



УДК 330.42: 314.88.001.18 (477.46)

Гук В. І.<sup>1</sup>

## **ПРОГНОЗ НАСЕЛЕННЯ ЧЕРКАСЬКОЇ ОБЛАСТІ НА 2013 РІК МЕТОДОМ АНАЛІЗУ АДИТИВНИХ КОМПОНЕНТ ЧАСОВОГО РЯДУ**

*У статті розглянуто проблему демографічного прогнозування чисельності населення. В роботі використовується підхід, що базується на аналізі часових рядів. По кроках описано алгоритм розрахунку інтервального прогнозу чисельності населення методом аналізу адитивних компонент часового ряду.*

**Ключові слова:** адитивні компоненти часового ряду, прогнозування, тренд, сезонна компонента, ряд Фур'є, циклічна компонента.

### **ВСТУП**

Останнім часом в Україні спостерігається тенденція до постійного скорочення чисельності населення. Взагалі чисельність населення залежить від народжуваності, смертності, структури населення, масштабів та інтенсивності міграційних процесів. Врахувати всі ці фактори для створення математичної моделі та прогнозу чисельності населення практично неможливо. Тому динаміка

---

<sup>1</sup> Рецензент – Савченко С. О., д. е. н., доцент



чисельності та складу населення характеризується значним ступенем невизначеності, стохастичності.

Способом вирішення проблеми демографічного прогнозування є використання інтервальних прогнозів за допомогою часових рядів, коли немає необхідності описувати функціональну залежність чисельності населення від будь-яких змінних, крім часу.

Для прогнозування значень часових рядів використовуються як найпростіші методи, наприклад, прогнозування на основі тренду [1], так і досить складні адаптивні методи прогнозування [2], які засновані на представленні часового ряду окремою реалізацією випадкового процесу і моделюванні цього процесу на підставі аналізу автокореляційної функції. Демографічне прогнозування характеризується відсутністю різких змін членів часового ряду, що пов'язано з великою інерційністю такого соціального показника, як чисельність населення регіону [3]. Тому для прогнозування можна використовувати відносно простий і наочний метод аналізу компонент часового ряду [4]. Відповідно до цього методу часовий ряд  $Y_t$  представляється у вигляді суми окремих компонент: тренда  $T_t$ , сезонної  $S_t$ , циклічної  $C_t$  і випадкової  $\varepsilon_t$  компоненти. Таке представлення називають адитивною моделлю часового ряду і записують у вигляді:

$$Y_t = T_t + S_t + C_t + \varepsilon_t. \quad (1)$$

Поділ часового ряду на компоненти, визначення математичних формул для прогнозного розрахунку невідповідних компонент і знаходження закону розподілу випадкової компоненти і її числових параметрів є досить складним завданням [5].

#### **ПОСТАНОВКА ЗАВДАННЯ**

Метою є розробка єдиного формалізованого математичного методу прогнозування чисельності населення на підставі послідовного виділення та аналізу компонент часового ряду на прикладі задачі прогнозування чисельності населення Черкаської області.

#### **РЕЗУЛЬТАТИ ДОСЛІДЖЕННЯ**

Статистичні дані по чисельності населення Черкаської області доступні на сайті Державної служби статистики України [6]. Дані наводяться помісячно з лютого 2003 по жовтень 2012 року. Часовий ряд, тобто залежність чисельності населення  $Y$  від часу  $t$ , можна представити у вигляді графіка, наведеного на рисунку 1.

