

ДИСКРЕТНЕ ФУР'Є-ПРОДОВЖЕННЯ ЯК АЛГОРИТМ ПРОГНОЗУВАННЯ ФІНАНСОВО-ЕКОНОМІЧНИХ ЧАСОВИХ РЯДІВ

Д.М. ЧАБАНЕНКО

Запропоновано методика прогнозування часових рядів на основі виявлення частотного спектра. Методика полягає в апроксимації часового ряду на проміжку навчання аналітичною функцією, яка містить тренд та комбінацію гармонічних коливань. Для оцінювання параметрів моделі використовуються нелінійні методи оптимізації. За допомогою експериментальної апробації показано ефективність запропонованого методу для прогнозування рядів фінансово-економічної динаміки.

ВСТУП

Прогнозування часових рядів економічної динаміки є надзвичайно актуальною задачею в дослідженні складних фінансово-економічних систем. Специфіка фінансових часових рядів полягає в наявності стилізованих фактів [1]: негаусівський розподіл прибутковостей, кластеризація волатильності, флікер-шум ($1/f$ — шум) та ін., що вказує на складність структури системи, що досліджується. Такі часові ряди не завжди адекватно відтворюються сучасними методами прогнозування. Тому питання розробки принципово нових підходів до вирішення задачі прогнозування фінансово-економічних часових рядів є актуальними.

Мета роботи — побудова рядів фінансово-економічної динаміки прогнозної моделі на основі нелінійної апроксимації часового ряду сумою гармонічних функцій.

ПОСТАНОВКА ЗАДАЧІ: ЗАВДАННЯ ДОСЛІДЖЕННЯ

Задача прогнозування часових рядів різної природи, зокрема фінансово-економічних, є надзвичайно актуальною з огляду на кількість публікацій з цієї тематики. У роботі розглядаються підходи до побудови моделі трендової складової часового ряду та екстраполяції отриманої аналітичної функції в часі [2], моделювання процесу, що породжує ряд на основі технології нейронних мереж [3], побудова багатофакторних регресійних [4] чи авторегресійних [5] моделей та їх удосконалення — метод групового урахування аргументів (МГУА) та його нечіткий варіант [6]. На відміну від зазначених вище методик, у цій роботі запропоновано методи прогнозування часових рядів на основі частотного спектра.

Наявність характерної циклічної поведінки економічних систем було помічено та пояснено М.Д. Кондратьєвим та рядом інших видатних еконо-

містів [7]. Їхні роботи дали поштовх для цього дослідження, основною ідеєю якого є ідентифікація та моделювання циклічних коливань у фінансово-економічних часових рядах та практичне використання побудованих моделей під час прогнозування часових рядів фінансово-економічної природи.

АНАЛІЗ ОСНОВНИХ ПУБЛІКАЦІЙ ЩОДО ПРОБЛЕМИ ДОСЛІДЖЕННЯ

Відома значна кількість робіт, в яких пропонується досліджувати частотний спектр сигналу та будувати прогноз на основі виявлених частотних закономірностей.

Методи, які засновані на перетворенні Фур'є, дають можливість аналізувати амплітудно-частотну характеристику часового ряду, але не дають можливості виразити у вигляді окремої змінної фази сигналу, яка є важливою для побудови прогнозу. Також частота сигналу, що аналізується, може бути лише кратною проміжку дискретизації, що не дає можливість налаштуватися на частоту, яка не є кратною проміжку дискретизації. Під час використання достатньо великої кількості гармонік, проблема вирішується, але такий підхід, на думку автора, не є оптимальним [8].

Вейвлет-аналіз [9] дає інформацію як про амплітуди сигналу, так і про фазу (зміщення дочірнього вейвлета) та частково долає першу із зазначених вище проблем. Але має місце збитковість інформації в перетворенні, тому що зберігається не основна частота, а Фур'є-спектр цієї частоти.

