

УДК 51-77

**БАРАНОВСЬКИЙ О.М.**, аспірант НТУУ “КПІ”,  
**КАЧИНСЬКИЙ А.Б.**, доктор технічних наук,  
**ДОБРОВОЛЬСЬКИЙ Є.Л.**, аспірант НТУУ “КПІ”,  
**ЛАНДЕ Д.В.**, доктор технічних наук

## **ХАРАКТЕРИСТИКИ ІНФОРМАЦІЙНИХ ПОТОКІВ, ПОВ’ЯЗАНИХ З РЕФОРМУВАННЯМ ПОДАТКОВОЇ СИСТЕМИ УКРАЇНИ**

***Анотація.** Робота присвячена дослідженню властивостей інформаційних потоків, що відображають прояви громадянської активності населення під час подій податкового реформування в Україні в 2010 – 2011 роках. В статті розглядається процес отримання та первісної обробки часових рядів спостережень, що характеризують інформаційні потоки, а також здійснюється вивчення їх характеристик за допомогою методу нелінійної динаміки.*

***Ключові слова:** контент-моніторинг, інформаційний потік, фрактальний аналіз.*

***Аннотация.** Работа посвящена исследованию свойств информационных потоков, отражающих проявления гражданской активности населения во время событий налогового реформирования в Украине в 2010 – 2011 годах. В статье рассматривается процесс получения и первичной обработки временных рядов, характеризующих информационные потоки, а также осуществляется изучение их характеристик с помощью метода нелинейной динамики.*

***Ключевые слова:** контент-мониторинг, информационный поток, фрактальный анализ.*

***Summary.** The work is about researching of the properties of information flows, reflecting the manifestations of civic engagement of the population during the events of the tax reform in Ukraine in 2010 – 2011 respectively. This article discusses the process of obtaining and processing of time series describing the flow of information, as well as by the study of their properties by the method of nonlinear dynamics.*

***Keywords:** content-monitoring, information flow, fractal analysis.*

***Постановка проблеми.** При вивченні соціальних процесів однією з актуальних проблем є прогнозування майбутніх подій або явищ. Традиційно складання прогнозів спирається на гуманітарні науки і тому базується на вузькому наборі фактів та явищ, що ускладнює застосування математичних методів. Одним з найперспективніших методів моделювання соціальних процесів є дослідження подій, явищ або процесів в інформаційному середовищі, зокрема в мережі Інтернет, що стала одним з основних джерел новин.*

*Об’єктами дослідження обрано інформаційні потоки української частини мережі Інтернет, що відображали прояви громадянської активності в період прийняття Податкового кодексу України впродовж 2010 – 2011 років. Використовувалась класифікація проявів громадянської активності населення, запропонована Т. Жиро [1].*

***Метою статті** є дослідження властивостей інформаційних потоків (хаотичність, випадковість, ергодичність), наявності та довжини пам’яті потоку, прогнозування тенденції подальшого розвитку.*

***Виклад основних положень.** В попередній роботі [2] для формування відповідного інформаційного потоку авторами було розглянуто принципи побудови вербальних моделей та отримання даних за допомогою методів контент-аналізу та системи InfoStream.*

*В даній роботі досліджувались інформаційні потоки щодо наступних типів політичної активності населення: страйк, лобізм, демонстрація, бойкот, судовий процес, тиск та протест, політичне насилля, протест та делегітимізація, громадянська непокора в контексті проблеми реформування податкової системи України.*

Для системи InfoStream було розроблено вербальні моделі, які дозволяють отримати показник кількості статей, що були зафіксовані системою у визначений день дослідження та відповідають визначеному типу активності населення. Використані вербальні моделі для кожного виду політичної активності наведені в Табл. 1.

Таблиця 1

Вид активності	Вербальна модель
Демонстрація	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ)) &(митинг пикет мітинг пікет демонстрац) &(country.UA geo.UA*)
Лобізм	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))&(Лобі Лобби) &(country.UA geo.UA*)
Судовий процес	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))&((Судов~процес) (Судебн~процесс)) &(country.UA geo.UA*)
Страйк	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))&(забастов страйк) &(country.UA geo.UA*)
Тиск та протест	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))&((давлен тиск)&(протест)) &(country.UA geo.UA*)
Бойкот	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))&(мовчанка бойкот) &(country.UA geo.UA*)
Громадська непокоря	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))& ((граждан~неповиновен) (громад~непокоря)) &(country.UA geo.UA*)
Делегітимізація	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))&(Делегітиміз Делегитимиз) &(country.UA geo.UA*)
Політичне насилля	((подат~кодекс) (налогов~кодекс)  (подат~реформ) (налогов~реформ))&((полити~насил) (політич~насил)) &(country.UA geo.UA*)

