

П.С. КНОПОВ

Інститут кібернетики ім. В.М. Глушкова НАН України, Київ, Україна,
e-mail: *knopov1@yahoo.com*.

А.С. КОРХІН

Придніпровська державна академія будівництва та архітектури, Дніпро, Україна,
e-mail: *a.s.korkhin@gmail.com*.

**ВИЗНАЧЕННЯ КУСКОВО-ЛІНІЙНОГО ТРЕНДУ
НЕСТАЦІОНАРНОГО ЧАСОВОГО РЯДУ НА ОСНОВІ
ІНТЕЛЕКТУАЛЬНОГО АНАЛІЗУ ДАНИХ.
II. МАШИННІ ЕКСПЕРИМЕНТИ ТА РОЗВ'ЯЗАННЯ
ПРАКТИЧНОЇ ЗАДАЧІ^{1,2}**

Анотація. Наведено результати апробації методу побудови кусково-лінійного тренду, що може як мати розриви у точках перемикань, так і бути неперервним у цих точках, тобто являти собою лінійний сплайн. Розглянуто приклад застосування методу для побудови лінійної регресії з перемиканнями, яка має дві незалежні змінні з трендом. Розв'язано задачу апроксимації сплайном часового ряду логарифмів числа інфікованих COVID-19 в Україні.

Ключові слова: тренд, регресія, точка перемикання, сплайн, обчислення у реальному часі.

ВСТУП

Нестационарні часові ряди застосовують для дослідження економічних рядів, кількості інфікованих в епідеміології, рядів спостережень температури поверхні Землі та ін. Їхні тренди потрібні як для аналізу спостережуваних явищ, так і для прогнозування (нерідко в першу чергу). З цією метою, якщо механізм явища невідомий або дуже складний, застосовують метод чорної скриньки, коли явище описують деякою формулою, що має сенс тренду, аргументом якої є час. Це дає змогу використовувати формулу для прогнозування. Її коефіцієнти оцінюють за передісторією.

Тренд може мати підйоми, плато та спади. Аналітичний вираз, що їх визначає, достатньо складний, може містити нелінійності. Якщо до того ж потрібно підвищувати адекватність виразу для тренду, додаючи обмеження на коефіцієнти формули тренду, то сформульована задача оцінювання стає складною. Як приклад, пошлемося на роботи [1, 2].

Як альтернативу можна запропонувати підхід, що зводиться до опису тренду кусково-лінійною функцією з невідомими точками перемикання. Метод її побудови розглянуто в [3].

У цій статті описано апробацію методу, а також розв'язано практичну задачу з його використанням.

1. ПОЗНАЧЕННЯ ТА ТЕРМІНОЛОГІЯ

У зручному для розуміння вигляді наведемо короткий опис методу, що викладений у [3].

¹Роботу виконано в межах наукового проекту РК № 0122U000552 «Комплексний аналіз робастних профілактичних та адаптивних заходів управління продовольством, енергією, водою та соціальною сферою в умовах системних ризиків та наслідків COVID-19».

²Продовження. Початок в № 1, т. 60, 2024.

Розглянемо часовий ряд

$$y_t = \mathbf{x}'_t \alpha_{ii}^0 + \varepsilon_t, \quad i=1, 2, \dots, \quad t=1, 2, \dots, \quad (1)$$

де штрих позначає транспонування, $\mathbf{x}_t = [1 \ t]'$, $\alpha_{ii}^0 \in \mathbb{R}^2$ — дійсні невідомі величини параметрів регресії, ε_t — випадкова величина. При цьому

$$\alpha_{ii}^0 = \alpha_i^0 = \text{const} \quad \text{для } t \in I_i^0 = [t_{i-1}^0 + 1, t_i^0], \quad i=1, 2, \dots, \quad t_0^0 = 0, \quad (2)$$

де t_i^0 — i -та точка перемикавання, в ній виконується перехід з однієї прямої на іншу пряму.

Вважаємо, що випадкові величини ε_t , $t=1, 2, \dots$, мають нульове математичне сподівання та нормально розподілені.

Точки перемикань оцінюватимемо послідовно двома способами за надходженням нових спостережень.

Спосіб 1. Нехай \hat{t}_i , $i=0, 1, 2, \dots$, — остання оцінювана точка перемикавання, причому $\hat{t}_0 = t_0^0 = 0$. Вважаємо, що на інтервалі часу $[\hat{t}_i + 1, T]$ всі точки містяться на одній прямій. Якщо у точці T відбувається перемикавання, то нова пряма визначається на інтервалі часу $[T + 1, T + L]$.

Отже, є два варіанти: 1) на інтервалі $[\hat{t}_i + 1, T + L]$ дані описуються однією лінійною регресією; 2) на цьому ж інтервалі є дві лінійні регресії на підінтервалах $[\hat{t}_i + 1, T]$ і $[T + 1, T + L]$. Вибирають той варіант, для якого загальна сума квадратів залишків значно менша, ніж ця ж сама величина для іншого варіанта. У разі вибору другого варіанта визначають наступну точку перемикавання: $\hat{t}_{j+1} = T$. Для іншого випадку аналізують наступну точку, а саме $T + 1$ і т.д.

Для вибору одного з двох варіантів потрібно перевірити, чи виконується нерівність

$$F(\hat{t}_i + 1, T + L) \leq F_p(q_1, q_2), \quad i=0, 1, \dots, \quad (3)$$

де F -відношення $F(\hat{t}_i + 1, T + L)$ уведено для зіставлення сум квадратів залишків двох варіантів за формулою (9) з [3]; $F_p(q_1, q_2)$ — верхня $100p$ %-ва точка розподілу Фішера, $q_1 = n$, $q_2 = T - \hat{t}_i + L - 2n$, $n=2$.

Якщо нерівність (3) виконується, то нульову гіпотезу про відсутність у точці T перемикавання не відхиляють на $100p$ %-му рівні, тобто вибирається перший варіант. В іншому випадку приймається гіпотеза про наявність перемикавання у точці T і приймається другий варіант.

Нерівність (3) може не виконуватися на деякому інтервалі часу J_p . Тоді як точку перемикавання вибирають період часу, якому відповідає перший максимум $F(\hat{t}_i + 1, T + L)$ на J_p .

Спосіб 2. Так саме, як і для способу 1, розглянемо ситуацію, коли з точки $\hat{t}_i + 1$ у точку T проведено одну пряму, і потрібно визначити, чи існує у T перемикавання. Для цього спрогнозуємо залежну змінну $\hat{y}_{T+\tau}$ на $\tau=1, 2, \dots, L$ періодів часу, продовжуючи цю пряму з точки T уперед. Визначимо довірчий інтервал для прогнозованої величини, який накріє її з ймовірністю $1 - p$, за формулою

$$\hat{y}_{T+\tau} - t_p(q) \hat{\sigma}_f(T + \tau) \leq y_{T+\tau} \leq \hat{y}_{T+\tau} + t_p(q) \hat{\sigma}_f(T + \tau), \quad (4)$$

де $t_p(q)$ — це 100 p %-ва точка розподілу Стьюдента з числом ступенів свободи $q = T - \hat{t}_i - 2$; $\hat{\sigma}_f(T + \tau)$ — оцінка середньоквадратичної похибки прогнозу точки розподілу Стьюдента з числом ступенів свободи $\hat{\sigma}_f^2(T + \tau) = \hat{\sigma}^2(1 + \mathbf{x}'_{T+\tau} \hat{\mathbf{K}} \mathbf{x}_{T+\tau})$, де $\hat{\mathbf{K}}$ — оцінка кореляційної матриці щодо $\boldsymbol{\alpha}^0$ — багатовимірного параметра регресії (див. (2)), отримана за спостереженнями на інтервалі часу $[\hat{t}_i + 1, T]$.