У роботі [10] подано алгоритм дискретного Фур'є-продовження в якості методики прогнозування низькочастотної складової часового ряду. Запропонований алгоритм дозволяє подолати зазначені вище проблеми, виділяючи моночастотні гармонічні складові, які легко екстраполюються в часі.

МЕТОДИКА ДИСКРЕТНОГО ФЕР'Є-ПРОДОВЖЕННЯ

В цій роботі подано низькочастотну апроксимуючу функцію виду:

$$y_{\text{abs}}(t) = a + bt + \sum_{i=1}^m c_i \sin(d_i t + e_i), \quad (1)$$

або для відносного масштабу:

$$y_{\text{rel}}(t) = a e^{bt} \prod_{i=1}^m c_i \sin(d_i t + e_i). \quad (2)$$

Значення параметрів моделі $a, b, c_1, c_2, \dots, c_m, d_1, \dots, d_m, e_1, \dots, e_m$ визначаються за допомогою мінімізації наступного критерію оптимальності (функціонал нев'язки):

$$F(a, b, c_1, \dots, c_m, d_1, \dots, d_m, e_1, \dots, e_m) = \sum_{i=1}^n (|y_i - y_{\text{abs}}(t_0 + i\Delta t)|)^k, \quad (3)$$

або для відносного варіанта:

$$F(a, b, c_1, \dots, c_m, d_1, \dots, d_m, e_1, \dots, e_m) = \sum_{i=1}^n \left(\left| 1 - \frac{y_i}{y_{\text{rel}}(t_0 + i\Delta t)} \right| \right)^k, \quad (4)$$

де $\{y_i\}$ — дискретно заданий часовий ряд, що прогнозується, k — параметр. При $k=1$ критерії (3) та (4) відповідають мінімальності відхилень за модулем, при $k=2$ — мінімальності квадратів відхилень, при $k \rightarrow \infty$ зазначені вище критерії відповідають критеріям мінімальності максимальних відхилень реальної та апроксимуючої кривої.

Під час розв'язування задачі оптимізації необхідно задати початкові оцінки параметрів, які оптимізуються, а також накласти обмеження на їх значення. Суттєвим обмеженням необхідно накласти на частоту d_i . У випадку короткої навчальної вибірки можлива ситуація, коли найкраща апроксимація відповідає невеликій частині повного гармонічного коливання. При цьому продовження цієї аналітичної кривої може бути не характерним для ряду, що досліджується. Тому необхідною має бути умова відповідності не менш ніж половини повного коливання довжині навчальної вибірки n , тобто

частота коливання не може бути нижчою величини $\frac{\pi}{n}$, де n — довжина

вибірки навчання. Також обмеження має бути накладене на високі частоти, що пояснюється: по-перше, складністю розв'язування задач оптимізації для високих частот, пов'язаними з наявністю великої кількості хибних локальних мінімумів; по-друге, можливою некоректністю такого наближення на високих частотах. Емпірично було вибрано обмеження 10 коливань за час навчальної вибірки.

Оскільки число параметрів функції F зростає зі збільшенням числа гармонік m , пропонується ітераційна апроксимація по одній (або двох) гармоніках, обчислення залишків та застосування наступної ітерації до більш високочастотних залишків для одночастинного наближення:

$$r_i(t) = r_{i-1}(t) - (a + bt + c_i \sin(d_i t + e_i)), \quad (5)$$

або для m -частинного наближення:

$$r_i(t) = r_{i-1}(t) - \left(a + bt + \sum_{j=1}^m c_{ij} \sin(d_{ij} t + e_{ij}) \right). \quad (6)$$

Мінімізація нелінійного функціонала нев'язки (3) або (4) пов'язана з труднощами, які спричинені існуванням декількох локальних мінімумів функції F у просторі значень параметрів. Для подолання цієї проблеми пропонується здійснювати оптимізацію з використанням декількох початкових оцінок значень параметрів. В якості початкових наближень для коефіцієнтів тренду вибираються коефіцієнти лінійного тренду, визначені за допомогою МНК. Вибирається 2 початкових значення для фази: 0 та π радіан. Це дає можливість уточнити величину фази апроксимуючої функції як в бік збільшення, так і в бік зменшення. Запропоновано декілька початкових значень для частоти, обраних із рівномірною сіткою в інтервалі між мінімальною та максимальною частотою. Емпірично вибрано кількість початкових значень частоти для низькочастотних коливань (перша ітерація, апроксимація безпосередньо вхідного ряду) $nf=3$, для середньочастотних (друга та подальші ітерації) $nf=5$.

