «Автофільтр» (функція Excel) до отриманих діапазонів. 4. Сортуємо за зростанням (спаданням) стовбець 2 з випадково згенерованими величинами. В результаті стовбець 2 містить перемішані дані базового ряду ( $y_i$ ).

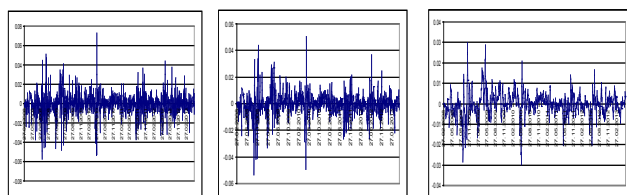
У якості контрольної групи виступили щоденні ціни закриття, тобто нефільтровані дані, а також ряд випадково згенерованих величин. В якості критерію адекватності отриманих даних виступали значення показника Херста для перемішаних даних.

За результатами застосування методик генерації на даних індексу Доу-Джонса можна зробити висновок, що емпіричні значення показника Херста (табл. 3) цілком відповідають теоретичним значенням (для штучно згенерованих даних) для кожного з аналізованих рядів даних.

Таблиця 3  
Результати розрахунку експоненти Херста для різних типів даних

	Випадкові	Персистентні	Антиперсистентні
Значення показника	0.51	0.99	0.08
Значення показника для перемішаних даних	0.53	0.5	0.54

Першим етапом перевірки був візуальний аналіз рядів даних: не фільтрованих та фільтрованих за одним із запропонованих методів. Результати такої перевірки за методом простої ковзної середньої (рис. 1) свідчать, що принципово характер поведінки цінового ряду не змінюється в результаті фільтрації, втім рівень «шуму» стає значно меншим. З позиції фрактальної теорії, візуально фрактальна розмірність зменшується.

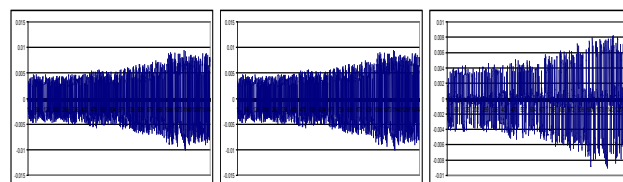


Нефільтровані дані UX      Фільтровані за допомогою SMA 2      Фільтровані за допомогою SMA 5

Рис. 1. Візуальна інтерпретація нефільтрованого та фільтрованих рядів даних SMA

Для того, щоб підтвердити, що реальні характеристики цінового ряду не змінюються і ми лише нейтралізуємо зайві «шуми», нами було проведено аналіз фільтрації випадково згенерованих величин.

У теорії фрактальна розмірність випадково генерованого ряду не має принципово змінюватись від використання процедур фільтрації. Однак, візуальний аналіз (рис. 2) показав, що випадковий ряд теж зменшує фрактальну розмірність в результаті фільтрації.



Випадкові дані      Випадкові дані, фільтровані за допомогою SMA 2      Випадкові дані, фільтровані за допомогою SMA 5

Рис. 2. Візуальна інтерпретація нефільтрованого та фільтрованих рядів випадкових даних SMA

Подібна візуалізація була проведена і для інших методів фільтрації даних. При цьому отримані висновки є аналогічними методу SMA.

Для того, щоб підтвердити візуальний аналіз конкретними числовими характеристиками, нами було розраховано значення показника Херста для кожного з варіантів фільтрації (табл. 4).

Як засвідчив проведений аналіз, фільтрація базових даних призводить до штучного завищення значень показника Херста. Чим більший період усереднення (фактично це означає більший рівень фільтрації даних), тим більше значення показника Херста.

Про штучність результатів свідчить поведінка показника Херста у випадку фільтрації випадкових величин. Якщо б фільтрація була адекватним методом підвищення якості базового ряду даних, то показник Херста, розрахований для ряду випадкових величин не мав б автоматично та пропорційно збільшуватись залежно від періоду усереднення.

Таким чином загальний висновок щодо необхідності та доцільності фільтрації базових даних негативний. Тому гіпотезу про доцільність фільтрації даних вважаємо відкинутою.

Ще однією модифікацією методології R/S аналізу, що перевірялась нами є використання при розрахунку експоненти не кратної кількості груп для загального діапазону даних, а всіх можливих дільників. При цьому базовий ряд за необхідності зменшувався на кілька елементів для забезпечення подільності.