Якщо обидва прогнози не потрапляють у свої довірчі інтервали, то схоже, що в точці T є перемикання. Точніше, якщо виконується рівність

$$\varphi_P(T) = 0, \quad (5)$$

то згідно зі сказаним вище гіпотеза $H_0: \hat{t}_{i+1} = T$ не відхиляється на 100 P %-му рівні. В іншому випадку вона відхиляється.

Тут $P = p^2$, де ймовірність p визначена в (4); функція

$$\varphi_P(T) = \varphi_{1P}(T) + \varphi_{2P}(T), \quad 0 \leq \varphi(T) \leq 2, \quad (6)$$

де

$$\varphi_{iP}(T) = \begin{cases} 1, & \text{якщо } y_{T+i} \text{ задовольняє (4),} \\ 0, & \text{якщо } y_{T+i} \text{ не задовольняє (4),} \end{cases} \quad i = 1, 2.$$

Для стислості будемо називати нерівність (3) критерієм 1, а функцію $\varphi_P(T)$ — критерієм 2. Один із них вважатимемо головним, інший — допоміжним.

Для підвищення надійності висновку про наявність перемикання у точці T вимагатимемо, щоб цей висновок одночасно збігався за обома критеріями. Під збігом критеріїв будемо розуміти таке. Нехай період часу $T^{(1)}$ за критерієм 1 вважатимемо точкою перемикання. Таке рішення вважатимемо прийнятим за обома критеріями, якщо за критерієм 2 точкою перемикання є така точка $T^{(2)}$, що $T^{(1)} - a \leq T^{(2)} \leq T^{(1)} + a$, $0 \leq a \leq (L - 1)$.

У всіх розв'язаних нижче задачах $L = 2$, $n = 2$. Тоді наведена вище нерівність матиме вигляд

$$T^{(1)} - a \leq T^{(2)} \leq T^{(1)} + a, \quad 0 \leq a \leq 1. \quad (7)$$

Звідси за умови $a = 0$ отримаємо $T^{(1)} = T^{(2)} = T$.

За умовою (7) можна увести точки перемикань з деяким надлишком. Подальша перевірка лінійних гіпотез дає змогу вилучити зайві точки.

Уведемо ще одне апріорне обмеження на точки перемикань з метою забезпечення мінімальної надмірності спостережень для оцінювання коефіцієнтів прямих, які створюють тренд. Вимагаємо, щоб виконувалась нерівність

$$|t_i^0 - t_{i-1}^0| \geq cn, \quad i = 1, 2, \dots, \quad (8)$$

де $c = 3$, $n = 2$.

У розв'язаних нижче задачах використовують звичайно 5 %-й і 1 %-й рівні значущості. Причому рівні значущості за обома критеріями не обов'язково збігаються для однієї і тієї ж точки перемикання. Тому ставимо таку ж мету, що і для (7): знайти максимальну кількість точок перемикань, які можна корегувати, вилучаючи незначущі точки (для яких не відхиляється гіпотеза про відсутність у них перемикань).

Метод з [3] дає змогу знайти точки перемикань, не гарантуючи неперервність тренду. Процес обчислення точок перемикань можна зобразити у вигляді «хвилі», що рухається за новими даними на відстані в L періодів часу.

Якщо за апріорною інформацією тренд є неперервною функцією часу, то будується сплайн з уточненням точок перемикань, використовуючи метод, на-

ведений у [3, розд. 3]. Побудову сплайну можна зобразити у вигляді другої хвилі, яка рухається за першою хвилею з відставанням, що дорівнює відстані між двома останніми оціненими точками перемикачів. Оцінені точки перемикачів у тренді у разі наявності розривів (перша хвиля) можна уточнювати поступовим виявленням зайвих точок. Для цього розглядають оцінки трьох точок перемикачів, що слідує одна за одною, $\hat{t}_{i-2}, \hat{t}_{i-1}, \hat{t}_i$, причому $\hat{t}_i - \hat{t}_{i-1} < cn$, де $c=3, n=2$. Тому умова (8) не виконується.

Вилучати (або не вилучати) перемикачів з розривом у точці \hat{t}_{i-1} вирішують, перевіряючи лінійну гіпотезу (докладніше див. [3, п. 2.3]). Вилучення зайвих точок перемикачів сплайну виконується за результатами перевірки лінійної гіпотези (див. [3, розд. 3]).

2. МАШИНИЙ ЕКСПЕРИМЕНТ 1

Розглянемо сплайн, заданий на інтервалі часу довжиною $\Theta=62$, який має на цьому інтервалі три точки перемикачів: $t_1^0=15, t_2^0=33, t_3^0=45$. Параметри чотирьох прямих, що утворюють сплайн:

$$\alpha_1^0 = \begin{bmatrix} 5 \\ 0.1 \end{bmatrix}, \alpha_2^0 = \begin{bmatrix} 8.75 \\ -0.15 \end{bmatrix}, \alpha_3^0 = \begin{bmatrix} -2.95 \\ 0.2 \end{bmatrix}, \alpha_4^0 = \begin{bmatrix} 1.55 \\ 0.1 \end{bmatrix}.$$

До сплайну додано нормально розподілений некорельований у часі випадковий шум $\varepsilon_t \sim N(0, \sigma^2)$, $t=1, 2, \dots, \Theta$, $\sigma^2=0.16$, якому відповідає рівень 7% (відношення σ до середньої величини y_t на інтервалі спостереження надано у відсотках).

Оцінимо точки перемикачів сплайну, використовуючи критерії 1 та 2 (повний опис і обґрунтування наведено у [3, пп. 2.1, 2.2]). У табл. 1 наведено результати оцінювання t_1^0 . Нехай згідно з формулою (3) $i=0, \hat{t}_0=0, q_1=2, q_2=T-\hat{t}_0-1$. У цій і подальших таблицях жирним шрифтом позначено такі величини F -кри-

Таблиця 1. Результати оцінювання t_1^0

T	$F(\hat{t}_0+1, T+L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
3	0.016	19	99	2	2
4	1.204	9.552	30.817	2	2
5	0.291	6.944	18	2	2
6	1.193	5.786	13.274	2	2
7	1.603	5.143	10.925	2	2
8	1.415	4.737	9.547	1	2
9	0.214	4.459	8.649	0	2
10	0.852	4.257	8.022	0	1
11	1.727	4.103	7.559	1	1
12	0.025	3.982	7.206	2	2
13	1.826	3.885	6.927	2	2
14	4.418	3.806	6.701	1	1
15	5.815	3.739	6.515	0	0
16	3.477	3.682	6.359	1	1
17	3.148	3.633	6.226	1	1
18	3.444	3.592	6.112	0	0

терію, які більші, ніж табличні показники, жирний шрифт з підкресленням величини F -критерію показує, що відповідна точка є точкою перемикання. Жирним шрифтом також виокремлено нульові значення функції $\varphi_P(T)$ та знайдені оцінки точок перемикань.

Якщо вважати критерій 1 головним, а критерій 2 використовувати для контролю висновків, то згідно з табл. 1 на 5 %-му рівні відповідно до формули (3) приймається гіпотеза про те, що період часу $T = 15$ є точкою перемикання. Цей висновок підтверджується індикаторною функцією $\varphi_P(15) = 0$, що визначає точку перемикання, яка задовольняє умову (7), для $a = 0$, $P = 0.01, 0.05$. Тому вважатимемо $\hat{t}_1 = 15$.

