

УДК 681.513

МОДЕЛЮВАННЯ БАГАТОВИМІРНИХ ЦИКЛІЧНИХ ПРОЦЕСІВ ЗА ДВОРІВНЕВИМ АЛГОРИТМОМ МГУА

Ю.В. Костенко

*Міжнародний науково-навчальний центр інформаційних технологій
i систем НАН та МОН України*

astrid@irtc.org.ua

Одним із важливих етапів на шляху розробки інтегрованого середовища обробки та зберігання інформації в задачах індуктивного моделювання є проектування засобів інтеграції зовнішніх даних у сховище. У статті розглядаються ключові проблеми, що виникають на цьому етапі, в тому числі проаналізовано структури вхідних даних за різними параметрами.
Ключові слова: індуктивне моделювання, метод групового урахування аргументів, структури даних, обробка та зберігання інформації, інтегроване середовище.

One of the most important steps towards developing an integrated environment of storing and handling information in the tasks of inductive modeling is the design of the integration of external data in the repository. The article discusses key issues arising at this stage, including the structure of input data analyzed by various parameters.

Keywords: *inductive modeling, group method of data handling, structures of data, handling and storing of information, integrated environment.*

Одним из важнейших этапов на пути разработки интегрированной среды обработки и хранения информации в задачах индуктивного моделирования является проектирование средств интеграции внешних данных в хранилище. В статье рассматриваются ключевые проблемы, возникающие на этом этапе, в том числе проанализированы структуры входных данных по различным параметрам.

Ключевые слова: индуктивное моделирование, метод группового учета аргументов, структуры данных, обработка и хранение информации, интегрированная среда.

ВСТУП

Розглядається задача моделювання коливальних процесів, що мають характерний циклічний, або сезонний тренд: наприклад, середньомісячна температура повітря має характерні максимуми в літні та мінімуми в зимові місяці. Однак ці максимуми і мінімуми рік від року кількісно не збігаються, тому процеси з такого роду сезонними коливаннями називають циклічними на відміну від строго періодичних процесів. До циклічних відносяться насамперед природні процеси з постійною тривалістю циклу (рік або доба), мінливість яких визначається впливом якихось додаткових факторів. До них можна віднести також деякі економічні (продаж сезонних товарів тощо) і технологічні процеси.

Є певний досвід моделювання природних циклічних процесів із застосуванням алгоритмів МГУА в класі дворівневих моделей [1], заснованих на критерії балансу прогнозів. Тут розглянуто досить загальний алгоритм

МГУА, що дозволяє синтезувати як дворівневі, так і багато інших видів лінійних прогнозуючих моделей.

1. Дворівневі моделі циклічних процесів

Розглянемо спочатку одновимірний циклічний процес без зовнішніх впливів, тобто окремо взятий часовий ряд. Нехай він поданий значеннями спостережень y_k , усередненими на деяких дискретних інтервалах — тараках, а індексація цих тараків k . Нехай також у циклі процесу міститься рівно K тараків, так що загальна довжина початкової вибірки дорівнює $K \times M$ тараків, де M — кількість циклів. Тоді основою запису дворівневих моделей є перехід від одновимірного ($(k=1, K \times M)$) до двовимірного (t, T) відліку часу, де $t = \overline{1, K}$, $T = \overline{1, M}$. Прикладами можуть бути такі відліки тарак-цикл: година-дoba ($K = 24$), місяць-рік ($K = 12$), квартал-рік ($K = 4$). При цьому середньотактові y_t і середньоциклові Y_T значення процесу називаються відповідно змінними нижнього і верхнього рівнів усереднення.

