



УДК 004.6+519.2

РОЗРОБКА ЕВОЛЮЦІЙНОГО МЕТОДУ ДЛЯ ПРОГНОЗУВАННЯ ЧАСОВИХ РЯДІВ

Мулеса О. Ю.¹, Снитюк В. Є.²¹ Державний вищий навчальний заклад «Ужгородський національний університет», м. Ужгород, Україна² Київський національний університет імені Тараса Шевченка, м. Київ, УкраїнаORCID: ¹ <http://orcid.org/0000-0002-6117-5846>, ² <http://orcid.org/0000-0002-9954-8767>E-mail: ¹ oksana.mulesa@uzhnu.edu.ua, ² snytyuk@gmail.com

Copyright © 2020 by author and the journal "Automation of technological and business - processes.

This work is licensed under the Creative Commons Attribution International License (CC BY).

<http://creativecommons.org/licenses/by/4.0>DOI: <https://doi.org/10.15673/atbp.v12i3.1854>

Анотація. Процеси прийняття рішень щодо діяльності об'єктів господарювання, як правило, пов'язані з необхідністю аналізу основних показників їх діяльності. Виявлення тенденцій зміни числових показників в часі дозволяє робити припущення щодо їх майбутніх значень. Такі задачі можна звести до задач прогнозування часових рядів, які полягають у дослідженні законів зміни значень ряду та, на основі заданого критерію точності, знаходження прогнозних значень. Аналітичний огляд сучасних наукових публікацій показав, що задача прогнозування часових рядів є актуальною. Існує багато досліджень присвячених розробці ефективних гібридних методів прогнозування, в основі яких містяться декілька інших методів.

Дослідження присвячене розробці прогнозної моделі, яка використовує кращі властивості базових моделей прогнозування, дозволяє підвищити точність прогнозу та його волатильність.

В ході дослідження було розроблено еволюційний метод прогнозування на основі базових моделей прогнозування. Для обчислення прогнозних значень будується оптимізаційна модель, в яку входять прогнозні значення, обчислені за допомогою базових моделей. Параметри моделі можуть бути визначені за допомогою генетичного алгоритму. Критеріями якості прогнозної схеми були відносна похибка прогнозування, а також волатильність прогнозу. Такий підхід дозволяє зменшити відхилення прогнозних значень від точних.

Виконано експериментальну верифікацію розробленого методу прогнозування. Виконано порівняльний аналіз результатів роботи розробленого методу та інших методів прогнозування для часового ряду «Кількість хворих на СНІД». Показано, що використання прогнозної схеми дозволяє як підвищити точність прогнозу, так і покращити його волатильність.

Abstract. Usually, decision-making processes regarding the activities of economic entities require core activity metrics analysis. Identification of trends in changes in numerical indicators over time allows us to assume their future values. Such tasks can be reduced to forecasting of numerical sequences, which based on investigation of rules of changing values of sequence and discovering forecasting values by defined precision criteria. Analytical overview of recent scientific papers shows that forecasting numerical sequences is topical. There are lots of research of development an effective hybrid forecasting methods based on several of other methods.

Research is about development of forecasting model which use best properties of base forecasting models, allows us to improve forecasting precision and its volatility.

During the research an evolutionary method of forecasting based on base forecasting models was developed. To compute forecasting values we build optimisation model which includes forecasting values taken from base models. Parameters of model can be defined by genetic algorithm. The quality criteria of forecasting system was relative error of forecasting and volatility of forecast. Such approach allows us to reduce a deviation of forecasting values against exact values.

Experimental verification of the proposed forecasting method was done. A comparative analysis of the results of the developed method and other forecasting methods for the time series "Number of AIDS patients" was performed. It is shown that the use of the forecast scheme allows both to increase the accuracy of the forecast and to improve its volatility.

Ключові слова: часовий ряд, еволюційний метод, точність прогнозу, волатильність прогнозу, прогнозна схема.

Keywords: time series, evolutionary method, forecast accuracy, forecast volatility, forecast scheme.



1. Вступ

Процеси прийняття рішень щодо діяльності об'єктів господарювання, як правило, пов'язані з необхідністю аналізу основних показників їх діяльності. Виявлення тенденцій зміни числових показників в часі дозволяє робити припущення щодо їх майбутніх значень. Такі задачі можна звести до задач прогнозування часових рядів, які полягають у дослідженні законів зміни значень ряду та, на основі заданого критерію точності, знаходження прогнозних значень.