Якщо навпаки: критерій 2 — головний, критерій 1 — допоміжний, то в жодній з точок $T = 9, 10$, в яких $\varphi_{0.05}(T) = 0$, не може бути перемикання, оскільки згідно з критерієм 1 не може бути відхилена гіпотеза про відсутність перемикань у зазначених точках. Згідно з табл. 1 тільки для точки $T = 15$, яка за умови $a = 0$ задовольняє формулу (7), відповідно до двох критеріїв приймається гіпотеза про існування у цій точці перемикання.

Перейдемо тепер до оцінювання t_2^0 . У цьому випадку в (3) вважатимемо $i = 1$, $\hat{t}_1 = 15$, $q_1 = 2$, $q_2 = T - \hat{t}_1 - 1$, $T \geq 16$. Результати розрахунку наведено у табл. 2.

За критерієм 1 згідно з табл. 2 для 5 %-го рівня значущості є інтервал $J_{0.05} = [33, 34, \dots, 39]$ значень T , для яких виконується умова (3). Для 1 %-го

Таблиця 2. Результати оцінювання t_2^0

T	Кількість спостережень після \hat{t}_1	$F(\hat{t}_i + 1, T + L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
18	3	2.923	19	99	2	2
19	4	0.073	9.552	30.816	2	2
20	5	0.003	6.944	18	2	2
21	6	0.132	5.786	13.274	1	1
22	7	4.349	5.143	10.925	0	0
23	8	2.251	4.737	9.547	1	1
24	9	0.451	4.459	8.649	2	2
25	10	0.129	4.256	8.021	2	2
26	11	1.778	4.103	7.559	2	2
27	12	0.456	3.982	7.206	1	1
28	13	2.839	3.885	6.927	1	1
29	14	0.306	3.805	6.701	2	2
30	15	0.010	3.738	6.515	2	2
31	16	0.846	3.682	6.359	2	2
32	17	0.337	3.634	6.226	2	2
33	18	5.865	3.592	6.112	2	2
34	19	6.898	3.555	6.013	1	1
35	20	14.090	3.522	5.926	0	0
36	21	8.731	3.493	5.849	0	0
37	22	12.412	3.467	5.780	0	0
38	23	12.865	3.443	5.719	0	0
39	24	16.093	3.422	5.664	0	0

рівня значущості має місце інтервал $J_{0.01} = [34, 35, \dots, 39]$, причому $J_{0.01} \subset J_{0.05}$. Перший максимум $F(\hat{t}_i + 1, T + L)$ на обох інтервалах відповідає $T = 35$. До того ж $\varphi_{0.05}(35) = \varphi_{0.01}(35) = 0$. Отже, за критерієм 1 маємо $\hat{t}_2 = 35$.

З табл. 2 випливає, що за критерієм 2 приймається гіпотеза про наявність у точці $T = 22$ перемикавання, а за критерієм 1 не відхиляється гіпотеза про відсутність у цій точці перемикавання (всі міркування з урахуванням (7) для $a = 0$ на 5 %-му і 1 %-му рівнях). Отже, дійшли висновку, що у точці $T = 22$ перемикавання немає.

За критерієм 2 приймається гіпотеза на 5 %-му і 1 %-му рівнях про наявність у точці $T = 35$ перемикавання, що не суперечить такому висновку за критерієм 1, оскільки для $T = 35$ виконується нерівність (3) для зазначених вище рівнів значущості.

Отже, за обома критеріями на 5 %-му і 1 %-му рівнях приймається гіпотеза про існування перемикавання в точці $T = 35$. Точки, які можуть мати перемикавання за періоди часу $T > 35$, у табл. 2 не розглянуто, оскільки вони належать наступній прямій (табл. 3).

Для оцінювання третьої точки перемикавання задамо у (3), $i = 2$, $\hat{t}_2 = 35$, $q_1 = 2$, $q_2 = T - \hat{t}_2 - 1$, де $T \geq 36$. Результати розрахунку наведено у табл. 3.

За результатами спостережень, наведеними у табл. 3, виявлено, що на 1 %-му рівні за критерієм 2 відсутні точки перемикавань. Вони з'являються, починаючи з рівня 2 %.

Згідно з критерієм 1 у табл. 3 наявні дві точки: $T = 49$ і $T = 51$, в яких можливе перемикавання. На 5 %-му і 2 %-му рівнях значущості приймається гіпотеза про те, що $\hat{t}_3 = 49$, водночас в (7) $a = 0$. У цьому випадку критерій 2 не суперечить такому висновку. Тоді точку $T = 51 > 49$ не розглядають у табл. 3, вона потрапляє до наступної прямої, що утворює тренд.

За критерієм 2 на 5 %-му рівні маємо три точки: $T = 39, 42, 49$ — кандидати на перемикавання. На 2 %-му рівні таких точок дві: $T = 39, 49$. Крім точки $T = 49$,

Таблиця 3. Результати оцінювання t_3^0

T	Кількість спостережень після \hat{t}_2	$F(\hat{t}_2 + 1, T + L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
38	3	1.24	19	99	1	2
39	4	6.711	9.552	30.817	0	0
40	5	0.944	6.944	18	1	2
41	6	1.075	5.786	13.274	1	2
42	7	1.262	5.143	10.925	0	2
43	8	2.385	4.737	9.547	2	2
44	9	1.995	4.459	8.649	1	2
45	10	3.334	4.256	8.022	2	2
46	11	0.969	4.103	7.559	1	1
47	12	3.545	3.982	7.206	1	1
48	13	2.454	3.885	6.927	1	1
49	14	7.123	3.806	6.701	0	0
50	15	2.85	3.739	6.515	1	1
51	16	4.286	3.682	6.359	1	1
52	17	1.381	3.634	6.226	1	1

Таблиця 4. Результати оцінювання t_4^0

T	Кількість спостережень після \hat{t}_3	$F(\hat{t}_3 + 1, T + L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
52	3	0.072	19	99	2	2
53	4	2.83	9.552	30.817	2	2
54	5	0.936	6.944	18	2	2
55	6	1.036	5.786	13.274	1	2
56	7	0.137	5.143	10.925	1	2
57	8	0.258	4.737	9.547	0	2
58	9	0.623	4.459	8.649	1	1
59	10	0.852	4.256	8.022	2	2

решта точок для $a=0$ в (7) суперечать критерію 1. Отже, за обома критеріями справджується оцінка $\hat{t}_3 = 49$ для 5 %-го и 2 %-го рівнів.

Розглянемо тепер інтервал часу $T \geq 50$, що відповідає $i=3$, $\hat{t}_3 = 49$, $q_1 = 2$, $q_2 = T - \hat{t}_3 - 1$. Він представлений у табл. 4, де наведено обидва критерії для $T \in [50, 51, \dots, 59]$. Для останніх в інтервалі спостереження точок $T = 60, 61, 62$ ці критерії не обчислювали, оскільки не визначено рівнів часового ряду для $T > 62$.

Згідно з табл. 4 немає жодної точки, в якій можливе перемикання за критерієм 1. Те саме можна сказати про критерій 2 на 1 %-му рівні значущості. На 5 %-му рівні значущості критерій 2 вказує на точку $T = 57$, як кандидата на перемикання. Але цю властивість точки не підтверджує критерій 1 на обох рівнях значущості. Отже, за обома критеріями не знайдено на 5 %-му і 1 %-му рівнях жодної точки, в якій могло бути перемикання.