#### **Алгоритм розрахунків складається з наступних кроків:**

**Крок 1.** Вибір виду тренда, побудова формули тренда і знаходження залишку часового ряду без тренда.

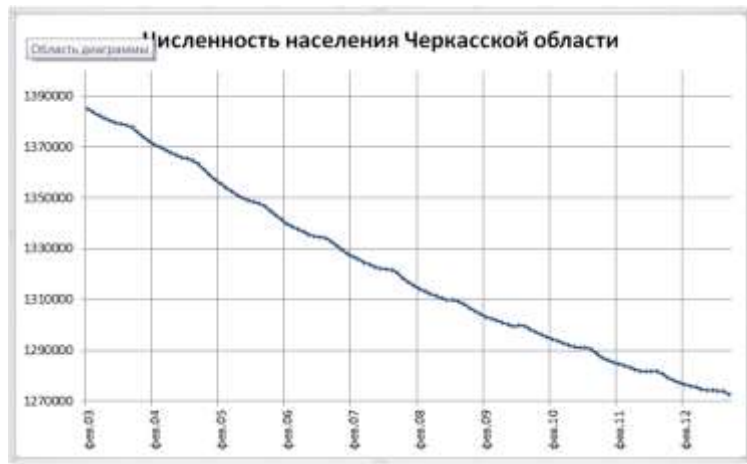


Рис. 1. Графік часового ряду чисельності населення Черкаської області (за даними Державної служби статистики України)

Виходячи з виду графіка, наведеного на рисунку 1, можна припустити, що тренд має поліноміальний вигляд і близький до лінійного. Формули різних трендів і коефіцієнт детермінації  $R^2$ , який характеризує величину достовірності апроксимації, наводяться в табл. 1.

Таблиця 1

Порівняння точності апроксимації даних різними видами трендів

Вид тренду	Поліноміальний тренд				Степеневий тренд	Експоненціальний тренд
	лінійний	квадратичний	кубичний	4-го ступеню		
Формула тренду	$y = -990,77t + 1 \cdot 10^6$	$y = 3,7608t^2 - 1434,5t + 1 \cdot 10^6$	$y = 0,0061t^3 + 2,6824t^2 - 1383,4t + 1 \cdot 10^6$	$y = -0,0005t^4 + 0,1347t^3 - 7,104t^2 - 1124,2t + 1 \cdot 10^6$	$y = (1 \cdot 10^6) e^{-(7E-4)t}$	$y = (1 \cdot 10^6) e^{0,025t}$
Коефіцієнт детермінації $R^2$	0,9862	0,9992	0,9992	0,9994	0,9885	0,849

Джерело: власна розробка



З таблиці видно, що найкращим слід вважати поліноміальний тренд другого порядку, тобто квадратичний тренд зі значенням коефіцієнта детермінації  $R^2 = 0,9992$ . Кубічний тренд має таке ж значення коефіцієнта детермінації, а поліноміальний тренд 4-го порядку хоча й має більше значення коефіцієнта детермінації  $R^2 = 0,9994$ , але відмінність між коефіцієнтами незначна і ускладнювати формулу тренда немає сенсу. Перевагу квадратичного тренду перед лінійним видно і на графіку, наведеному на рисунку 2.

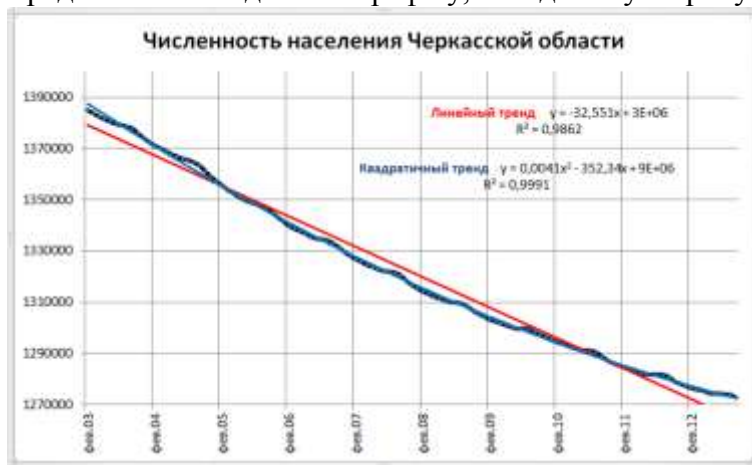


Рис. 2. Порівняння лінійного і квадратичного трендів часового ряду

Джерело: власна розробка

Для визначення коефіцієнтів квадратичного тренду

$$\varphi(t) = a_0 \cdot t^2 + a_1 \cdot t + a_2 \quad (2)$$

необхідно розв'язати систему нормальних рівнянь методу найменших квадратів

$$\begin{cases} a_0 \cdot \sum_{k=1}^n t_k^4 + a_1 \cdot \sum_{k=1}^n t_k^3 + a_2 \cdot \sum_{k=1}^n t_k^2 = \sum_{k=1}^n t_k^2 \cdot y_k \\ a_0 \cdot \sum_{k=1}^n t_k^3 + a_1 \cdot \sum_{k=1}^n t_k^2 + a_2 \cdot \sum_{k=1}^n t_k = \sum_{k=1}^n t_k \cdot y_k \\ a_0 \cdot \sum_{k=1}^n t_k^2 + a_1 \cdot \sum_{k=1}^n t_k + a_2 \cdot n = \sum_{k=1}^n y_k \end{cases} \quad (3)$$