Приклад отриманих первісних результатів для виду політичної активності “Тиск та протест” відображено на Рис. 1. Графічне відображення коефіцієнтів автокореляції для ряду, який відповідає типу активності “Тиск та протест”, свідчить про наявність розподілу кореляційних властивостей за сімома рівнями значень, що відповідають семи дням тижня (Рис.2). Таким чином, існує необхідність вилучення тижневого тренду з даних для отримання особливостей реальної поведінки інформаційного потоку.

Для видалення періодичної складової авторами був використаний запропонований в [3] метод зваженої “ковзної середньої”. Таким чином був отриманий новий ряд:

$$\xi'_i = \frac{1}{7} \sum_{j=i-3}^{j=i+3} \frac{\xi_j}{N_j}, \tag{1}$$

де:  $\xi_j$  – кількість статей, що були відібрані за вербальною моделлю, за j-ий день дослідження,  $N_j$  – загальна кількість статей, що було зафіксовано системою за j-ий день дослідження.

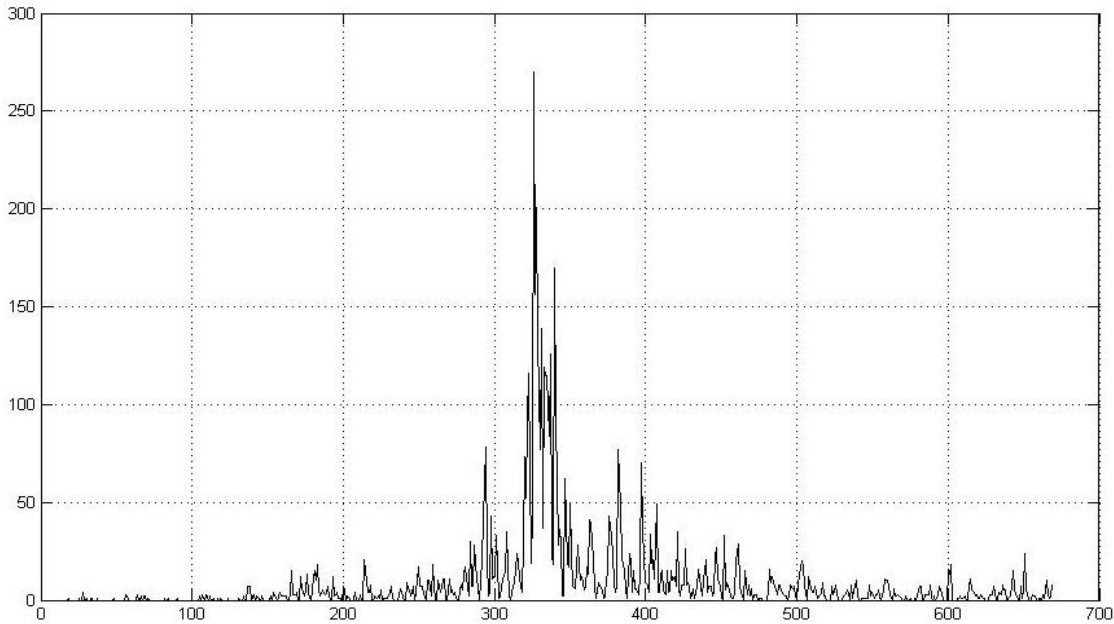


Рис. 1. Кількість публікацій, що відповідають типу активності “Тиск та протест”