У цих позначеннях загальний вид дворівневої одновимірної моделі буде таким:

$$y_{tT} = a_{0t} + \sum_{j=1}^{SY} a_{jt} y_{t-j, T} + \sum_{r=1}^{QY} d_{rt} y_{t, T-r} + \sum_{h=1}^{HY} g_{ht} Y_{T-h+1}, \quad t = \overline{1, K}, \quad (1)$$

де SY , QY — число враховуваних запізнюваних значень по тараках і циклах відповідно для y_{tT} а HY — число значень змінної верхнього рівня Y_T . Модель (1) являє собою систему різницевих рівнянь, для якої необхідно задати також відповідні початкові значення, причому вихід кожного рівняння є входом для наступного, а вихідна величина $y_{k,T}$ є входом для рівняння з $y_{1,T+1}$ — обчислення виконуються циклічно (рекурентно). Структура моделі (1), званої «комплектом рівнянь» [1], та її коефіцієнти визначаються за алгоритмом МГУА. Склад аргументів цієї моделі може бути розширений за розрахунок введення ковзного усереднення таракових значень, тобто в праву частину (1) можна додати вираз

$$\sum_{s=1}^{FY} C_{st} \lambda_{tp_s}, \quad \lambda_{tp_s} = \frac{1}{p_s} \sum_{i=2}^{Ps} y_{t-i, T} \quad (2)$$

де FY — число ковзних середніх значень λ_{tp_s} , p_s — інтервал (число тараків) усереднення. Доцільно вводити члени цього типу для зменшення числа враховуваних запізнювань SY , QY і HY .

Якщо спостерігаються, крім y_{tT} , також зовнішні впливи w_{tT}^i $i = \overline{1, NW}$, загальна модель набуде вигляду:

$$\begin{aligned}
y_{tT} = & a_{0t} + \sum_{j=1}^{SY} a_{jt} y_{t-j,T} + \sum_{r=1}^{QY} d_{rt} y_{t,T-r} + \sum_{h=1}^{HY} g_{ht} Y_{T-h+1} + \\
& + \sum_{i=1}^{NW} \left(\sum_{j=1}^{SW} b_{ijt} w_{t-j+1,T}^i + \sum_{r=1}^{QW} f_{irt} w_{t,T-r}^i \right), \quad t = \overline{1, K},
\end{aligned} \tag{3}$$

де SW, QW — число врахуваних запізнювань для зовнішніх впливів по тактах і циклах відповідно. Очевидно, що в (3) можна ввести також ковзні середні цих впливів, а також їхні середньоциклові значення W_t .

Якщо моделюється багатовимірний процес, тобто сукупність NY взаємозалежних цикліческих процесів $y_{tT}^l, l = \overline{1, NY}$, то модель у вигляді системи NY комплектів по K рівнянь можна записати так (аналог системи (1)):

$$y_{tT}^l = a_{0t}^l + \sum_{i=1}^{NY} \left(\sum_{j=1}^{SY} a_{ijt}^l y_{t-j,T}^l + \sum_{r=1}^{QY} d_{irt}^l y_{t,T-r}^i \sum_{h=1}^{HY} g_{ih}^l Y_{T-h+1}^i \right), l = \overline{1, NY}, t = \overline{1, K}, \tag{4}$$

тобто загалом у цій системі $K \times NY$ рівнянь. Не важко записати також багатовимірний аналог (3) для випадку, коли на модельований процес впливає векторний зовнішній вплив $w_{tT}^i, i = \overline{1, NW}$.

Вище записані тільки деякі основні варіанти загального вигляду дворівневих моделей. Очевидно, що, комбінуючи інші сполучення чисел NY, SY, QY, FY, HY , а також аналогічні NW, SW , тощо, можна побудувати багато інших варіантів.

Цікаво зазначити, що модель із двовимірним відліком часу можна розглядати формально як деяку дискретну модель об'єкта з розподіленими параметрами, що змінюється як у часі τ , так і в просторі x , причому має місце аналогія t і τ , T і x .

Для аналізу властивостей моделей виду (1), (3), (4) їх варто звести до якого-небудь відомого виду різницевих рівнянь. Наприклад, у простому випадку $SY < K, QY = 0, HY = 1$ систему (1) можна перетворити або в одновимірне рівняння $y_k = a_0(k) + a_1(k)y_{k-1} + g(k)\tilde{y}_k$ з періодичними параметрами (де \tilde{y}_k — кусочно-постійна на довжині циклу функція), або у векторне рівняння вигляду $y_T = a_0 + Ay_{T-1} + gY_T$ з постійними параметрами, де $y_T = (y_{1T}, y_{2T}, \dots, y_{kT})^T$ [1]. При цьому можна встановити, що у випадку стійкої моделі (1) її розв'язками (у початковому часі k) можуть бути: періодична функція з періодом K ; періодична функція з періодом, більшим K ; неперіодична незатухаюча функція (в залежності від властивостей змінної Y_T). В [1] отримано оцінки часу перехідних процесів як для вектора y_T в цілому, так і для окремих його