Отже, можна виснувати, що на 5 %-му і 1 %-му рівнях значущості на інтервалі спостереження є три точки перемикань: $\hat{t}_1 = 15$, $\hat{t}_2 = 35$, $\hat{t}_3 = 49$ (дві останні точки відповідають також 1 %-му рівню). Ці величини задовольняють обмеження (8), тому вважаємо, що вони демонструють зміни тренду, а не є випадковими відхиленнями. Значення оцінок точок перемикання достатньо близькі до істинних значень — відносна похибка менша 10 %.

Тепер побудуємо сплайн за алгоритмом з розд. 3 [3]. Результати наведено у табл. 5, а сплайн — на рис. 1.

Згідно з табл. 5 (рядки 1 і 2) мінімізація за параметрами тренду з істинними точками перемикань уможливило зменшення S . Схожий результат отриманий для лінійної регресії: сума квадратів залишків, що відповідає оцінкам її параметрів, менша за таку саму суму для істинних параметрів. Цей факт засвідчує порівняння S для істинної моделі та моделі у другому рядку з моделлю, отриманою мінімізацією S за точками перемикань та параметрів тренду (рядок 7).

Зазначимо, що оцінки точок перемикань за критеріями 1 та 2 помітно покращуються після з'єднання їх сплайном. Порівняємо точки (15, 35, 49) (рядок 4) і точки (14, 34, 43) (рядок 7) з істинними точками перемикань (15, 33, 45). Причому сума квадратів залишків, що відповідає сплайну з точками перемикань (14, 34, 43), менша, ніж така сама сума для істинних точок перемикань (рядки 1, 2), що має місце для хорошої моделі оцінювання.

Таблиця 5. Результати побудови сплайну з трьома точками перемикань

Номер рядка	Модель	\hat{t}_1	\hat{t}_2	\hat{t}_3	Сума квадратів залишків S	Коментар
1	Істинна	15	33	45	7.570	—
2	Сплайн	15	33	45	6.236	—
3	Можливі розриви тренду у точках перемикань	15	35	49	5.395	—
4	Сплайн	15	35	49	6.664	—
5	(26) з [3]	14	35	49	5.577	Використано обмеження, що гарантує неперервність тренду в першій точці перемикання
6	(28) з [3] для $i = 2$	14	34	49	5.840	Використано два обмеження, що гарантують неперервність тренду в першій та другій точках перемикань
7	(28) з [3] для $i = 3$	14	34	43	5.754	Використано три обмеження, що гарантують неперервність тренду в першій, другій і третій точках перемикань

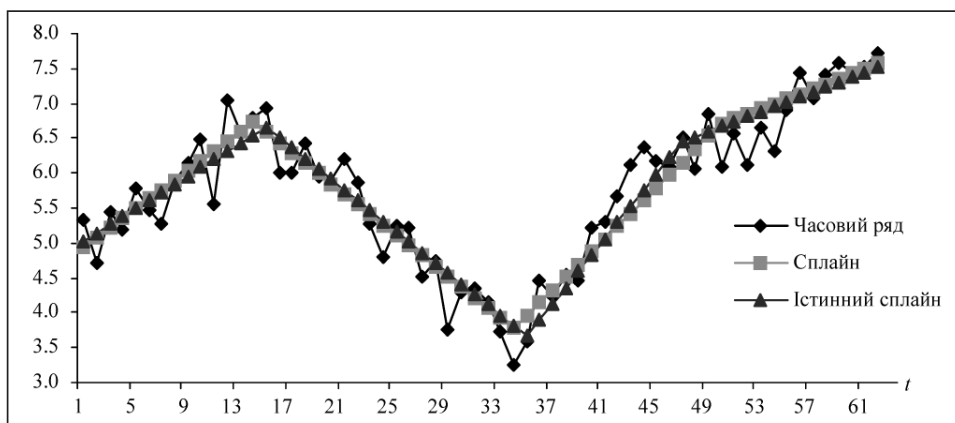


Рис. 1. Сплайн, побудований у машинному експерименті 1

Різниця між сплайнами істинної (рядок 1) та оціненої (рядок 7) моделей має середньоквадратичну похибку 0.108, середнє істинного сплайну 5.74; звідси маємо, що коефіцієнт варіації різниці дорівнює 1.87 %. Отже, тренд визначено з високою точністю.

3. МАШИНИЙ ЕКСПЕРИМЕНТ 2

Постає питання: чи можна методами виявлення точок перемикань побудувати лінійну регресію з перемиканнями, в якій незалежні змінні є функціями часу. З цією метою розглянемо приклад з роботи [5]. Він полягає в побудові регресії з перемиканнями $y_t = \mathbf{x}'_t \boldsymbol{\alpha}^0 + \mathbf{z}'_t \boldsymbol{\beta}^0 + \varepsilon_t$, $i = 1, \dots, k^0 + 1$, де $k^0 = 5$ — істинна кількість точок перемикань, $t = 1, 2, \dots, 58$, для випадку, коли $\mathbf{x}_t = [x_{t1} \ x_{t2}]'$, $\sigma = 0.5$, що відповідає достатньо високому рівню шуму, а саме 11.73 %. Кожна незалежна змінна — компонента \mathbf{x}_t , являє собою суму лінійного тренду з нормально розподіленим шумом. Ці змінні сильнокорельовані —

Таблиця 6. Результати аналізу інтервалу [1, 30]

T	Кількість спостережень після попередньої точки перемикання \hat{t}_*	$F(\hat{t}_* + 1, T + L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
6	6 (відлік з $T=1$)	365.128	5.786	13.274	0	0
20	14 (відлік з $T=7$)	12.692	3.522	5.926	1	2
21	15	15.639	4.463	4.463	1	1
22	16	108.951	3.467	5.78	0	0
23	17	16.242	3.443	5.719	0	0
25	3 (відлік з $T=23$)	45.932	3.403	5.614	1	2
26	4	8.584	3.385	5.568	0	0
27	5	9.564	3.369	5.526	0	0
30	3 (відлік з $T=28$)	10.320	3.328	5.420	0	0

коефіцієнт кореляції дорівнює 0.996, що свідчить про сильну мультиколінеарність. Регресія має розриви в точках перемикань: $t_1^0 = 6$, $t_2^0 = 22$, $t_3^0 = 34$, $t_4^0 = 38$, $t_5^0 = 44$.

Побудуємо регресію за часовими рядами змінних, використовуючи метод з роботи [3, розд. 2], а саме критерій 1 та 2, а також вилучаючи зайві точки. Як головний виберемо критерій 1, а як допоміжний — критерій 2. У табл. 6 наведено результати аналізу інтервалу [1, 30] — першої частини інтервалу спостереження. Тут розглянуто лише точки, які задовольняють критерію 1, а також відповідні величини критерію 2.

Згідно з табл. 6 маємо $\hat{t}_1 = 6$, $\hat{t}_2 = 22$, $\hat{t}_3 = 27$. Оскільки $\hat{t}_3 - \hat{t}_2 = 5 < 6$, то умова (8) не виконується. Тому відповідно до п. 2.3 з [3] розглянемо трійку точок: \hat{t}_{i-2} , \hat{t}_{i-1} , \hat{t}_i , де $i=3$, і перевіримо гіпотезу про наявність перемикання в точці $t=22$. Маємо $F(\hat{t}_{i-2} + 1, \hat{t}_i) = F(\hat{t}_1 + 1, \hat{t}_3) = F(7, 27) = 93.733$, що набагато більше за $F_{0.01}(q_1, q_2) = 6.226$, де $q_1 = 2$, $q_2 = \hat{t}_i - \hat{t}_{i-2} - 4 = 16$. Отже, для 1 %-го рівня немає підстав вважати, що в точці $T=22$ відсутнє перемикання.