Для розв'язання задачі прогнозування на основі часових рядів можливим є застосування великої кількості різних методів прогнозування. До відомих класів методів прогнозування можна віднести регресійні методи, методи на основі авторегресії, методи, які дозволяють враховувати сезонність, нейромережні методи тощо. Проте, відомою особливістю проблеми прогнозування є те, що, як правило, неможливо розробити уніфікованих рекомендацій щодо того, який метод прогнозування є ефективнішим для кожного конкретного часового ряду. Таким чином, тягар вибору методу прогнозування лягає на аналітика. Тому, актуальною є розробка методу, який дозволить би в автоматичному режимі враховувати кращі результати різних методів при прогнозуванні заданих числових показників.

2. Аналіз літературних даних і постановка проблеми

Аналітичний огляд сучасних наукових публікацій показав, що задача прогнозування часових рядів є актуальною. В основі досліджень, при цьому, лежать різні концепції. Так в [1] розглядається нейромережний метод прогнозування на основі моделі авторегресійного інтегрованого ковзного середнього. В роботі відзначено, що використання зазначеного гібридного підходу в роботі нейронної мережі, дозволяє досягти більш високої точності прогнозування. Роботи [2, 3] присвячені розробці гібридної моделі, яка дозволяє розбивати часовий ряд на його лінійні та нелінійні компоненти. Таким чином, авторам вдалося поєднати можливості трьох різних моделей. Показано, що така модель дозволяє підвищити точність прогнозування. [4] містить метод прогнозування часових рядів з використання ймовірнісної нейронної мережі, яка складається з декількох шарів машини Больцмана. Розроблений метод був застосований також для короткострокового прогнозування хаотичних рядів. [5] присвячена аналізу стратегій машинного навчання для прогнозування часових рядів. В [6] міститься метод умовного прогнозування часових рядів на основі адаптації WaveNet. Показано, що такий метод добре підходить для задач регресійного типу навіть у випадку, коли ряд є коротким.

Також, існує ряд досліджень, присвячених розробці ефективних гібридних методів прогнозування, в основі яких містяться декілька інших методів. Так, наприклад, в [7] виконано порівняльний результатів прогнозування часового ряду різними відомими методами прогнозування та гібридним методом, який включає в себе два-три інші методи. Відзначено, що різні комбінації гібридної моделі є найбільш ефективні, в порівнянні з звичайними методами. Роботи [8-10] присвячені розробці комбінованої схеми на основі базових прогнозних моделей. В основі запропонованої моделі лежить ідея побудови прогнозної схеми для визначення прогнозного значення у формі опуклої лінійної комбінації результатів прогнозування часового ряду базовими методами. Такий підхід в ряді випадків дозволяє покращити результати прогнозування, проте, він має ряд обмежень при застосуванні.

Таким чином, розробка прогнозних моделей, які використовують кращі властивості базових моделей прогнозування, дозволяє підвищити точність прогнозу та його волатильність.

3. Мета і завдання дослідження

Метою дослідження є розробка ефективної моделі прогнозування часових рядів. Для досягнення поставленої мети були розв'язані такі задачі:

- побудувати математичну модель задачі прогнозування часових рядів;
- розробити гнучку прогнозу схему на основі базових моделей, основні параметри якої визначаються еволюційним методом для підвищення точності прогнозування та волатильності прогнозу;
- здійснити експериментальну верифікацію отриманих результатів та виконати порівняльний аналіз.

4. Еволюційний метод прогнозування часових рядів

Розглянемо задачу прогнозування на основі часових рядів у такій постановці:

Нехай $v_1, v_2, \dots, v_t, \dots, v_n$ – деякий дискретний часовий ряд без пропусків довжини n . Значення часового ряду фіксуються в дискретні моменти часу $i = \overline{1, n}$. Задача прогнозування часових рядів полягає у побудові вирішального правила F , яке дозволяє знайти оцінку значення ряду в майбутні періоди часу, що являє собою прогноз в заданій точці $n+m$ з кроком прогнозу m ($m \geq 1$) [11]. Тоді, формально, залежність прогнозного значення φ_{n+m}^0 від ретроспективних даних можна записати:

$$\varphi_{n+m}^0 = F(v_{n-k+1}, v_{n-k+2}, \dots, v_n, m),$$

де k – глибина передісторії, m – крок прогнозування.