компонентів. У принципі перехід до стандартного вигляду векторно-матричних рівнянь з наступним аналізом можливий і для моделей вигляду (3), (4). Однак для ідентифікації та наступного прогнозування рекурентна форма моделей циклічних процесів зручніша.

2. Про інші класи лінійних моделей

З наведених вище рівнянь, прийнявши $K=1$, можна як окремі випадки одержати інші добре відомі класи моделей (при цьому, очевидно, необхідно відразу покласти також $QY = HY = QW = HW = 0$, а замість двовимірного часу (t, T) використовувати одновимірне k). Наприклад, (1) перетворюється у звичайну авторегресійну модель

$$y_k = a_0 + \sum_{j=1}^{SY} a_j y_{k-j}, \quad (5)$$

яка при врахуванні зовнішніх впливів узагальнюється в аналог (3):

$$y_k = a_0 + \sum_{j=1}^{SY} a_j y_{k-j} + \sum_{i=1}^{NW} \sum_{j=1}^{SW} b_{ij} w_{k-j+1}^i \quad (6)$$

При багатовимірному виході $NY > 1$ одержуємо звичайну векторну різницеву модель

$$y_k^l = a_0^l + \sum_{i=1}^{NY} \sum_{j=1}^{SY} a_{ij}^l y_{k-j}^i + \sum_{i=1}^{NW} \sum_{j=1}^{SW} b_{ij}^l w_{k-j+1}^i, \quad l = \overline{1, NY}. \quad (7)$$

У рівняння (5) — (7) можуть бути додані члени у вигляді ковзних середніх значень як залежних y_k^l , так і незалежних w_k^l величин. Очевидно також, що при $SY = 0, SW = 1$ ми одержимо систему регресійних рівнянь, що описують статичний об'єкт:

$$y^l = a_0^l + \sum_{i=1}^{NW} b_i^l w^i, \quad l = \overline{1, NW}, \quad (8)$$

тобто при різних сполученнях чисел NY, SY, NW, SW можна одержати класи моделей (5) - (8) та їхні різновиди.

Таким чином, ми фактично одержали зручну класифікацію лінійних моделей, що охоплює більшість варіантів однорівневих (одновимірних і багатовимірних) і дворівневих (одновимірних і багатовимірних) моделей, що зустрічаються на практиці.

3. Алгоритм МГУА для дворівневого моделювання циклічних процесів

Призначений для побудови моделей природних і виробничих циклічних процесів з незмінною довжиною циклу. Моделі формуються у вигляді системи лінійних різницевих рівнянь каскадного типу: число рівнянь рівно числу інтервалів часу (тактів), на які ділиться період процесу. Прогнозування здійснюється шляхом циклічного переходу від одного рівняння системи до наступного. Розглядаються тільки окремо взяті процеси без урахування додаткових впливів. Приклади двовимірного відліку часу (період-такт): доба-година, рік-місяць, рік-квартал і т.д. Структура кожного рівняння та його коефіцієнти визначаються за комбінаторним алгоритмом СОМВІ.

Структура алгоритму відбиває основні етапи розв'язання задач моделювання в описаних вище класах моделей.

Вибір того або іншого класу моделей визначається заданням відповідних значень параметрів K (дво- або однорівневі моделі) і NY, SY, QY, FY, HY , які можуть приймати нульові значення (але не всі одночасно). При цьому число еквівалентних аргументів у будь-якому рівнянні моделі дорівнює

$$N = S + Q + F + H, \quad (9)$$

так що поєднання значень зазначених параметрів повинне бути таким, щоб N відповідало можливостям алгоритму, за яким ідентифікується структура окремого рівняння: застосування комбінаторного алгоритму СОМВІ вимагає обмеження $N \leq 16$.