Тому шукатимемо наступну точку перемикання, починаючи з $\hat{t}_3 = 27$. Такою точкою згідно з табл. 6 буде $\hat{t}_4 = 30$, причому $\hat{t}_4 - \hat{t}_3 = 3 < 6$ — умова (8) не виконується. Знову скористаємося результатами роботи [3, п. 2.3] і розглянемо трійку точок: \hat{t}_{i-2} , \hat{t}_{i-1} , \hat{t}_i , де $i=4$, перевіримо гіпотезу про відсутність перемикання у точці $T=27$. Маємо $F(\hat{t}_{i-2} + 1, \hat{t}_i) = F(\hat{t}_2 + 1, \hat{t}_4) = F(23, 30) = 4.773$, $F_{0.01}(q_1, q_2) = 18$, де $q_1 = 2$, $q_2 = \hat{t}_i - \hat{t}_{i-2} - 4 = 4$. Отже, на 1 %-му рівні гіпотеза про відсутність в точці $T=27$ перемикання не відхиляється. Це означає, що $\hat{t}_2 = 22$ — остання знайдена точка перемикання і потрібно рухатися далі, починаючи з $T=28$, збільшуючи значення T , для знаходження \hat{t}_3 . Отримані результати наведено в табл. 7.

Згідно з табл. 7 маємо $\hat{t}_3 = 34$. Ця величина не потребує уточнення, оскільки $\hat{t}_3 - \hat{t}_2 = 12 > 6$. Інші точки перемикання визначали за даними табл. 8, де наведено періоди часу, для яких виконувався критерій 1.

Таблиця 7. Результати визначення \hat{t}_3

T	Кількість спостережень після \hat{t}_2	$F(\hat{t}_2 + 1, T + L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
28	6	0.098	3.354	5.488	2	2
29	7	0.318	3.34	5.453	2	2
30	8	0.289	3.328	5.42	2	2
31	9	0.167	3.316	5.39	2	2
32	10	7.787	3.305	5.362	2	2
33	11	5.071	3.295	5.336	1	1
34	12	65.220	3.285	5.312	0	0
35	13	19.128	3.276	5.289	0	0

Таблиця 8. Результати аналізу інтервалу [35, 58]

T	Кількість спостережень після попередньої точки перемикання \hat{t}_*	$F(\hat{t}_* + 1, T + L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
37	3 (відлік з $T=35$)	5.062	3.259	5.248	1	1
38	4	223.572	3.252	5.229	0	0
42	4 (відлік з $T=39$)	70.735	3.226	5.163	1	2
43	5	0.875	3.22	5.149	1	1
44	6	38.849	3.214	5.136	0	0
45	7	3.757	3.209	5.123	0	0
47	3 (відлік з $T=45$)	11.05	3.200	5.099	1	1

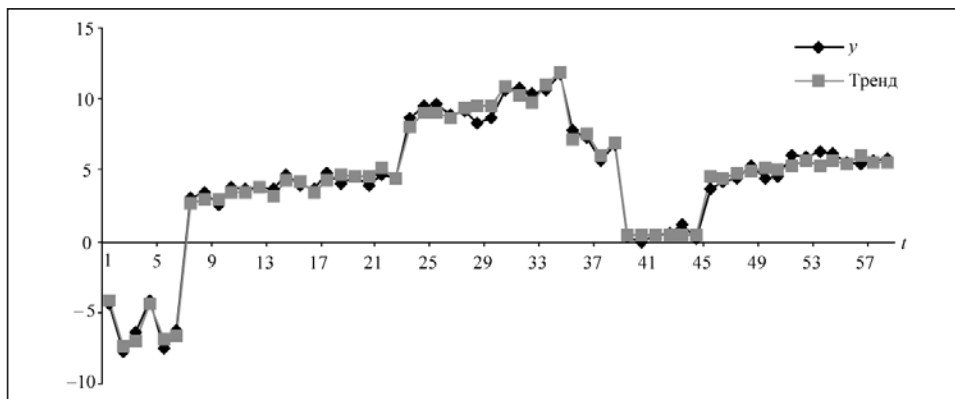


Рис. 2. Часовий ряд залежної змінної і лінійна регресія, що її апроксимує (машинний експеримент 2)

З перших двох рядків табл. 8 випливає, що $\hat{t}_4 = 38$, але $\hat{t}_4 - \hat{t}_3 = 4 < 6$. Однак уточнювати $\hat{t}_3 = 34$ через це не потрібно, бо в точці $T = 34$ маємо стрибок залежної змінної, який зумовив велике значення F -критерію під час перевірки гіпотези про відсутність перемикання у зазначеній точці. З наступних чотирьох рядків $42 \leq T \leq 45$ отримаємо $\hat{t}_5 = 44$. Для точок перемикань $\hat{t}_4 = 38$ і $\hat{t}_5 = 44$ висновки отримано на 1 %-му рівні значущості. Оскільки $\hat{t}_5 - \hat{t}_4 = 6$, корегувати ці точки перемикань немає потреби.

Отже, для $a=0$ в (7) отримано оцінки п'ятьох точок перемикань: $\hat{t}_1 = 6$, $\hat{t}_2 = 22$, $\hat{t}_3 = 34$, $\hat{t}_4 = 38$, $\hat{t}_5 = 44$, які збігаються з істинними величинами. Такий збіг можна пояснити помітними стрибками залежної змінної в перших чотирьох точках перемикань. Методами з [5, 6], які застосовувалися для розв'язання цієї задачі, отримано менш точні результати. На рис. 2 наведено лінійну регресію та часовий ряд залежної змінної, який вона апроксимує.

4. ЗГЛАДЖУВАННЯ РЯДУ ЛОГАРИФМІВ ЧИСЛА ІНФІКОВАНИХ COVID-19 В УКРАЇНІ В 2020 Р., З ЯКОГО ВИЛУЧЕНІ ТИЖНЕВІ КОЛИВАННЯ

Природу тижневих коливань кількості інфікованих і засіб їхнього видалення розглянуто у [7]. У роботі [8] описано модель епідемії COVID-19 в дискретному часі. Модель має змістовний сенс, оскільки вона враховує процеси інфікування, лікування тощо. У роботі [8] показано, що можна підвищити точність ідентифікування, використовуючи як вихідні дані не кількість інфікованих за добу, а часовий ряд логарифмів чисел, що позначають цю кількість.

Такий ряд, побудований у великому масштабі, візуалізує тренд у вигляді кусково-лінійної функції (рис. 3). Ряд охоплює період з 12.04.2020 до 06.08.2021, що відповідає нумерації періодів часу (доба) від 1 до 482. Розглянемо початкову частину цього ряду з 12.04.2020 до 20.08.2020 (час T змінюється від 1 до 131). У цій частині ряду амплітуди коливань рівнів ряду значно менші, що дає змогу перевірити можливість ідентифікації невеликих сплесків методами, описаними у [3]. На рис. 3 візуалізовано аномальну точку, що відповідає періоду часу $T = 51$, рівень ряду складає 6.880, який було скореговано до рівня 6.2.

У табл. 9 наведено результати аналізу розглянутого початкового інтервалу, що характеризується (особливо для малих T) значною кількістю аномальних спостережень, які не такі великі у порівнянні з точкою $T = 51$. Таку появу можна пояснити становленням систем тестування на COVID-19 та обліку інфікованих. Тому на відміну від задач, розв'язаних в [3, розд. 2, 3], де у формулі (7) $a=0$, тут ми допускаємо, що в сумнівних випадках у цій формулі $a=1$. Це означає, що головний критерій і допоміжний критерій можуть належати сусіднім періодам часу, а не тому самому періоду, як в розд. 2, 3.