ЕКСПЕРИМЕНТАЛЬНІ РЕЗУЛЬТАТИ

Результати апроксимації логарифму індексу Доу-Джонса (DJ) наведено на рис. 1. Як видно з рис. 1, точність апроксимації досить висока. Така апроксимація дозволяє виокремити низькочастотну складову ряду, що прогнозується, та отримати стаціонарні з точки зору статистики залишки (рис. 2). Отримані залишки піддаються аналогічній апроксимації, у результаті чого виділяється наступна низькочастотна складова, наявна в ряді. Таким чином, виконуючи деяку кількість ітерацій цієї процедури, отримаємо частотний спектр, який містить виокремлені на кожній ітерації низькочастотні складові.

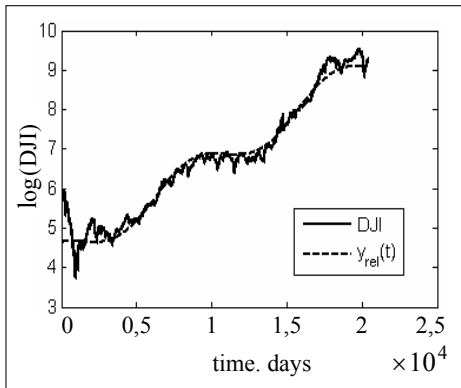


Рис. 1. Індекс Доу-Джонса (ітерація 1)

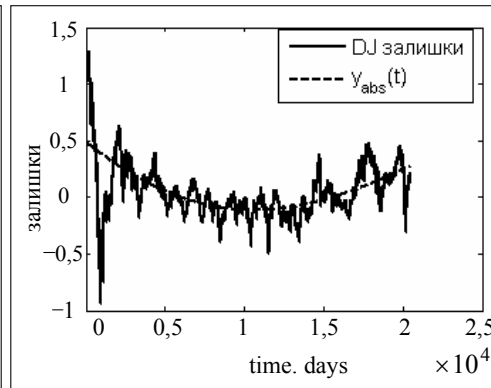


Рис. 2. Залишки та їх повторна апроксимація (ітерація 2)

Цей метод був запропонований авторами в роботі [10] та показав достатньо високі прогностичні можливості на спокійних фінансово-економічних часових рядах, що не містили кризових явищ. Були виявлені погіршення ефективності методики за наявності кризових явищ на навчальній чи прогнозній вибірці, які спричиняли помилки визначення частоти коливань, нестабільність цих частот та різкі падіння величини, що прогнозується. На нашу думку, такі явища можуть бути описані резонансами в коливаннях, що наявні в часовому ряді, для виявлення яких необхідно звернути увагу на випадки накладання декількох коливань.

БАГАТОЧАСТИННЕ НАБЛИЖЕННЯ АПРОКСИМУЮЧИХ ФУНКЦІЙ

Для покращення прогностичних результатів пропонуємо включити в апроксимуючу функцію для однієї ітерації дві синусоїди та проводити оптимізацію, змінюючи параметри одночасно обох гармонік (двочастинне наближення апроксимуючих функцій). Припускається, що такий підхід дозволить виявити складніші закономірності, які були пропущені через накладання частот, так звані «биття».

У табл. 1 наведено параметри функції (1) при одно- та двочастинному наближенні, а також значення середнього відхилення MSE. Даніми для прогнозування є індекс Доу-Джонса та індекс «Standard & Poors 500» (S&P). Цифрою 1 або 2 позначено варіанти одно- чи двочастинного наближення, також вказана перша чи друга гармоніка.