Початкові дані задаються своїми середньотактовими (детальними) значеннями, тобто число значень змінної, що вводяться, дорівнює $M \times K$, де M - число циклів спостережень. При цьому задається також поділ M на три частини MA, MB й MC і число циклів прогнозування T .

Загальна структура моделі у вигляді системи K різницевих одновимірних дворівневих рівнянь має вигляд:

$$\begin{aligned} y_{tT} = & a_0 t + \sum_{j=1}^S a_{jt} y_{t-j,T} + \sum_{r=1}^Q a_{rt} y_{t,T-r} + \sum_{h=1}^P q_{ht} Y_{T-h+1} + \\ & + \sum_{q=1}^{SS} C_{qt} \left(\frac{1}{\sigma_q} \sum_{i=1}^{\sigma_q} y_{t-i,T} \right), \quad t = \overline{1, K}; T = \overline{1, M}, \end{aligned} \quad (10)$$

де : M – кількість періодів процесу;

K – кількість тактів у періоді;

y_{tT} – середнє значення вихідної змінної на t -му такті;

P – кількість аргументів у вигляді середньоперіодних значень;

SS – кількість аргументів у вигляді ковзних середніх;

σ – інтервал (число тактів) усереднення.

Оптимальна складність структури цієї системи визначається в результаті виконання таких операцій:

1 етап

1. Послідовно для кожного такту процесу відповідно до заданого набору аргументів (S, Q, P, SS) формується масив даних для вихідної змінної і аргументів і за алгоритмом COMBI отримується набір із C рівнянь.

2. Після отримання таких наборів рівнянь для перших L (з K) послідовних тактів виконується повний перебір можливих комбінацій рівнянь (всього таких комбінацій $C^{**}L$) і відбирається G кращих варіантів цих комплектів за частковим критерієм балансу

$$B_L = \sum_{i=1}^{MA+MB} (Y_i - \frac{1}{L} \sum_{j=1}^L \hat{y}_{ij})^2, \quad (11)$$

де \hat{y}_{ij} - результат покрокового інтегрування комплекту з L рівнянь.

3. Далі для наступного такту формується відповідний масив даних і за алгоритмом COMBI отримується набір із C рівнянь. Потім з раніше отриманих G кращих за частковим критерієм балансу комплектів відбираються $L-1$ останні за порядком слідування тактів рівняння і виконується перебір цих неповних комплектів спільно із C варіантами рівнянь, отриманих для поточного такту. При цьому знову за частковим критерієм балансу відбираються G кращих комплектів. Цей процес повторюється для всіх подальших тактів. Всього перебирається $C^{**}L + C * G * (K - L)$ варіантів. В результаті отримується G комплектів по K рівнянь.

II етап

З G комплектів відбирається CB кращих за основним критерієм балансу

$$B = \sum_{i=1}^{MA+MB} (Y_i - \frac{1}{K} \sum_{j=1}^K \tilde{y}_{ij})^2 \quad (12)$$

де \tilde{y}_{ij} - результат покрокового інтегрування оцінюваного повного комплекту. Ефективність кращих комплектів рівнянь перевіряється потім на всіх частинах початкової вибірки даних і виконується прогноз процесу на задане число періодів уперед за відомих прогнозів середньоперіодних значень Y_i .

Алгоритм реалізовано у вигляді обчислювальної програми на мові С.

Опис вхідних параметрів

$M, M1, M2$ – число періодів відповідно в навчальній, перевірочній і екзаменаційній частинах початкової вибірки даних;

$N1$ - число тактів у періоді процесу;

T – число періодів прогнозу за отриманими моделями;

C – число комплектів, сформованих для кожного такту (за програмою COMBI);

L – довжина (число тактів) ковзного інтервалу;

G – число комплектів моделей, відібраних на 1 етапі алгоритму;

CB – число кращих за критерієм балансу комплектів моделей;

PR – ключ управління виведенням інформації на екран: при $PR=0$ – значення управлюючих параметрів, масив початкових даних і результати розрахунків у вигляді відібраних комплектів моделей, розбаланс цих моделей, помилки контрольного прогнозу і результат «чистого прогнозу» (при $T>0$); при $PR=1$ додатково виводяться сформовані за COMBI C моделей для кожного такту та їх помилки;

S, Q – число аргументів у вигляді запізнювань відповідно по тактах і по періодах;

MP – число аргументів у вигляді середньоперіодних значень процесу;

F – число аргументів у вигляді ковзних середніх;

$FF[F]$ – масив, елементи якого вказують довжини інтервалів ковзного усереднення;

$RF[N1*MC]$ – масив середньотактових значень процесу (початкова вибірка даних, $MC=M+M1+M2$).