Розглянемо метод побудови еволюційного методу прогнозування на основі базових моделей прогнозування. Для цього на першому етапі виберемо систему базових моделей прогнозування M_1, M_2, \dots, M_q . Серед них можуть бути регресійні, нейромережні та інші методи прогнозування.

Зафіксуємо значення кроку прогнозування m ($m \geq 1$) та глибини передісторії k . Обчислимо відповідні прогнозні значення для відомих елементів ряду обраними базовими методами та запишемо їх у таблицю (табл.1):



Таблиця 1 - Результати прогнозування базовими прогнозними моделями

Номер елемента	Прогнозні моделі			
	M_1	M_2	...	M_q
$k + m$	$\varphi_{k+m}^{(1)}$	$\varphi_{k+m}^{(2)}$	$\varphi_{k+m}^{(q)}$
$k + m + 1$	$\varphi_{k+m+1}^{(1)}$	$\varphi_{k+m+1}^{(2)}$	$\varphi_{k+m+1}^{(q)}$
...
n	$\varphi_n^{(1)}$	$\varphi_n^{(2)}$	$\varphi_n^{(q)}$

Побудуємо прогнозну схему у вигляді функції для обчислення прогнозного значення часового ряду в момент часу p ($k + 2m \leq p \leq n + m$), за правилом:

$$G(p) = \alpha_1 \varphi_p^{(1)} + \alpha_2 \varphi_p^{(2)} + L + \alpha_q \varphi_p^{(q)}, \quad (1)$$

де α_j ($j = \overline{1, q}$) – ваговий коефіцієнт прогнозної моделі M_j для обчислення прогнозного значення часового ряду в момент часу p ($k + 2m \leq p \leq n + m$).

Для обчислення значень параметрів схеми α_j ($j = \overline{1, q}$) при заданому значенні p побудуємо функціонал:

$$L(\alpha_1, \alpha_2, \dots, \alpha_q) = \sum_{i=k+m}^{p-m} \left| \frac{G(i) - v_i}{v_i} \right|, \quad (2)$$

та застосуємо генетичний алгоритм для розв'язання задачі мінімізації функції:

$$L(\alpha_1, \alpha_2, \dots, \alpha_q) \rightarrow \min. \quad (3)$$

Нехай $(\alpha_1^*, \alpha_2^*, \dots, \alpha_q^*) \in \text{Arg min } L(\alpha_1, \alpha_2, \dots, \alpha_q)$. Тоді, для визначення точності отриманих прогнозних значень, на наступному етапі обчислимо величини:

$$\Delta_j = \sum_{i=k+m}^{p-m} \left| \frac{\varphi_i^{(j)} - v_i}{v_i} \right|, \quad j = \overline{1, q},$$

та

$$\Delta_0 = L(\alpha_1^*, \alpha_2^*, \dots, \alpha_q^*).$$

У випадку, якщо виконується умова:

$$\forall j \in \{1, 2, \dots, q\} \quad \Delta_0 \leq \Delta_j,$$

то можна зробити висновок, що застосування схеми (1) із заданими параметрами $(\alpha_1^*, \alpha_2^*, \dots, \alpha_q^*)$ дозволяє підвищити точність прогнозу у порівнянні з базовими прогнозними моделями M_1, M_2, \dots, M_q .

Іншим критерієм якості прогнозної схеми є волатильність отриманого прогнозу, яка може бути оціненою середньоквадратичним відхиленням за формулами (4)-(5):

$$\sigma_j = \sqrt{\frac{\sum_{i=k+m}^{p-m} (\varphi_i^{(j)} - v_i - \delta_j)^2}{p - k - 2m + 1}}, \quad j = \overline{1, q}, \quad (4)$$

$$\sigma_0 = \sqrt{\frac{\sum_{i=k+m}^{p-m} (G(i) - v_i - \delta_0)^2}{p - k - 2m + 1}}, \quad (5)$$

де

$$\delta_j = \frac{1}{p - k - 2m + 1} \sum_{i=k+m}^{p-m} (\varphi_i^{(j)} - v_i), \quad j = \overline{1, q},$$

$$\delta_0 = \frac{1}{p - k - 2m + 1} \sum_{i=k+m}^{p-m} (G(i) - v_i).$$

Таким чином, прогнозне значення φ_p може бути обчислене за формулою:

$$\varphi_p = \alpha_1^* \varphi_p^{(1)} + \alpha_2^* \varphi_p^{(2)} + L + \alpha_q^* \varphi_p^{(q)}.$$