Визначаючи за допомогою сумування відповідних величин коефіцієнти системи нормальних рівнянь (3), отримаємо наступну систему:

$$\begin{cases} 4479124299 \cdot a_0 + 47651409 \cdot a_1 + 540735 \cdot a_2 = 699515133370 \\ 47651409 \cdot a_0 + 540735 \cdot a_1 + 6903 \cdot a_2 = 148893381 \\ 540735 \cdot a_0 + 6903 \cdot a_1 + 117 \cdot a_2 = 154632498 \end{cases} \quad (4)$$



Після розв'язання системи лінійних алгебраїчних рівнянь знаходимо значення коефіцієнтів квадратичного тренду:

$$a_0 = 3,760780495; \quad a_1 = -1434,538572 \quad a_2 = 1,388901984 \cdot 10^6 \quad (5)$$

Таким чином, уточнена формула квадратичного тренду має вигляд:

$$\varphi(t) = 3,760780495t^2 - 1434,538572 \cdot t + 1388901984 \quad (6)$$

Графік залишку часового ряду, отриманого після видалення квадратичного тренду, наведений на рисунку 3.



Рис. 3. Залишок часового ряду після видалення тренду

Джерело: власна розробка

**Крок 2.** Виділення сезонної компоненти часового ряду. Фур'є-аналіз періодичної функції, представлення сезонної компоненти у вигляді декількох головних гармонік ряду Фур'є і знаходження залишку часового ряду без тренду і сезонної компоненти.

Сезонна компонента часового ряду – це коливальна складова ряду з періодом 12 місяців. Згідно з методом простої середньої [4], для виділення сезонної компоненти слід усереднити дані про чисельність населення по місяцях за всі доступні для аналізу роки. Після проведення Фур'є-аналізу отримуємо тригонометричний многочлен для апроксимації сезонної компоненти часового ряду у вигляді

$$\begin{aligned} S(t) = & \frac{a_0}{2} + a_1 \cdot \cos(\omega \cdot t) + b_1 \cdot \sin(\omega \cdot t) + \\ & + a_2 \cdot \cos(2\omega \cdot t) + b_2 \cdot \sin(2\omega \cdot t) + \\ & + a_3 \cdot \cos(3\omega \cdot t) + b_3 \cdot \sin(3\omega \cdot t) + \dots + \\ & + a_n \cdot \cos(n\omega \cdot t) + b_n \cdot \sin(n\omega \cdot t) \end{aligned} \quad (7)$$



де  $\omega = \frac{2\pi}{T}$

$$a_j = \frac{2}{n} \sum_{k=0}^{n-1} y_k \cos\left(j \frac{2\pi k}{n}\right) \quad b_j = \frac{2}{n} \sum_{k=0}^{n-1} y_k \sin\left(j \frac{2\pi k}{n}\right)$$

Після відкидання гармонік в формулі (7), амплітуда яких складає менше, ніж 5 % від амплітуди першої гармоніки, отримуємо сезонну компоненту, графік якої наведений на рис. 4.



Рис. 4. Сезонна компонента часового ряду, виділена методом Фур'є  
Джерело: власна розробка

До формули розкладу Фур'є, яка використана для побудови графіку, входять перша, друга та третя гармоніки. Старші гармоніки дають незначний внесок, що видно з табл. 2, і тому ними можна знехтувати.