Результати розрахунків, що виводяться на екран

«Розбаланс відібраних комплектів» - значення критерію балансу для відібраних комплектів рівнянь. Для кожного з CB комплектів виводиться така інформація:

«Модель (комплект)» - номер чергового з CB комплекту «Поліноми оптимальної складності» - коефіцієнти рівнянь, що входять в описуваний комплект.

«Апроксимація» - розраховані значення процесу, що відповідають навчальній і перевірочній частинам початкової вибірки даних.

«Помилка апроксимації» - розбаланс на частині $M+M1$ початкової вибірки, нормований на суму квадратів середньоперіодних значень.

«Прогнозування» - перевірка на екзаменаційній послідовності (при $M2 > 0$) і прогноз на T періодів (при $T > 0$).

«ОШО, ОШС, ОШМ» – значення середньоквадратичної помилки контрольного прогнозування, нормованої трьома способами: на суму квадратів табличних

даних, на суму квадратів відхилень табличних даних від середнього; на суму квадратів відхилень від середніх значень для кожного такту.

«Екзаменаційна послідовність» - відповідні табличні дані.

«Контрольний прогноз» - результати контрольного прогнозування, що відповідають періодам екзамену.

«Прогноз» - («чистий прогноз») результати прогнозування на T періодів уперед.

«СКО ПО СМЛ» - середньоквадратичне відхилення прогнозу від середніх за всі періоди значень процесу для кожного такту, нормоване на суму квадратів цих значень.

«РОЗБАЛАНС» - значення критерію балансу для комплекту, що перевіряється.

При $PR > 0$ додатково виводиться:

hc – (інтервал початкового запізнювання) – число початкових періодів початкової вибірки, які використовуються для розрахунку початкових значень запізнюваних аргументів;

mg та $M1$ – скориговане значення сумарного числа періодів навчання і перевірки, а також скориговане число періодів перевірки;

„ТАКТ N” – номер поточного модельованого такту;

« $Y =$ » « $X =$ » – сформовані для поточного такту масиви даних вихідної величини і аргументів.

Вибір значень управлюючих параметрів

Розділення початкової вибірки даних на частини $M - M1 - M2$ слід робити згідно з рекомендаціями Інструкції з використання алгоритму COMBI.

Числа $N1$ і T визначаються умовами початкової задачі.

Числа S, Q, MP, F, FF , що визначають структуру моделей, задаються дослідником початкової задачі.

Значення PR вибирається залежно від необхідного обсягу інформації, що виводиться на екран.

Значення решти управлюючих параметрів повинні задовольняти таким обмеженням:

1. Число моделей, сформованих для кожного такту C - свобода вибору для програми COMBI - не повинне перевищувати числа варіантів моделей, сформованих у COMBI:

$$C < 2 ** (S + Q + F + MP)$$

2. Довжина ковзного інтервалу L не може перевищувати числа тактів в періоді

$$1 \leq L \leq N1$$

3. Число комплектів моделей G , відібраних на 1 етапі алгоритму, не може перевищувати величини $C * * L$.

4. Число відібраних комплектів тактових моделей на виході алгоритму CB не може перевищувати кількості комплектів, відібраних на 1 етапі.

Програмно реалізовано новий дворівневий алгоритм МГУА, призначений для моделювання і прогнозування часових рядів (цикліческих процесів) у класі дворівневих різницевих моделей. Такі моделі адекватні цикліческим процесам із суттєвою сезонною компонентою, що є типовим для економічних та екологічних задач. Клас дворівневих моделей є узагальненням відомих стандартних (однорівневих) класів моделей типу регресії та авторегресії, а також звичайної лінійної регресії.