Експериментальна верифікація розробленого методу. Для експериментальної верифікації було розглянуто часовий ряд з даними про кількість хворих на СНІД на початок року в Закарпатській області [12]. Часовий ряд складався з 23 елементів. Прогнозування кількості хворих на заданій території дозволяє здійснювати планування діяльності кадрового складу медико-соціальних закладів, для максимального охоплення хворих потрібними їм послугами. Важливим при такому прогнозуванні є забезпечення мінімального відхилення прогнозних значень від реальних. Точність прогнозу, при цьому, є менш важливою.

За базові моделі було вибрано модель авторегресії з кроком передісторії $k = 4$ та модель тренду у вигляді поліноміальної функції третього порядку. Прогнозування здійснювалося на 1 крок вперед.

Результати обчислень наведені в табл.2. та на рис.1.

Таблиця 2 - Результати прогнозування значень часового ряду «Кількість хворих на СНІД» базовими моделями

Рік	Елементи часового ряду	Метод авторегресії	Модель тренду
1997	8	-	-
1998	35	-	-
1999	51	-	-
2000	63	-	-
2001	77	-	-
2002	94	-	-
2003	91	-	-
2004	77	-	-
2005	73	-	-
2006	100	81,1748	49,016
2007	109	117,9946	99,3452
2008	136	113,2296	127,3312
2009	154	151,6245	168,5805
2010	185	170,5247	198,5224
2011	221	210,8523	236,5665
2012	251	251,9838	280,3528
2013	288	282,4377	318,5382
2014	322	322,6126	358,7616
2015	370	359,3304	396,0834
2016	421	417,1556	442,9315
2017	484	476,5829	496,1607
2018	571	554,905	558,6563
2019	627	652,4457	647,8408

Примітка: в кожному ряді виділено краще прогнозне значення

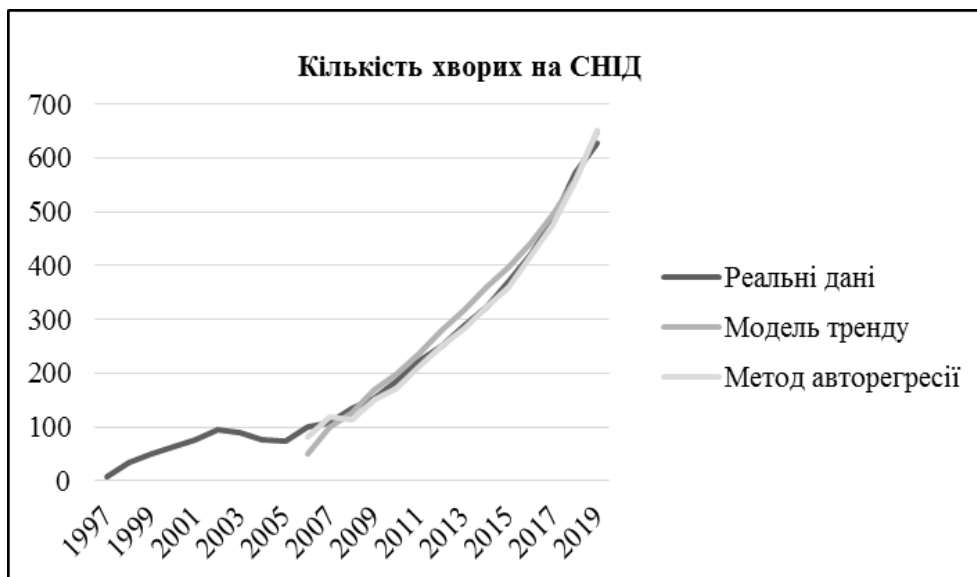


Рис. 1 - Результати прогнозування кількості хворих на СНІД різними методами



Для покращення ефективності прогнозу на наступному етапі було застосовано схему (1) у такому виді:

$$L(\alpha_1, \alpha_2) = \sum_{i=10}^m \left| \frac{\alpha_1 \varphi_i^{(1)} + \alpha_2 \varphi_i^{(2)} - v_i}{v_i} \right|, \quad (6)$$

де $\varphi_i^{(1)}$ – результат прогнозування i -го значення ряду методом авторегресії, $\varphi_i^{(2)}$ – результат прогнозування i -го значення поліноміальною моделлю тренду, $m = \overline{14, 22}$.