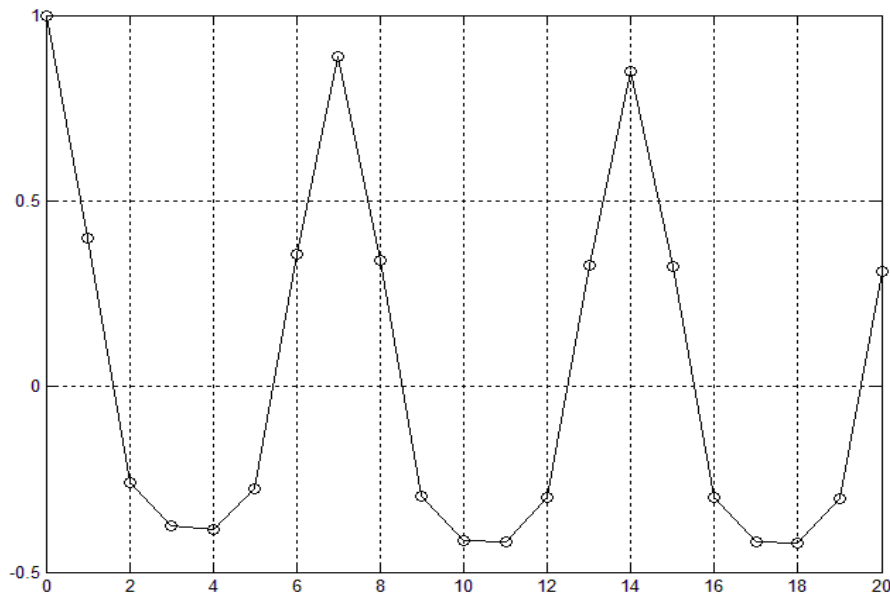


Рис. 2. Коефіцієнти автокореляції ряду, що відповідає типу активності “Тиск та протест”

На Рис. 3 зображено графічне представлення згладженого ряду, що відповідає типу активності “Тиск та протест”.

Графічне відображення коефіцієнтів автокореляції для згладженого ряду представлено на Рис. 4 та не містить піків та явно виражених гармонічних складових. Тобто в новому ряду не існує відкритої або скритої тижневої періодичності.

Вважається, що видалення періодичної складової дозволить досліджувати особливості реальної поведінки інформаційних потоків, що відповідають проявам політичної активності населення, та виключити вплив періодичності електронних ЗМІ [3]. Наведений вище алгоритм обробки був застосований для всіх отриманих рядів.

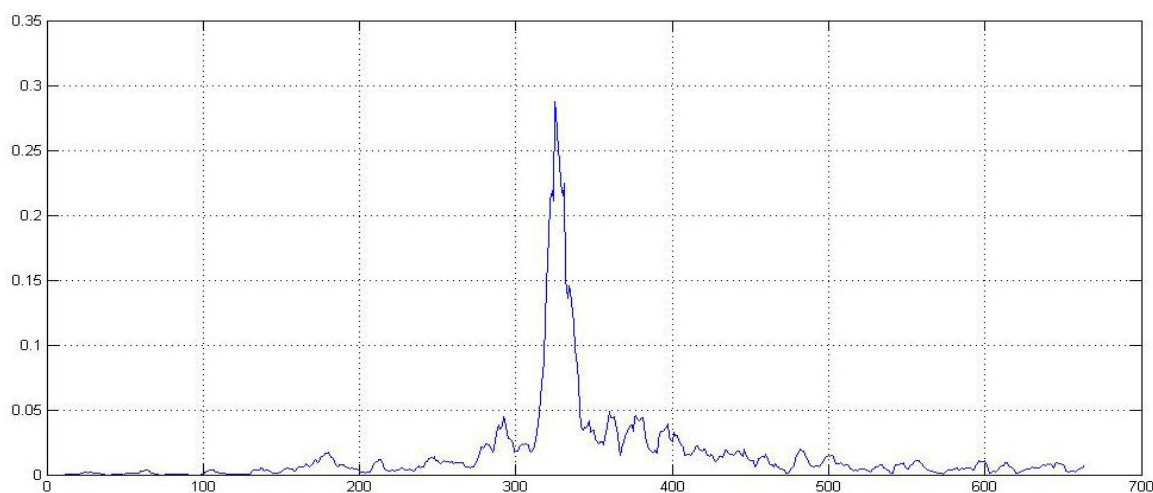


Рис. 3. Згладжений ряд, що відповідає типу активності “Тиск та протест”