5. Приклади розв'язання практичних задач

Нижче наведено результати розв'язання двох задач, що ілюструють застосування алгоритму у випадку окремих часових рядів, тобто одновимірних цикліческих процесів без урахування зовнішніх впливів ($NY=1$, $NW=0$). В усіх задачах допоміжними критеріями були $\Delta^2(B)$ (етап 2) і ε_{AB}^2 (етап 3).

Приклад 1.

Зміна квартальних значень ВВП України у 1996-2003 pp.

Дані за $M=7$ років при $MA=3$, $MB=4$, $MC=0$, при цьому $K=4$.

Параметри структури рівнянь: $S=2$, $Q=1$, $F=0$, число аргументів $N=4$.

Якість моделі відбита на рис. 3.7, причому помилка апроксимації 3.403020E-007.

Зазначимо, що на графіку подано результат інтегрування з початкових умов, що відповідають першій добі даних.

Поліноми оптимальної складності:

такт N 1:	269.364	0.000	0.000	0.000	0.779
такт N 2:	-730.409	0.366	0.000	0.000	0.632
такт N 3:	-2007.402	0.000	0.000	0.000	1.217
такт N 4:	13.227	-0.553	-1.972	0.000	3.568

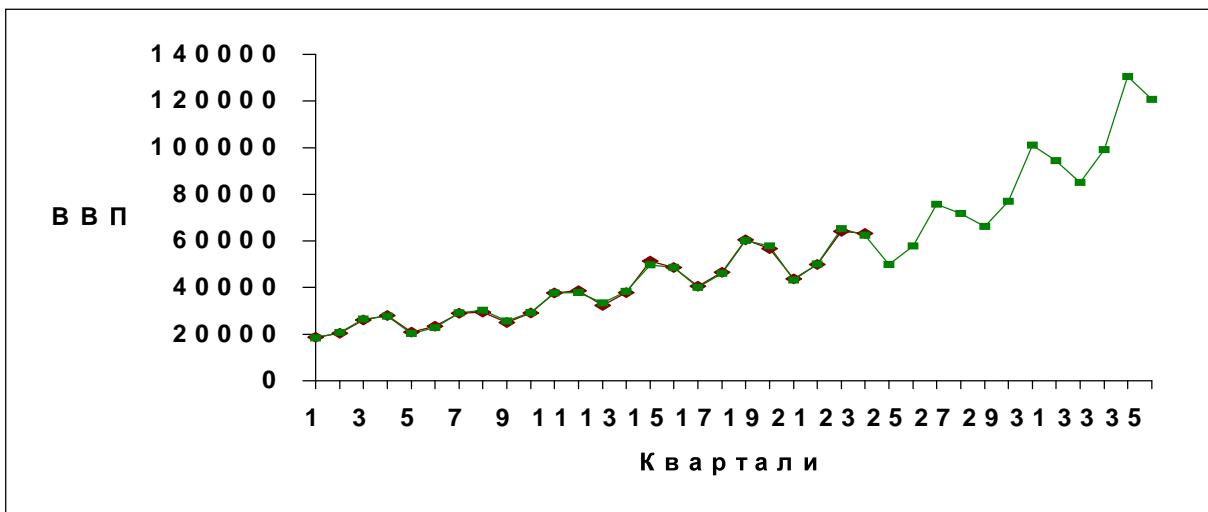


Рис. 1. Моделювання зміни квартальних значень ВВП України

Приклад 2. Зміна квартальних значень відношення загального обсягу інвестицій до ВВП у 1996-2003 рр.

Дані за $M = 7$ років при $MA = 3$, $MB = 4$, $MC = 0$, при цьому $K=4$. Параметри структури рівнянь: $S=2$, $Q=1$, $F=0$, число аргументів $N=4$. Якість моделі відбита на рис. 3.8, причому помилка апроксимації 3.403020E-007. Зазначимо, що на графіку подано результат інтегрування з початкових умов, що відповідають першій добі даних.