У табл. 9 наведено точки, де, можливо (за критерієм 1 або за критерієм 2) є перемикання (головним є критерій 1).

Зазначимо, що тільки у перших двох рядках табл. 9 маємо ситуацію, коли в першому рядку міститься головний критерій, а в другому — допоміжний критерій 2. Інакше кажучи, на інтервалі [1, 131] один раз умова (7) виконувалася

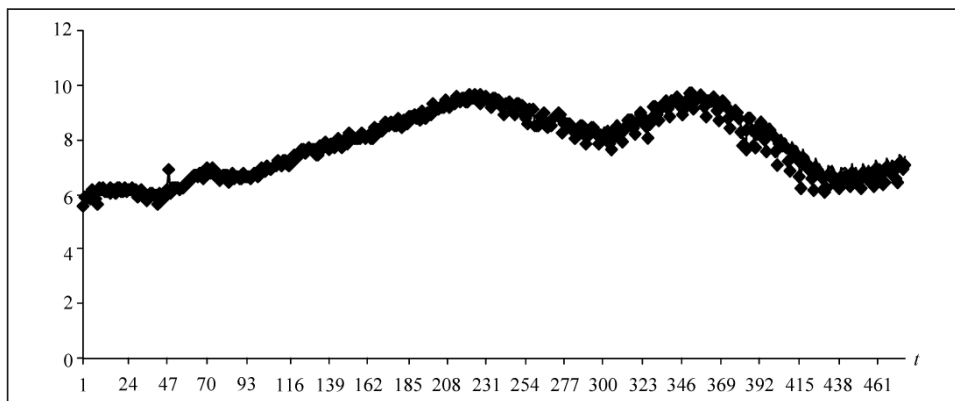


Рис. 3. Згладжений часовий ряд логарифмів числа інфікованих в Україні без урахування тижневих коливань з 12.04.2020 до 06.08.2021

Таблиця 9. Результати аналізу згладженого часового ряду логарифмів числа інфікованих в Україні (без урахування тижневих коливань) з 12.04.2020 до 20.08.2020

Період часу T	Кількість спостережень після попередньої точки перемикання \hat{t}_*	$F(\hat{t}_* + 1, T + L)$	$F_{0.05}(q_1, q_2)$	$F_{0.01}(q_1, q_2)$	$\varphi_{0.05}(T)$	$\varphi_{0.01}(T)$
6	6 (відлік з $T = 1$)	7.895	5.786	13.274	1	1
7	7	2.464	5.143	10.925	0	0
35	29 (відлік з $T = 7$)	5.624	3.34	5.453	1	1
36	30	6.084	3.328	5.420	0	0
37	31	3.915	3.316	5.390	0	1
78	43 (відлік з $T = 37$)	8.785	3.214	5.163	0	1
103	23 (відлік з $T = 79$)	5.856	3.403	5.614	0	0
118	13 (відлік з $T = 104$)	4.244	3.739	6.515	1	2
119	14	2.672	3.634	6.359	0	1
121	3 (відлік з $T = 119$)	34.780	19	99	0	0

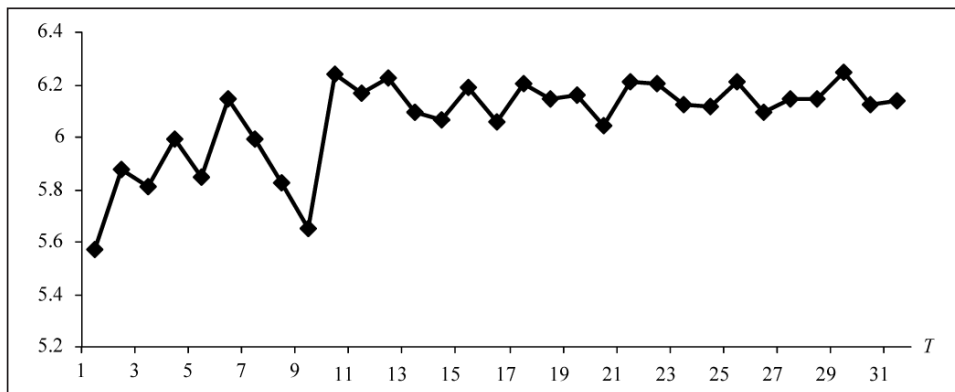


Рис. 4. Початок згладженого часового ряду логарифмів числа інфікованих в Україні без урахування тижневих коливань

для $a = 1$, в інших випадках ця умова виконувалася для $a = 0$. На рис. 4 показано скупчення аномальних точок на початку інтервалу $[1, 131]$.

Отже, для рядків $T = 6$ і $T = 7$, для яких в (7) $a = 1$, маємо $\hat{t}_1 = 6$ на 5 %-му рівні значущості. Для рядків $T = 35$, $T = 36$, $T = 37$ отримуємо $\hat{t}_2 = 36$ на 1 %-му рівні значущості. Для рядка $T = 78$ маємо $\hat{t}_3 = 78$ також на 1 %-му рівні значущості. Критерій 2 підтверджує коректність цього висновку на 5 %-му рівні. Відповідно для рядка $T = 103$ отримуємо $\hat{t}_4 = 103$ на 1 %-му рівні.

Розглянемо тепер рядки $T = 118$, $T = 119$. Вважатимемо у (7) $a = 1$, тоді за принципом визначення максимально можливої кількості точок перемикання отримуємо на 5 %-му рівні значущості $\hat{t}_5 = 118$.

Для останнього рядка табл. 9 на 5 %-му рівні $\hat{t}_6 = 121$. Тоді $\hat{t}_6 - \hat{t}_5 = 3 < 6$, що означає невиконання умови (8) і може призвести не до відстеження тренду, а відтворення випадкових збурень. Для дослідження цього процесу скористаємося методом з роботи [3, п. 2.3].

Розглядаючи трійку точок \hat{t}_4 , \hat{t}_5 , \hat{t}_6 , перевіримо гіпотезу про те, що \hat{t}_5 — помилкова або зайва точка перемикання в регресії, яка має не більше однієї точ-

ки перемикання та задана на інтервалі $[\hat{t}_4 + 1, \hat{t}_6]$ з фіксованими кінцями. Відповідно до [3, п. 2.3] маємо величину F -критерію $F(\hat{t}_{i-2} + 1, \hat{t}_i) = F(104, 121) = 4.195$, де $i = 6$. Нульова гіпотеза H_0 про відсутність перемикання в точці $T = 118$ не відхиляється на 1 %-му рівні, оскільки $F(104, 121) < F_{0.01}(q_1, q_2) = 6.515$, де $q_1 = 2$, $q_2 = \hat{t}_i - \hat{t}_{i-2} - 4 = 121 - 103 - 4 = 14$. На підставі цього можна дійти висновку, що введення точки $T = 118$ як кандидата на перемикання застосуванням $a = 1$ у (7) було перестрахованням.

Отже, маємо останню оцінену точку перемикання $\hat{t}_4 = 103$ для $T \leq 118$. Далі рухаємося від точки $T = 119$ до кінця інтервалу $[1, 131]$. При цьому жодної точки — кандидата на перемикання, у тому числі й $T = 121$, не було виявлено. Отже, отримано оцінки точок перемикання: $\hat{t}_1 = 6$, $\hat{t}_2 = 36$, $\hat{t}_3 = 78$, $\hat{t}_4 = 103$.