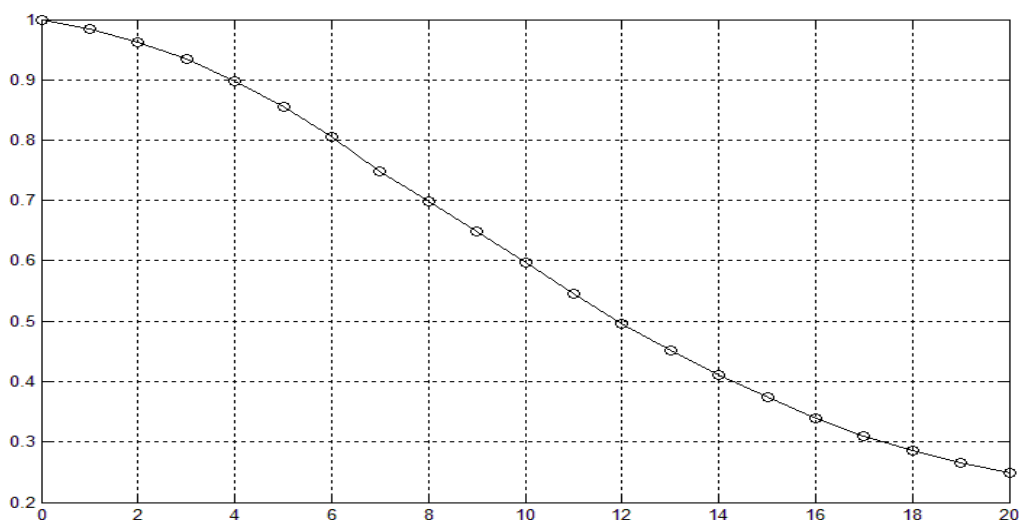


Рис. 4. Коефіцієнти автокореляції згладженого ряду, що відповідає типу активності “Тиск та протест”

Наступним кроком дослідження було визначення природи отриманих рядів. Згідно з [4] спостереження показують, що в будові і функціонуванні систем, що пов’язані з соціумом, спостерігається загальносистемна властивість самоподібності. Тобто системи однаково влаштовані в широкому діапазоні просторових, тимчасових або кількісних масштабів, що обумовлено деякою мірою масштабної симетрії. Тому малі фрагменти системи можуть бути подібні цілій системі. Для аналізу самоподібності використовують R/S-аналіз. Поняття фракталу було введено Бенуа Мандельбротом у 1970-ті роки [5]. Термін походить від латинського fractus, прикметника від дієслова frangere – ламати, розбивати на частини. Велика кількість спостережень природних процесів приводить до часових залежностей або рядів вимірів. Фрактальна статистика – галузь знань, що активно розвивається та володіє потужними інструментами опису складних самоподібних об’єктів і процесів. Поява цього підходу дає нове наближення для опису складних об’єктів на мові математики та дозволяє ввести кількісні характеристики для самоподібності, уявлення про яку на якісному рівні давно увійшло в різноманітні сфери знань. Часові послідовності вимірів таких величин, як температура, сток рік, кількість

опадів, індекси фінансових ринків, кількість сонячних плям можливо досліджувати за допомогою методу нормованого розмаху, або методу Херста [5].

Формула Херста (2) вперше була застосована при аналізі сумарних витрат води в річці Ніл за тривалий період часу.

$$R/S = (A * N)^H, \quad (2)$$

де:  $R$  – розмах між максимальним та мінімальним значеннями показника;  $R$  – середньоквадратичне відхилення;  $A$  – масштабна постійна, що залежить від масштабу використовуваних чисел;  $N$  – обсяг вибірки або час спостережень;  $H$  – показник Херста.

Варіанти оцінки показника Херста:  $H = 0,5$ ;  $0 \leq H < 0,5$ ;  $0,5 < H < 1$ .

При  $0,5 < H < 1$  ряд є персистентним, або трендостійким, тобто, якщо ряд зростав у попередній період, то ймовірніше, що він буде зберігати цю тенденцію. Трендостійкість поведінки, або сила персистентності, збільшується при наближенні  $H$  до одиниці.

При  $0 \leq H < 0,5$  ряд є антиперсистентним, тобто, якщо ряд зростав у попередній період, то ймовірніше, буде спадати в наступному періоді. Стійкість такої поведінки залежить від того, наскільки  $H$  наближується до нуля.

Чим ближчий  $H$  до 0,5, тим більше події є випадковими та некорельованими.

Для кожного з рядів, що відображають визначені види політичної активності, було обчислено показник Херста та побудовані (Рис. 5) в подвійних логарифмічних координатах криві залежності нормованого розмаху від часу спостережень. Нахил апроксимуючої прямої і визначає показник Херста.

Отримані результати свідчать про те, що ряди, які досліджуються авторами, є динамічними, не випадковими та мають відносно довгу пам'ять. Всі отримані показники Херста істотно перевищують 0,5.