Таблиця 1. Результати двочастинного наближення індексів

Індекс	A	b	c_1	d_1	e_1	c_2	d_2	e_2	MSE
S&P2	$2,75 \cdot 10^{-4}$	3,040	0,0741	$1,62 \cdot 10^{-3}$	5,87	0,383	$6,63 \cdot 10^{-4}$	-0,506	105,28
S&P2 1 гар- моніка	$2,75 \cdot 10^{-4}$	3,040	0,0741	$1,62 \cdot 10^{-3}$	5,87	0	0	0	184,11
S&P2 2 гар- моніки	$2,75 \cdot 10^{-4}$	3,04	0	0	0	0,383	$6,63 \cdot 10^{-4}$	-0,506	112,46
S&P1	$2,74 \cdot 10^{-4}$	3,04	–	–	–	0,375	$6,61 \cdot 10^{-4}$	-0,483	113,9
S&P1 2 ітер	$1,19 \cdot 10^{-6}$	-0,0066	0,0735	$1,63 \cdot 10^{-3}$	5,78				105,55
DJ2	$2,03 \cdot 10^{-4}$	4,65	0,417	$4,10 \cdot 10^{-4}$	1,30	0,657	$5,94 \cdot 10^{-4}$	2,79	622,56
DJ2 1 гар- моніка	$2,03 \cdot 10^{-4}$	4,65	0	0	0	0,657	$5,94 \cdot 10^{-4}$	2,79	1960,47
DJ2 2 гар- моніки	$2,03 \cdot 10^{-4}$	4,65	0,417	$4,10 \cdot 10^{-4}$	1,30	0	0	0	1074,11
DJ1	$2,33 \cdot 10^{-4}$	4,31	0,402	$6,64 \cdot 10^{-4}$	2,15	–	–	–	1010,59
DJ1 2i	$-3,13 \cdot 10^{-5}$	1,36	1,18	$9,44 \cdot 10^{-5}$	-2,3	–	–	–	685,8

Порівняння відхилень MSE для двох методів, наведених в табл. 1, показує, що у випадку оцінювання коефіцієнтів одночасно для двох гармонік, мінімальні відхилення отримуються меншими, ніж у випадку повторного застосування апроксимації функції (1) до залишків. Також видно, що аналітичні функції при одно- та двочастинному наближенні не мають спільних частот, що свідчить про неможливість досягнення точності двочастинного наближення ітераційним застосуванням одночастинного.

На рис. 3 та 4 показано графіки апроксимуючих кривих із параметрами, які вказано в табл. 1. Чорними лініями (штриховою та пунктирною) показано криві апроксимації, здійснені згідно з одночастинним наближенням, отримані при першій та другій ітерації. Сірими лініями вказано результати двочастинного наближення: окремо перша ітерація та друга ітерація. Комбінація двох наведених ітерацій із 2-частинним наближенням зображена тонкою суцільною лінією (2 sintrend all).