На рис. 5 за результатами розв'язання задачі (21) у [3] представлено тренд у вигляді кусково-лінійної функції з розривами в точках перемикання. Аномальну точку, зображену на цьому рисунку та рис. 6, 7, у розрахунках враховували у відкорегованому вигляді.

Процес інфікування є неперервним у часі, тому побудуємо його тренд у вигляді сплайну. Результати розрахунків представлено у табл. 10 (аналогу табл. 5), відповідний графік — на рис. 6. Згідно з цим рисунком на початку епідемії в Україні зафіксовано два осередки інфекції (два сплески на графіку), зумовлені, мабуть, проведенням заходів з великою кількістю людей.

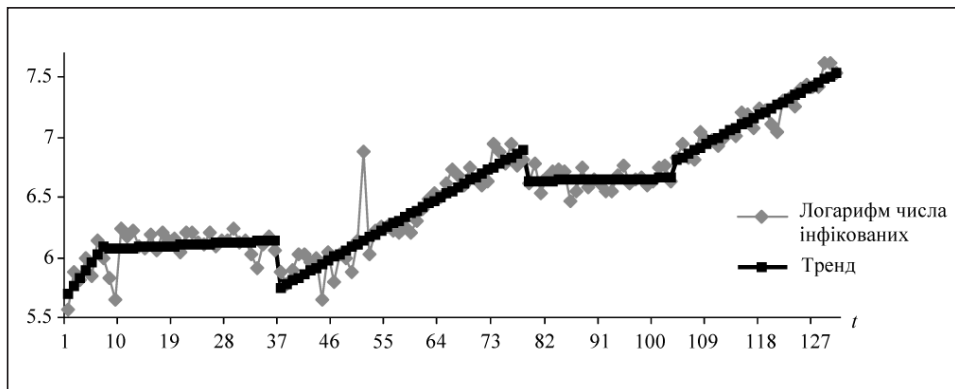


Рис. 5. Логарифм числа інфікованих в Україні без тижневих коливань та його тренд у вигляді кусково-лінійної функції (період з 12.04.2020 до 20.08.2020)

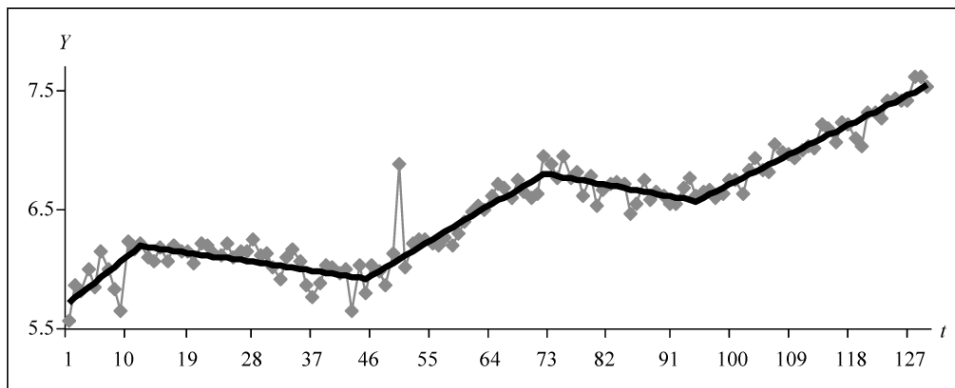


Рис. 6. Логарифм числа інфікованих в Україні без тижневих коливань та його тренд у вигляді сплайну (період з 12.04.2020 до 20.08.2020)

Таблиця 10. Результати розрахунків для побудови сплайну з чотирма точками перемикачів

Номер рядка	Модель	\hat{t}_1	\hat{t}_2	\hat{t}_3	\hat{t}_4	Сума квадратів залишків S	Коментар
1	Тренд із розривами (рис. 7)	6	35	78	103	1.378	—
2	(26) з [3]	12	46	78	103	1.315	Використано обмеження, що гарантує неперервність тренду в першій точці перемикачів
3	(28) з [3] для $i = 2$	12	46	78	103	1.316	Використано два обмеження, що гарантують неперервність тренду в першій та другій точках перемикачів
4	(28) з [3] для $i = 3$	12	46	73	103	1.393	Використано три обмеження, що гарантують неперервність тренду в трьох точках перемикачів
5	(28) з [3] для $i = 4$	12	46	73	96	1.346	Використано чотири обмеження, що гарантують неперервність тренду в чотирьох точках перемикачів

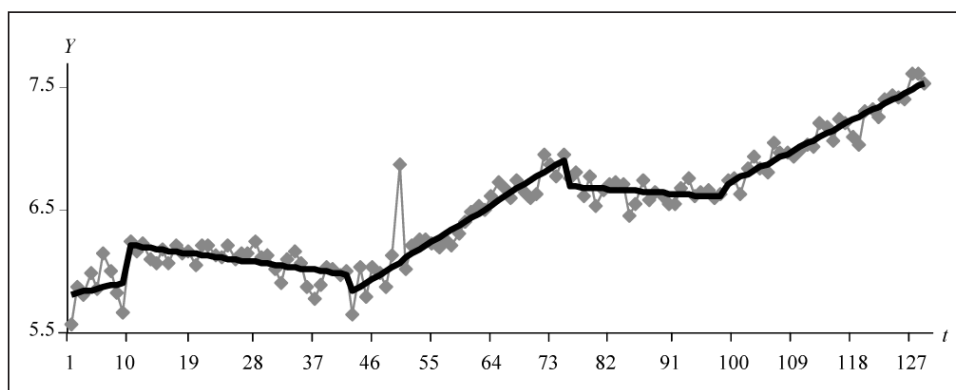


Рис. 7. Логарифм числа інфікованих в Україні без тижневих коливань та його тренд у вигляді покращеної кусково-лінійної функції (період з 12.04.2020 до 20.08.2020)

Згідно з табл. 10 уведення обмежень на параметри регресії, щоб отримати сплайн, не збільшило значення S , а, навпаки, незначно зменшило суму квадратів залишків. Це пояснюється тим, що варіювалися як параметри регресії, так й точки перемикачів.

Другий висновок з табл. 10: тренд з розривами не є оптимальним. На скінченному відрізку часу точки перемикачів, у яких є розриви, можна покращити методом, який використано для побудови тренду у вигляді сплайну згідно з табл. 10, але без вимоги неперервності тренду у точках перемикачів. На рис. 7 наведено логарифм числа інфікованих в Україні без тижневих коливань та його тренд у вигляді покращеної кусково-лінійної функції. Він відповідає оцінкам точок перемикачів: $\hat{t}_1 = 9$, $\hat{t}_2 = 43$, $\hat{t}_3 = 76$, $\hat{t}_4 = 100$. Сума квадратів залишків становить $S = 1.17$. Порівняно з рис. 5 на рис. 7 розриви в точках перемикачів, крім першого, зменшилися, що зумовило зменшення значення S .

Повернімося тепер до сплайну на рис. 6. Природно виникає питання, чи значущі максимуми цього сплайну? Можливо, вони є результатом випадкових змін часового ряду? Щоб відповісти на це питання, скористуємось наслідком 1 з роботи [3], вважаючи оцінені точки перемикаць сплайну фіксованими. Таке спрощення є некритичним унаслідок великих відстаней між цими точками, що зумовлює значну надмірність даних для оцінювання параметрів сплайну.

Перший максимум — у точці $t=12$. Якщо в цій точці немає перемикаць, тоді не відхиляють гіпотезу:

$$H_0: \alpha_{11}^0 = \alpha_{12}^0, \quad (9)$$

де $\alpha_{11}^0, \alpha_{12}^0$ — тангенси кутів нахилу першої та другої прямих сплайну на рис. 6, що стикаються в точці $t=12$.