Візуальний аналіз кривих для таких видів політичної активності, як демонстрації (показник Херста – 0,80), тиск та протест (0,86) та страйк (0,76) свідчить про те, що на часовому інтервалі приблизно в 160 діб спостерігається деякий цикл, який проявляється у відхиленні від лінійної залежності. Наявність однакового циклу в даних рядах свідчить про зв'язок між інформаційними потоками, що відображають вказані типи політичної активності. Цікаво також те, що при детальному аналізі можна зробити висновок про наявність циклу довжиною близько 30 діб в кожному з цих типів активності.

Щодо рядів, що відносяться до бойкоту (0,99), лобізму (0,99) та судових процесів (0,96), то криві вказують на наявність циклів зі зміщенням приблизно в 30 діб та їх закінчення близько 160 доби. Тобто на відміну від попередніх рядів цикли у даному випадку становлять близько 130 днів, але закінчуються одночасно з розглянутим у попередньому випадку.

Дуже схожу динаміку демонструють залежності для громадянської непокори (0,71) та делегітимізації (0,70). Цикли в даних рядах закінчуються близько 310 доби, але в ряді, що відповідає громадянській непокорі, цикли починаються зі зміщенням на 45 діб – близько 100 доби, а в делегітимізації зміщення початку циклу становить 55 діб.

Згідно з Е. Петерсом [6] перевірка ефективності  $R/S$ -аналізу може бути реалізована шляхом випадкового перемішування даних та застосування методу до перемішаного ряду. У новому ряді порядок елементів буде повністю відрізнятись від вихідного, а частотний розподіл залишиться незмінним. Таким чином, якщо елементи ряду є незалежними, то показник Херста має залишитись незмінним, бо відсутня кореляція між спостереженнями. Якщо ж існує довгострокова пам'ять, то перемішування даних знищить структуру системи, а  $H$  має наближуватись до 0,5.