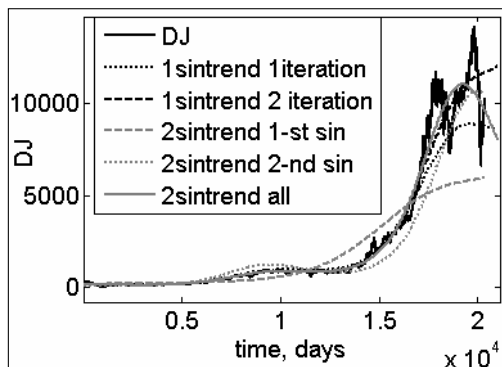


Рис. 3. Прогноз індексу Доу-Джонса

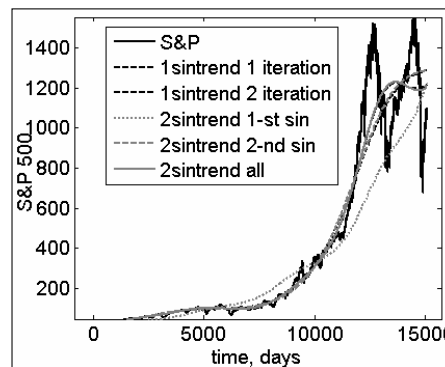


Рис. 4. Прогноз індексу S&P 500

Із рис. 3 та 4 бачимо, що двочастинна апроксимація є точнішою, хоча її частотні складові, взяті окремо, не дають точної апроксимації. Ці результати підтверджують актуальність досліджень, пов'язаних із багаточастинним наближенням гармонічної апроксимуючої функції.

Для прогнозування аналітичну суму гармонічних функцій екстраполюють у часі. Для порівняння прогнозних якостей методів одночастинного та двочастинного наближення, у табл. 2 наведено статистику відхилень прогнозів від реального продовження для різної довжини навчальної вибірки та різних моментів початку прогнозу. У табл. 2 критерієм якості прогнозу вибраний середній квадрат похибки MSE та MAPE:

$$MSE = \frac{1}{N} \sqrt{\sum_{i=1}^N (y_i - y(x_i))^2}, \quad (7)$$

$$MAPE = \frac{1}{N} \sum_{i=1}^N \frac{|y_i - y(x_i)|}{\bar{y}} 100\%. \quad (8)$$

Розрахунки прогнозів проводилися для 10-ти ітерацій при одночастинному наближенні та для 5-ти ітерацій при двочастинному. Довжина прогнозів — 1000-денних значень.

Таблиця 2. Порівняння прогнозів одночастинного та двочастинного наближення

Індекс	Довжина навчальної вибірки	Кількість гармонік в аналітичній функції	Похибка оцінювання на вибірці навчання		Похибки прогнозу	
			MSE	MAPE, %	MSE	MAPE, %
DJ	2000	1	60,894	3,1	525,287	18
DJ	2000	2	66,554	3,32	78,312	2,65
S&P	2000	1	3,254	2,27	33,436	10,25
S&P	2000	2	3,428	2,38	9,7438	2,66
DJ	3000	1	65,078	3,5	1520,36	37
DJ	3000	2	63,908	3,56	96,96	2,9
S&P	3000	1	4,531	2,97	50,318	22,3
S&P	3000	2	4,117	2,63	70,298	22,88
DJ	4500	1	56,721	3,93	1170,949	29
DJ	4500	2	53,735	3,883	691,24	24,7
S&P	4500	1	5,537	3,5	67,476	14,5
S&P	4500	2	4,86	3,22	96,584	24,99
DJ	7000	1	63,434	4,8	1533,841	27,4
DJ	7000	2	59,251	4,5	205,006	5,36
S&P	7000	1	7,102	4,3	284,586	35
S&P	7000	2	7,001	4,38	24,182	4,19
DJ	10500	1	71,28	6,1	584,353	12
DJ	10500	2	54,558	4,343	250,468	4,76
S&P	10500	1	11,389	5,5	289,896	28
S&P	10500	2	10,43	5,16	49,838	5,51

Табл. 2 містить усереднені результати прогнозування з різними довжинами навчальної вибірки (з 5-ьох експериментів із кожним індексом та вибраною довжиною навчальної вибірки) для одно- та двочастинного наближення. Ці результати показують, що в більшості випадків (95 %) двочастинне наближення дає кращі прогнози за критерієм середніх квадратів відхилень. Оптимальна довжина вибірки навчання для одночастинного варіанта коливається в межах від 2000 до 3000. При довжинах вибірки навчання 3000 отримані приклади прогнозів із $MAPE = 10,25\%$, що є дещо вищим за результати при інших довжинах вибірки. Це може бути пояснене наявністю постійної частоти коливань, яка залишилась незмінною на прогнозному фрагменті.