Значення α_1, α_2 на кожному кроці були знайдені за допомогою генетичного алгоритму для мінімізації функціоналу (6). Результати обчислень наведені в табл.3. Критеріями точності прогнозу були обрані відносна похибка та волатильність прогнозу у формі середньоквадратичного відхилення (4)-(5).

Таблиця 3 - Порівняння результатів прогнозування прогнозною схемою з іншими базовими моделями

Рік	Елементи часового ряду	Метод авторегресії	Модель тренду	Параметри моделі		Прогнозне значення
				α_1	α_2	
2010	185	170,5247	198,5224	0,95945054	0,052944097	174,1206
2011	221	210,8523	236,5665	1,171551501	-0,07885669	228,3695
2012	251	251,9838	280,3528	1,454850157	-0,36053295	265,5222
2013	288	282,4377	318,5382	1,48363889	-0,41159206	287,9278
2014	322	322,6126	358,7616	1,186310485	-0,14273517	331,5108
2015	370	359,3304	396,0834	1,344293004	-0,29879915	364,696
2016	421	417,1556	442,9315	1,045989452	-0,02676594	424,4849
2017	484	476,5829	496,1607	1,085873472	-0,06252759	486,485
2018	571	554,905	558,6563	1,036641888	-0,01913628	564,5472
2019	627	652,4457	647,8408	0,920765306	0,085053726	655,8506
Відносна похибка прогнозу		0,027136	0,102863			0,026486053
Середньоквадратичне відхилення		11,2493	12,9884			10,9777

Примітка: в кожному ряді виділено краще прогнозне значення

Обговорення результатів. Розроблений метод прогнозування часових рядів дозволяє покращити результати прогнозування за допомогою інших базових методів. Такий підхід, при розв'язанні реальних задач прогнозування часових рядів, дає можливість одночасно розглядати довільну кількість відомих методів прогнозування і розглядати результати їх роботи як окремо, так і в сукупності.

Експериментальна верифікація роботи розробленого методу показала, що при його вірному налаштуванні, вдається досягнути зменшення похибки прогнозування, а також покращення волатильності прогнозу, що дозволяє забезпечити зменшення відхилення прогнозного значення від точного.

Для розв'язання оптимізаційної задачі (2)-(3), крім генетичного алгоритму, можливим є також застосування інших еволюційних алгоритмів.

Висновки. Робота присвячена розробці еволюційного методу прогнозування часових рядів. В ході дослідження було побудовано математичну модель задачі прогнозування та розроблено метод прогнозування, який дозволяє обчислювати прогнозні значення, враховуючи результати прогнозування іншими методами. Результуючі значення, при цьому, обчислюються як згортка результатів, отриманих за допомогою обраних базових моделей, коефіцієнти якої обчислюються за допомогою генетичного алгоритму.



Виконано порівняльний аналіз результатів роботи розробленого методу та інших методів прогнозування для часового ряду «Кількість хворих на СНІД», який довів ефективність еволюційного методу прогнозування.

Список використаних джерел:

1. Khashei Mehdi, Mehdi Bijari. An artificial neural network (p, d, q) model for timeseries forecasting. *Expert Systems with applications*. 2010. 37.1. P. 479-489.
2. Khashei Mehdi, Mehdi Bijari. Which methodology is better for combining linear and nonlinear models for time series forecasting? 2012. P. 265-285.
3. Khashei Mehdi, and Mehdi Bijari. A novel hybridization of artificial neural networks and ARIMA models for time series forecasting. *Applied Soft Computing*. 2011. 11.2. P. 2664-2675.
4. Kuremoto Takashi, et al. Time series forecasting using a deep belief network with restricted Boltzmann machines. *Neurocomputing*. 2014. 137. P. 47-56.
5. Bontempi Gianluca, Souhaib Ben Taieb, Yann-Aël Le Borgne. Machine learning strategies for time series forecasting. *European business intelligence summer school*. Springer, Berlin, Heidelberg, 2012.
6. Borovykh Anastasia, Sander Bohte, Cornelis W. Oosterlee. Conditional time series forecasting with convolutional neural networks. *arXiv preprint arXiv*. 2017. 1703.04691.
7. Deb Chirag, et al. A review on time series forecasting techniques for building energy consumption. *Renewable and Sustainable Energy Reviews*. 2017. 74. P. 902-924.
8. Гече Ф., Мулеса О., Гече С., Вашкеба М. Розробка методу синтезу прогнозуючої схеми на основі базових прогнозуючих моделей. *Технологічний аудит та резерви виробництва*. 2015. Т. 3, N 2(23). С. 36-41. – Режим доступу : DOI : [10.15587/2312-8372.2015.44932](https://doi.org/10.15587/2312-8372.2015.44932).
9. Geche F., Batyuk A., Mulesa O., Vashkeba M. [Development of effective time series forecasting model](https://doi.org/10.1007/978-3-319-70581-1_26). *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*. 2015. Volume 4 Issue 12. P. 4377 – 4386.
10. Mulesa O., Geche F., Batyuk A., Buchok V. Development of Combined Information Technology for Time Series Prediction. In: *Shakhovska N., Stepashko V. (eds) Advances in Intelligent Systems and Computing II. CSIT 2017. Advances in Intelligent Systems and Computing*, vol 689. Springer, Cham. 2018. doi: https://doi.org/10.1007/978-3-319-70581-1_26
11. Кучанський О. Ю., Білощицький А. О. Прогнозування часових рядів методом селективного зіставлення зі зразком. *Восточно-Европейський журнал передових технологій*. 2015. 6 (4). С. 13-18.
12. Офіційний сайт Центру громадського здоров'я Міністерства охорони здоров'я України. http://old.phc.org.ua/pages/diseases/hiv_aids/statistics/hiv-aids-treatment

References:

1. Mehdi Khashei and Mehdi Bijari, "An artificial neural network (p, d, q) model for timeseries forecasting." in *Expert Systems with applications* 37.1 (2010), pp. 479-489.
2. Mehdi Khashei and Mehdi Bijari, "Which methodology is better for combining linear and nonlinear models for time series forecasting?" (2012), pp. 265-285.
3. Mehdi Khashei and Mehdi Bijari, "A novel hybridization of artificial neural networks and ARIMA models for time series forecasting." in *Applied Soft Computing* 11.2 (2011), pp. 2664-2675.
4. Takashi Kuremoto et al., "Time series forecasting using a deep belief network with restricted Boltzmann machines." in *Neurocomputing* 137 (2014), pp. 47-56.
5. Gianluca Bontempi et al., "Machine learning strategies for time series forecasting." in *European business intelligence summer school*. Springer, Berlin, Heidelberg, 2012.
6. A. Borovykh et al., "Conditional time series forecasting with convolutional neural networks." in *arXiv preprint arXiv:1703.04691* (2017).
7. Chirag Deb et al., "A review on time series forecasting techniques for building energy consumption." in *Renewable and Sustainable Energy Reviews* 74, pp. 902-924, 2017.
8. F. Geche et al., "Rozrobka metodu syntezu prohnozuiuchoi skhemy na osnovi bazovykh prohnozuiuchykh modelei" in *Tekhnolohichniy audyt ta rezervy vyrobnytstva*, N 2(23), pp. 36-41, 2015. DOI : [10.15587/2312-8372.2015.44932](https://doi.org/10.15587/2312-8372.2015.44932).
9. F. Geche F. et al., "[Development of effective time series forecasting model](https://doi.org/10.1007/978-3-319-70581-1_26)" in *International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)*, Volume 4 Issue 12, pp. 4377-4386, 2015.
10. O. Mulesa O. et al., "Development of Combined Information Technology for Time Series Prediction." In: *Shakhovska N., Stepashko V. (eds) Advances in Intelligent Systems and Computing II. CSIT 2017. Advances in Intelligent Systems and Computing*, vol 689. Springer, Cham., 2018. doi: https://doi.org/10.1007/978-3-319-70581-1_26
11. O. Kuchanskyi and A. Biloshchytskyi, "Prohnozuvannia chasovykh riadiv metodom selektyvnoho zistavlennia zi zrazkom" in *Vostochno-Evropeyskiy zhurnal peredovykh tekhnologiy*, 6 (4), pp. 13-18, 2015.
12. Офіційний сайт Центру громадського здоров'я Міністерства охорони здоров'я України. http://old.phc.org.ua/pages/diseases/hiv_aids/statistics/hiv-aids-treatment