Доречність даної новації обумовлена наступними міркуваннями:

Значення показника Херста для різних варіантів фільтрації даних

Таблиця 4

Дані	Нефільтровані	Метод фільтрації				Метод Ірвіна
		SMA (2)	SMA (5)	WMA (2)	WMA (5)	
Індекс UX	0.67	0.69	0.73	0.69	0.73	0.70
Перемішані дані UX	0.54	0.53	0.54	0.52	0.53	0.49
Випадкові дані	0.51	0.56	0.63	0.55	0.61	0.52
Перемішані випадкові дані	0.53	0.52	0.51	0.51	0.51	0.54

Значення показника Херста для стандартної методології (кратна кількість груп) та модифікованої (некратна кількість груп) для різних варіантів даних

Таблиця 5

	UX (ціни закриття)	Випадкові дані	UX (SMA 5)	UX (WMA 5)	UX (Ірвін)
Стандартна методика	0.67	0.51	0.73	0.73	0.7
Модифікована методика	0.7	0.55	0.78	0.77	0.73

- існує потреба у збільшенні кількості точок для побудови рівняння регресії, що має підвищити якість моделювання;

- не завжди буває достатня кількість даних для розрахунку експоненти за стандартним алгоритмом (мається на увазі умова кратності)

Як показали результати проведеного дослідження (табл. 5), використання некротної кількості груп штучно завищені отримані результати. Отриманий висновок підтверджує як аналіз реальних фінансових даних, так і випадково згенерованих даних.

Отже, загалом (при достатній кількості вхідних даних) недоречно використовувати некротну кількість груп. Тим не менш, у випадку відносно малих груп, а також коли рівняння регресії має

відносно низьку якість можна використовувати з урахуванням поправки на результат (мається на увазі отримані дані слід зменшити на 0,03-0,05 залежно від значень експоненти – чим більше вона тим на більше число зменшувати).

**Висновки і пропозиції.** Загальний висновок щодо запропонованих новацій до методології є негативним. Гіпотезу про доцільність фільтрації даних вважаємо відкинутою, оскільки фільтрація базових даних призводить до штучного завищення значень показника Херста. Чим більший період усереднення (фактично це означає більший рівень фільтрації даних), тим більше значення показника Херста. Отже, при проведенні R/S аналізу фінансових даних застосовувати попередню фільтрацію даних недоцільно.

#### Список літератури:

1. Гачков А.А. Рандомизированный алгоритм R/S-анализа финансовых рядов [Электронный ресурс] / А. А. Гачков. – Режим доступа : <http://www.math.spbu.ru/user/gran/soi5/Gatchkov5.pdf>.
2. Наконечний С.І. Економетрія : підручник / С.І. Наконечний, Т. О.Терещенко, Т. П. Романюк. – Вид. 3-тє, доп. та перероб. – К. : КНЕУ, 2004. – 520 с. – ISBN 966-574-630-8.
3. Сохацька О.М. Ф'ючерні ринки: глобальні тенденції та становлення в Україні : автореф. дис. на здобуття наук ступеня доктора екон. наук : спец. 08.05.01 «Світове господарство і міжнародні економічні відносини» / О.М. Сохацька. – Тернопіль, 2003. – 37 с.
4. Hurst H.E. Long-term Storage of Reservoirs. Transactions of the American Society of Civil Engineers / H. E. Hurst. – 1951. – 799 p.
5. Mandelbrot B., 1972. Statistical Methodology For Nonperiodic Cycles: From The Covariance To Rs Analysis // Annals of Economic and Social Measurement, Volume 1, number 3, 259-290.
6. Mandelbrot B.B. Robustness of the rescaled range R/S in the measurement of non-cycling long-run statistical dependence // Water Resources Research. 1969. V. 5. № 5. P. 967-988.
7. Peters Edgar E. Chaos and Order in the Capital Markets: A New View of Cycles, Prices, and Market Volatility/ Edgar E. Peters. – NY. : John Wiley and Sons, Inc, 1991 – 228 p.
8. Peters Edgar E. Fractal Market Analysis: Applying Chaos Theory to Investment and Economics / Edgar E. Peters – NY. : John Wiley & Sons, 1994. – 336 p.

**Пластун А. Л.**

Украинская академия банковского дела Национального банка Украины

#### АНАЛИЗ ЦЕЛЕСООБРАЗНОСТИ ПРЕДВАРИТЕЛЬНОЙ ФИЛЬТРАЦИИ ДАННЫХ В R/S АНАЛИЗЕ ФИНАНСОВЫХ РЫНКОВ

##### Резюме

Исследована целесообразность модификации методологии R/S анализа путем предварительной фильтрации данных. В качестве техник фильтрации использовались скользящие средние (простые и взвешенные) разных периодов, а также метод Ирвина. Как показал проведенный анализ, фильтрация данных приводит к искусственному завышению значений показателя Херста, поэтому модификация методологии R/S анализа путем предварительной фильтрации данных является нецелесообразной.

**Ключевые слова:** персистентность, R/S анализ, экспонента Херста, фильтрация данных, скользящие средние.

**Plastun O. L.**

Ukrainian academy of banking of the National bank of Ukraine

#### ABOUT THE NECESSITY OF DATA FILTRATION IN R/S ANALYSIS OF FINANCIAL MARKETS

##### Summary

The necessity of modifying the methodology of R/S analysis by preliminary data filtration is analyzed. As the filtering techniques moving averages (simple and weighted) of different periods and method Irvine are used. According to results data filtration artificially inflates the values of Hurst exponent. So modification methodology as R/S analysis by preliminary data filtration is unnecessary.

**Key words:** persistence, R/S analysis, Hurst exponent, data filtration, moving averages.