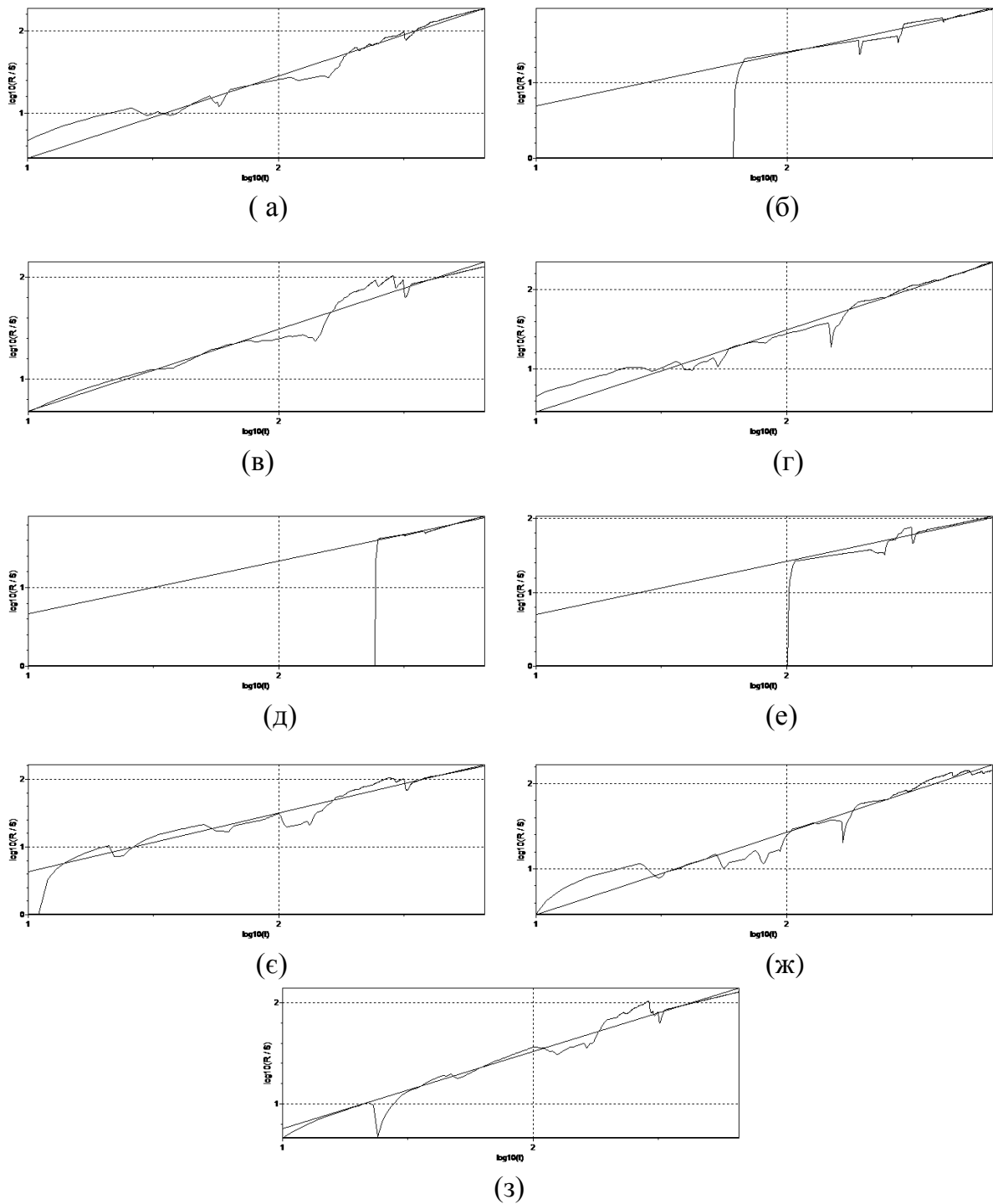


Рис. 5. Криві залежностей нормованого розмаху від часу спостережень: а – бойкот, б – делегітимізація, в – демонстрації, г – лобізм, д – політичне насилля, е – громадянська непокора, є – тиск та протест, ж – судовий процес, з – страйк.

Авторами статті було проведено обчислення показника Херста для випадковим чином перемішаних рядів. Наведені в Табл. 2 результати свідчать про те, що отримані ряди дійсно мали довгу пам’ять, а при перемішуванні була порушена структура процесу, знищений ефект довгої пам’яті, та спостереження стали незалежними.

Таке падіння величини  $H$  свідчить про істинність твердження Мандельброта про те, що R/S-аналіз ефективний незалежно від розподілу часового ряду [6].

Таблиця 2

Вид активності	Показник Херста	Відхилення
Демонстрація	0,42	±0,16
Лобізм	0,47	±0,16
Судовий процес	0,49	±0,10
Страйк	0,44	±0,16
Тиск та протест	0,42	±0,17
Бойкот	0,49	±0,13
Громадська непокора	0,53	±0,17
Делегітимізація	0,43	±0,11
Політичне насилля	0,47	±0,12

### **Висновки.**

В ході проведеного дослідження було зроблено висновок про те, що інформаційні потоки, що відображають прояви громадянської активності в період прийняття Податкового кодексу України, не є випадковими, а мають властивості реалізацій, породжуваних хаотичними системами.

Кожному з потоків властива наявність довгої пам'яті. Для окремих випадків було отримано показники довжини пам'яті для потоку. Зроблено висновок про наявність в деяких потоках циклів різної довжини.

Отримані показники Херста для кожного з інформаційних потоків, які розглядалися, свідчать, що останні є персистентними.

### **Використана література**

1. Жиро Т. Политология / Жиро Т. ; науч. ред., авт. предисл. С.В. Быков, И.И. Царьков. – Харьков : Гуманит. Центр, 2006. – 425 с.
2. Качинський А.Б., Добровольський Є.Л., Барановський О.М., Ланде Д.В. Прикладні аспекти застосування моделей взаємозв'язку “ЗМІ-соціум-державна політика” на прикладі реформування податкової системи України // Інформаційна безпека людини, суспільства, держави. – № 2(6) 2011. – 158 с.
3. Ландэ Д.В., Снарский А.А., Брайчевский С.М., Дармохвал А.Т. Моделирование динамики новостных текстовых потоков : сборник работ участников конкурса Интернет-математика 2007. – Екатеринбург : Изд-во Урал. ун-та, 2007. – 224 с.
4. Давыдов А.А. Системный подход в социологии: новые направления, теории и методы анализа социальных систем / А.А. Давыдов. – М. : КомКнига, 2005. – 328 с.
5. Федер Е. Фракталы / Е. Федер. – М. : Мир, 1991. – 254 с.
6. Петерс Э. Хаос и порядок на рынках капитала. Новый аналитический взгляд на циклы, цены и изменчивость рынка / Э. Петерс. – Москва: Мир, 2000. – 277 с.

~~~~~ \* \* \* ~~~~~