Таблиця 2

Амплітуди гармонік у відсотковому відношенні до першої гармоніки

Перша гармоніка	Друга гармоніка	Третя гармоніка	Четверта гармоніка	П'ята гармоніка	Шоста гармоніка
100,00 %	31,34 %	12,85 %	3,98 %	3,04 %	0,42 %

Формула сезонної компоненти має вигляд

$$S(t) = 0,8941835 - 264,88519\mathbb{B} \cdot \cos(0,52359876 \cdot t) - 885,18027\mathbb{A} \cdot \sin(0,52359876 \cdot t) - 219,10151\mathbb{B} \cdot \cos(1,047\mathbb{D}755 \cdot t) + 189,28050\mathbb{A} \cdot \sin(1,0471975 \cdot t) + 112,91276\mathbb{D} \cdot \cos(1,570\mathbb{D}633 \cdot t) - 36,834262\mathbb{A} \cdot \sin(1,570\mathbb{D}633 \cdot t) \quad (8)$$



Графік залишку часового ряду, отриманого після видалення сезонної компоненти, має вигляд, що наведений на рисунку 5.



Рис. 5. Залишок часового ряду після видалення сезонної компоненти  
Джерело: власна розробка

**Крок 3.** Виділення циклічної компоненти часового ряду методом простої середньої.

Наведений на рис. 5 графік залишку часового ряду має ще одну коливальну складову з періодом, що значно перевищує один рік. Після обчислення значень автокореляційної функції було визначено, що періодичність залишку часового ряду складає 72 місяці, тобто часовий ряд чисельності населення має коливальну складову з періодом 6 років. Формула цієї циклічної складової визначається аналогічно формулі сезонної складової та має наступний вигляд

$$\begin{aligned}
 C(t) = & -38,4700705 - \\
 & -88,1084492 \cdot \cos(0,08726643 \cdot t) - 286,003107 \cdot \sin(0,08726643 \cdot t) - \\
 & + 114,841932 \cdot \cos(0,174529 \cdot t) - 184,965777 \cdot \sin(0,174529 \cdot t) + \\
 & - 96,9171449 \cdot \cos(0,261794 \cdot t) + 41,4045188 \cdot \sin(0,261794 \cdot t) - \\
 & - 144,678508 \cdot \cos(0,349659 \cdot t) + 128,445490 \cdot \sin(0,349659 \cdot t)
 \end{aligned} \quad (9)$$

Графік залишку часового ряду, отриманого після видалення циклічної компоненти (9), має вигляд, що наведений на рисунку 6.

**Крок 4.** Виділення та аналіз випадкової компоненти. Знаходження оцінок числових параметрів випадкової величини.

Цей залишок можна вважати випадковою компонентою, яка має хаотичний характер. За допомогою критерію Пірсона було показано, що можна прийняти припущення про нормальний закон розподілу випадкової компоненти. Оцінки числових параметрів випадкової компоненти часового ряду наведені в табл. 3.



Рис. 6. Залишок часового ряду після видалення регулярних компонент  
Джерело: власна розробка

Таблиця 3

Оцінка числових параметрів випадкової компоненти часового ряду

Найменування параметру	Оцінка параметру
Математичне сподівання	-4,85841 E-14
Середнє квадратичне відхилення	439,3204595
Коефіцієнт ексцесу	-0,387013661
Коефіцієнт асиметрії	0,036139586

Джерело: власна розробка

**Крок 5.** Визначення відносних похибок прогнозування шляхом порівняння прогнозних і реальних значень часового ряду. Побудова точкового прогнозу.

Будуємо прогноз кожної компоненти часового ряду і сумарний прогноз чисельності населення області на 2013 рік. Для цього:

а) квадратичний тренд, сезонну складову та циклічну складову за формулами (6), (8) та (9) розраховуємо ще на 12 місяців 2013 року;

б) випадкову складову продовжуємо на 2013 рік за допомогою інструменту «Генерація случайных чисел» надбудови «Анализ данных» MS Excel при значеннях математичного сподівання і середнього квадратичного відхилення, які наведені в таблиці 3;

в) підсумовуємо всі компоненти часового ряду і в результаті отримуємо прогноз чисельності населення, графік якого наведено на рис. 7.

Для знаходження відносної похибки прогнозу виконується прогноз вже відомих даних, тобто виконується прогноз на 2012 рік на підставі даних на кінець 2011 року. Після порівняння розрахункових та фактичних даних встановлено, що максимальна похибка прогнозу складає 9,09 %, тобто не перевищує 10 %.