Поліноми оптимальної складності:

такт N 1:	0.037	0.000	0.000	0.000	0.666
такт N 2:	19.745	-0.455	0.047	0.000	0.000
такт N 3:	-3.047	0.230	-1.639	0.000	1.824
такт N 4:	1.815	-0.537	-1.126	0.000	2.988

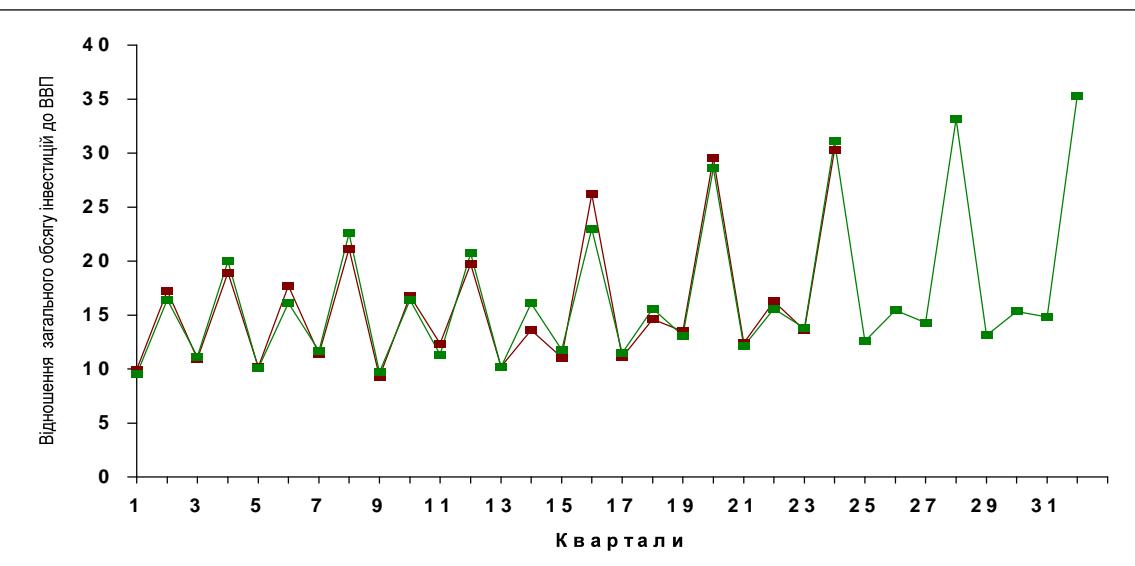


Рис. 2. Моделювання зміни квартальних значень відношення загального обсягу інвестицій до ВВП

Варто звернути увагу на той факт, що в обох прикладах у моделях відсутня залежність поточних квартальних значень від відповідних значень попереднього року, а також значень першого кварталу від будь-яких запізнювань. При цьому майже скрізь присутня залежність поточних квартальних значень від середньорічних, що свідчить про адекватність застосування дворівневих моделей.

Наведені приклади показують ефективність застосування описаного вище алгоритму для побудови прогнозуючих дворівневих моделей економічних цикліческих процесів. Ці моделі отримані на основі оброблення окремо взятих часових рядів. Урахування додаткових зовнішніх впливів дозволить поліпшити якість прогнозування.

Література

1. Ивахненко А.Г., Степашко В.С. Помехоустойчивость моделирования. – Киев: Наук. думка. – 1985. – 215 с.
2. Степашко В.С., Костенко Ю.В. Комбинаторно-селекционный алгоритм последовательного поиска модели оптимальной сложности // Праці І Міжнар. конф. з індуктивного моделювання, т.1, ч.1. – Львів ДНДІП, 2002. – С. 72-76.
3. Степашко В.С., Костенко Ю.В. О двухуровневых моделях природных циклических процессов (на примере ионосферного параметра МПЧ) // Четвертая украинская конференция по космическим исследованиям, Крым, Понизовка, 19-26 сентября 2004 г. Сборник тезисов. – К.: НКАУ, 2004. – С. 99.
4. Степашко В.С., Костенко Ю.В. Дворівневі моделі цикліческих процесів // 2-а Всеукраїнська науково-практична конференція „Проблеми електроніки та інформаційні технології”, 2–5 вересня 2010. – Львів-Чинадієво, 2010. – С. 29.