Для перевірки гіпотези (9) на інтервалі $I = [1, 131]$ розглянемо сплайн Ω з чотирма точками перемикаць, наведений у п'ятому рядку табл. 10. Сплайн $\bar{\Omega}$ з трьома точками перемикаць отримуємо з Ω за умов виконання (9), що є окремим випадком рівності (31) у [3].

Щоб перевірити гіпотезу (9), скористаємося нерівністю (32) з [3], де у розглядуваному випадку \mathbf{a}^- — вектор параметрів сплайну Ω , а \mathbf{a}^- — сплайну $\bar{\Omega}$. Маємо суми квадратів залишків: $\|\mathbf{Y} - \mathbf{X}\mathbf{a}^-\|^2 = 1.965$, $\|\mathbf{Y} - \mathbf{X}\mathbf{a}^-\|^2 = 1.346$, $k = m = 4$, що дає $N = 2(k + 1) = 10$, $v = 1$, $q_1 = v = 1$, $q_2 = \Theta - N + m = 125$. Звідси згідно з (32) у [3] отримуємо $F = 57.476$, $F_{0.01}(q_1, q_2) = 6.842$. Отже, на 1 %-му рівні гіпотеза H_0 відхиляється і приймається гіпотеза $H_1: \alpha_{11}^0 \neq \alpha_{12}^0$, що свідчить про наявність точки перемикаць в $t=12$. Цей висновок підтверджує правильність перестраховання введенням точки $\hat{t}_1 = 6$ у тренд з розривами, з якого отримано сплайн Ω .

Перевіримо тепер гіпотезу про наявність перемикаць у точці $t = 73$. Для цього розглянемо на інтервалі $I = [1, 131]$ сплайн $\bar{\Omega}$ з точками перемикаць $\hat{t}_1 = 12$, $\hat{t}_2 = 46$ і $\hat{t}_4 = 96$, вважаючи їх фіксованими. У цьому випадку висуваємо гіпотезу

$$H_0: \alpha_{12}^0 = \alpha_{13}^0. \quad (10)$$

Їй відповідають суми квадратів залишків $\|\mathbf{Y} - \mathbf{X}\mathbf{a}^-\|^2 = 4.026$, $\|\mathbf{Y} - \mathbf{X}\mathbf{a}^-\|^2 = 1.346$, величини N , v , q_1 , q_2 , $F_{0.01}(q_1, q_2)$ такі самі, як у попередньому випадку. Для наведених даних відповідно до (32) у [3] маємо $F = 248.839$. Як і під час перевірки попередньої гіпотези, гіпотеза (10) відхиляється на 1 %-му рівні з ще більшим відхиленням F від $F_{0.01}(q_1, q_2) = 6.842$. Отже, приймаються гіпотези про існування двох максимумів на рис. 6, які не випадкові.

Довжина прямих у сплайні в рядку 5 табл. 10 значно (особливо для прямих, окрім першої) більша, ніж два — кількість параметрів прямої. Якщо взяти до уваги той факт, що F значно перевищує табличну величину $F_{0.01}(q_1, q_2)$, то можна стверджувати, що неточність під час обчислення оцінок $\hat{t}_1 = 12$, $\hat{t}_2 = 46$ і $\hat{t}_4 = 96$ не є критичною для висновку про невідповідність максимумів на рис. 6.

ВИСНОВКИ

У роботі [3] запропоновано та обґрунтовано метод визначення тренду нестационарного часового ряду — поширена прикладна задача з різних галузей науки і має немалий практичний інтерес. У цій статті описано апробацію методу: проведено два машинних експерименти та розв'язано складну практичну задачу. Випробувано набір засобів оцінювання точок перемикаць і параметрів тренду, розглянутих у [3], як у разі побудови тренду з розривами в точках

перемикач, так і без них (останнє відповідає сплайну). Отримані результати є задовільними. Крім того, з машинного експерименту можна дійти висновку про можливість застосування методу для побудови лінійної регресії з перемикачними за однією змінною — часом. Подальші дослідження з цієї тематики можуть бути застосовані і для більш широких класів прикладних задач, а також для складніших задач, наведених у [9, 10].

СПИСОК ЛІТЕРАТУРИ

1. Knopov P.S., Korkhin A.S. Regression analysis under a priori parameter restrictions. Springer Optimization and Its Applications. Vol 54, New York, NY: Springer, 2011. <https://doi.org/10.1007/978-1-4614-0574-0>.
2. Korkhin A.S. Parameter estimation accuracy for nonlinear regression with nonlinear constraints. *Cybernetics and Systems Analysis*. 1998. Vol. 34, N 5. P. 663–672. <https://doi.org/10.1007/BF02667038>.
3. Кнопов П.С., Корхін А.С. Визначення кусково-лінійного тренду нестационарного часового ряду на основі інтелектуального аналізу даних. I. Опис та обґрунтування методу. *Кибернетика та системний аналіз*. 2024. Т. 60, № 1. С. 61–72.
4. Perron P., Zorta E. Estimation and inference of linear trend slope ratios with an application to global temperature data. *Journal of Time Series Analysis*. 2017. Vol. 38, N 5. P. 630–667.
5. Korkhin A.S. Constructing a switching regression with unknown switching points. *Cybernetics and Systems Analysis*. 2018. Vol. 54, N 3. P. 443–455. <https://doi.org/10.1007/s10559-018-0045-9>.
6. Korkhin A.S. An approximate method of constructing a switching regression with unknown switch points. *Cybernetics and Systems Analysis*. 2020. Vol. 56, N 3. P. 426–438. <https://doi.org/10.1007/s10559-020-00258-1>.
7. Knopov P.S., Korkhin A.S. Statistical analysis of the dynamics of coronavirus cases using stepwise switching regression. *Cybernetics and Systems Analysis*. 2020. Vol. 56, N 6. P. 943–952. <https://doi.org/10.1007/s10559-020-00314-w>.
8. Knopov P.S., Korkhin A.S. Dynamic models of epidemiology in discrete time taking into account processes with lag. *Int. J. Dynam. Control*. 2023. Vol. 11. P. 2193–2214. <https://doi.org/10.1007/s40435-023-01135-3>.
9. Golodnikov A.N., Knopov P.S., Pepelyaev V.A. Estimation of reliability parameters under incomplete primary information. *Theory and Decision*. 2004. Vol. 57, N 4. P. 331–344. <https://doi.org/10.1007/s11238-005-3217-9>.
10. Norkin V.I., Gaivoronski A.A., Zaslavsky V.A., Knopov P.S. Models of the optimal resource allocation for the critical infrastructure protection. *Cybernetics and Systems Analysis*. 2018. Vol. 54, N 5. P. 696–706. <https://doi.org/10.1007/s10559-018-0071-7>.

P.S Knopov, A.S. Korkhin

DETERMINING A PIECEWISE-LINEAR TREND OF A NON-STATIONARY TIME SERIES BASED ON INTELLIGENT DATA ANALYSIS.

II. MACHINE EXPERIMENTS AND PRACTICAL PROBLEM SOLUTIONS

Abstract. The article describes the results of the approbation of the method of constructing a piecewise-linear trend, which can have breaks at the switching points as well as be continuous at these points, i.e., represent a linear flow. An example of applying the method for constructing a linear regression with switches, which has two independent variables with a trend, is considered. The problem of spline approximation of the time series of logarithms of the number of infected people with COVID-19 in Ukraine is stated and solved.

Keywords: trend, regression, switch point, spline, real-time calculation.

Надійшла до редакції 11.08.2023