ВИСНОВКИ

У цій роботі запропоновано методіку прогнозування часових рядів на основі виявлення частотного спектра. Методика полягає в апроксимації часового ряду на проміжку навчання аналітичною функцією, яка містить тренд та комбінацію гармонічних коливань. Для оцінювання параметрів моделі використовуються нелінійні методи оптимізації для мінімізації функціоналу нев'язки. Експериментальна апробація показала ефективність запропонованого методу для прогнозування рядів фінансово-економічної динаміки. У порівнянні з відомими класичними підходами, ця технологія дає можливість передбачати моменти часу, в яких можлива зміна тенденції часового ряду, що є важливим у роботі трейдерів на фінансових ринках. У випадку високої інформативності отриманих прогнозів, про які говорять відповідність локальних мінімумів часових рядів прогнозу та реального продовження, класичні критерії оцінювання точності прогнозу (MSE , $MAPE$) не дають можливості виокремити цю перевагу методу в порівнянні з іншими методами прогнозування. Тому актуальним залишається питання розробки нових критеріїв точності прогнозів, які, по-перше, адекватно відобразили б співпадіння точок зміни тренду під час прогнозування, а по-друге, дозволили б спростити задачу пошуку оптимальних параметрів моделі. Похибки прогнозів цього методу можуть бути пояснені таким чином: наявністю фінансової кризи на проміжку навчання або прогнозування; похибками моделі частот, що спричиняють невідповідність екстремумів; похибками амплітуд коливань, які не є сталими і можуть збільшуватись під час кризи. Для подолання проблем, які були помічені, пропонується здійснити пошук оптимальних коефіцієнтів одночасно за двома гармоніками (двочастинне наближення) з метою отримання складніших комбінацій гармонічних коливань (накладання частот, «биття»), що покращує ефективність запропонованого методу. Продовженням роботи має бути удосконалення алгоритму для рядів зі змінними частотами.

ЛІТЕРАТУРА

1. Дербенцев В.Д., Сердюк О.А., Соловійов В.М., Шаранов О.Д. Синергетичні та екофізичні методи дослідження динамічних та структурних характеристик економічних систем: монографія. — Черкаси-Брама-Україна, 2010. — 300 с.

2. *Кендалл М., Стьюарт А.* Многомерный статистический анализ и временные ряды. — Глав. редакция физ.-математ. лит-ры изд-ва «Наука», 1976. — 375 с.
3. *Ежов А.А., Шумкий С.А.* Нейрокомпьютинг и его применения в экономике и бизнесе. — М.: МИФИ, 1998. — 224 с.
4. *Лукашин Ю.П.* Адаптивные методы краткосрочного прогнозирования временных рядов: учеб. пособ. — М.: Финансы и статистика, 2003. — 416 с.
5. *Бокс Дж., Дженкинс Г.* Анализ временных рядов. Прогноз и управление. — М.: Мир, 1974. — 197 с.
6. *Зайченко Ю.П.* Нечеткий метод группового учета аргументов при неопределенных входных данных // Системні дослідження та інформаційні технології. — 2007. — № 4. — С. 40–57.
7. *Кондратьев Н., Яковец Ю., Абалкин Л.* Большие циклы конъюнктуры и теория предвидения // Избранные труды. — М.: Экономика, 2002. — 768 с.
8. *Ronald N. Bravwell.* The Fourier Transform and its applications (Second edition, revised). — US: McGraw-Hill Book Company, 1978. — 464 p.
9. *Анушина Е.С., Поляхов Н.Д., Приходько И.А., Анушина Е.С., Хартян Е.В.* Вейвлет-теория в задачах прогнозирования // Изв. СПб ГЭТУ «ЛЭТИ». — 2008. — Вып. 4. — С 50–54.
10. *Сапцін В.М., Чабаненко Д.М.* Фур'є-продовження низькочастотних складових рядів економічної динаміки // Проблеми економічної кібернетики: тези доп. XIV Всеукр. наук.-метод. конф. (8–9 жовт.). — Х.: ХНУ ім. В.Н. Каразіна, 2009. — С. 132–133.

Надійшла 08.06.2010