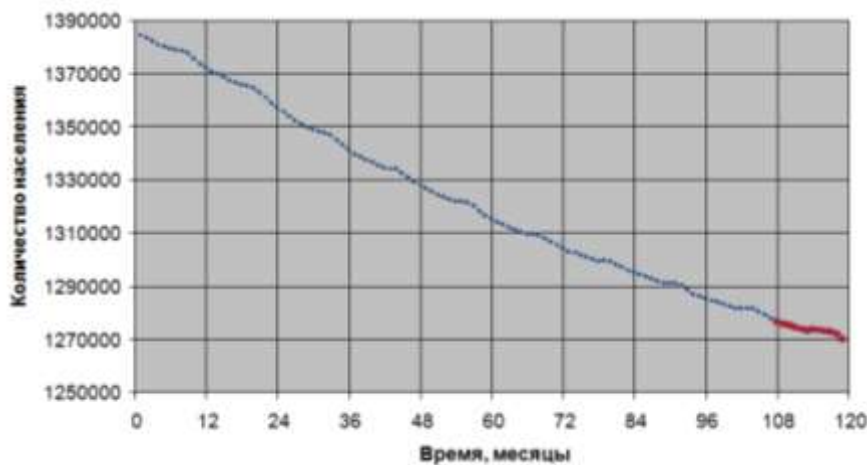


Рис. 7. Прогнозний графік зміни чисельності населення Черкаської області

Джерело: власна розробка

Віднімаючи від прогнозованої чисельності населення значення, що відповідає половині відносної похибки, отримуємо мінімальну прогнозовану чисельність населення, а додаючи відповідне значення, отримуємо максимальну прогнозовану чисельність населення. Ці дані наведені в таблиці 4.

Таблиця 4

Розрахунковий діапазон чисельність населення Черкаської області з урахуванням похибки прогнозування

Місяць	Мінімальне прогнозне значення	Максимальне прогнозне значення
Січень	1160770	1392900
Лютий	1159840	1391790
Березень	1157470	1391350
Квітень	1158820	1390570
Травень	1158310	1389960
Червень	1157710	1389240
Липень	1157990	1389570
Серпень	1157950	1389520
Вересень	1157560	1389060
Жовтень	1157310	1388760
Листопад	1156530	1387820
Грудень	1154920	1385880

Джерело: власна розробка



### **ВИСНОВКИ**

- 1) Запропонований алгоритм прогнозування на основі аналізу адитивних компонент часового ряду дозволяє з достатньою для практики точністю виконувати прогнозування соціально-економічних величин.
- 2) На підставі отриманих прогнозних даних видно, що в 2013 році збережеться тенденція до зменшення населення Черкаської області. Згідно виконаного прогнозу, населення Черкаської області в грудні 2013 року зменшиться і становитиме близько 1 270 400 осіб (з відносною похибкою 10 %). Тенденція до зменшення чисельності населення області спостерігатиметься і в найближчому майбутньому (щонайменше 5 років).

### **СПИСОК ЛІТЕРАТУРИ**

1. Кобелев Н. Б. Практика применения экономико-математических методов и моделей : учебно-практическое пособие / Н. Б. Кобелев. – М. : Финстатинформ, 2000. – 248 с.
2. Арженовский С. В. Статистические методы прогнозирования : учебное пособие для аспирантов / С. В. Арженовский, И. Н. Молчанов. – Ростов-на-Дону : Ростовский государственный экономический университет, 2001. – 74 с.
3. Касьяненко В. О. Моделювання та прогнозування економічних процесів. Конспект лекцій : навчальний посібник / В. О. Касьяненко, Л. В. Старченко. – Суми : Університетська книга, 2006. – 192 с.
4. Афанасьев В. Н. Анализ временных рядов и прогнозирование : учебник / В. Н. Афанасьев, М. М. Юзбашев. – М. : Финансы и статистика, 2001. – 228 с.
5. Методы и модели анализа временных рядов: Методические указания к лабораторным работам / Составитель С. И. Татаренко. – Тамбов : Тамбовский государственный технический университет, 2008. – 32 с.
6. Державна служба статистики України [Електронний ресурс]. – Статистична інформація. Демографічна ситуація. – Режим доступу : <http://www.ukrstat.gov.ua/> – Чисельність населення.

*Дата надходження до редакції – 20.03.2013 р.*