

УДК 519.766.4

JEL Classification: F59

DOI: 10.15587/2312-8372.2019.188188

**РОЗРОБКА КОМБІНОВАНОГО МЕТОДУ ПРОРОКУВАННЯ ДИСКРЕТНИХ ЧАСОВИХ РЯДІВ З НЕСТАЦІОНАРНІСТЮ ДЛЯ ПРОГНОЗУВАННЯ ПОПИТУ ТОВАРІВ ВІЙСЬКОВОГО ПРИЗНАЧЕННЯ**

**Кубів С. І., Баланюк Ю. В.**

**РАЗРАБОТКА КОМБИНИРОВАННОГО МЕТОДА ПРЕДСКАЗАНИЯ ДИСКРЕТНЫХ ВРЕМЕННЫХ РЯДОВ С НЕСТАЦИОНАРНОСТЬЮ ДЛЯ ПРОГНОЗИРОВАНИЯ СПРОСА ТОВАРОВ ВОЕННОГО НАЗНАЧЕНИЯ**

**Кубив С. И., Баланюк Ю. В.**

**DEVELOPMENT OF A COMBINED METHOD FOR PREDICTING DISCRETE TIME SERIES WITH NON-STABILITY FOR FORECASTING MILITARY GOODS DEMAND**

**Kubiv S., Balanyuk Y.**

*Об'єктом дослідження є модель виробничої системи товарів військового призначення з нестационарними процесами. У дослідженні часових рядів характеристик виробничої системи різні конкуруючі моделі, як правило, отримуються у виробничих умовах із стохастичними даними стосовно виходу продукції, що пояснюється проблемами вузьких місць. Отже, вибір найкращої моделі, що описує виробничу систему, стає складним та критичним, оскільки деякі моделі, які найбільш точно відповідають спостережуваним даним, можуть не передбачити майбутні значення відповідно до складності моделі. Це дослідження прагне продемонструвати процедуру вибору моделі у системі з випадковими даними за допомогою скоригованих вагових коефіцієнтів. У даній роботі представлено метод поєднання двох наборів прогнозів. Отримані вимірювання служать вхідними даними до функції автокореляції та часткової функції автокореляції для отримання порядку прогнозуючих моделей. Параметри моделі оцінювали та використовували для прогнозування та порівнювали з вихідними та перетвореними даними для отримання суми квадратів помилок у (SSE). Потім моделі були піддані оцінці адекватності та згодом були протестовані за критеріями Akaike та Schwarz. Два окремі набори прогнозів даних часового ряду об'єднані для формування комбінованого набору прогнозів. Слід зазначити, що, коли кожен набір прогнозів містить деяку незалежну інформацію, комбіновані прогнози можуть дати покращення. Запропонований метод комбінування прогнозів дозволяє змінювати вагові коефіцієнти, що може призвести до кращих прогнозів. Основний висновок*

полягає в тому, що набір прогнозів може призвести до меншої середньоквадратичної помилки, ніж будь-який з початкових прогнозів. Минулі помилки кожного з початкових прогнозів використовуються для визначення ваги для приєднання цих двох оригінальних прогнозів при формуванні комбінованих прогнозів. Однак результативність прогнозу може змінюватися з часом.

**Ключові слова:** модель прогнозування, дискретний часовий ряд, випадкові вихідні дані, комбінований метод прогнозу.

Объектом исследования является модель производственной системы товаров военного назначения с нестационарными процессами. В исследовании временных рядов характеристик производственной системы различные конкурирующие модели, как правило, получаются в производственных условиях со стохастическими данными по выходу продукции, что объясняется проблемами узких мест. Итак, выбор лучшей модели, описывающей производственную систему, становится сложным и критическим, поскольку некоторые модели, которые наиболее точно соответствуют наблюдаемым данным, могут не предусмотреть будущие значения в соответствии со сложностью модели. Это исследование стремится продемонстрировать процедуру выбора модели в системе со случайными данными с помощью скорректированных весовых коэффициентов. В данной работе представлен метод сочетания двух наборов прогнозов. Полученные измерения служат входными данными с функцией автокорреляции и частичной функцией автокорреляции для получения порядка прогнозирующих моделей. Параметры модели оценивали и использовали для прогнозирования и сравнивали с исходными и преобразованными данными для получения суммы квадратов ошибок в (SSE). Затем модели были подвергнуты оценке адекватности и впоследствии были протестированы по критериям Akaike и Schwarz. Два отдельных набора прогнозов данных временного ряда объединены для формирования комбинированного набора прогнозов. Следует отметить, что, когда каждый набор прогнозов содержит некоторую независимую информацию, комбинированные прогнозы могут дать улучшение. Предложенный метод комбинирования прогнозов позволяет изменять весовые коэффициенты, что может привести к улучшению прогнозов. Основной вывод заключается в том, что набор прогнозов может привести к меньшей среднеквадратичной ошибке, чем любой из начальных прогнозов. Прошлые ошибки каждого из начальных прогнозов используются для определения веса для присоединения двух оригинальных прогнозов при формировании комбинированных прогнозов. Однако результативность прогноза может меняться со временем.

**Ключевые слова:** модель прогнозирования, дискретный временной ряд, случайные выходные данные, комбинированный метод прогноза.

## 1. Вступ

Процеси планування виробництва товарів військового призначення залежать від прогнозу попиту. Якість планування залежить від точності цього

прогнозу. Цей взаємозв'язок добре задокументований [1–3]. Специфіка продукції цієї галузі обумовила необхідність розробки відповідних підходів в прогнозуванні, оскільки на нього впливає низка чинників. Виробники озброєння усвідомлюють значні ризики, вони потребують достовірних методів прогнозування попиту на свою продукцію, які були б працездатними на ранніх стадіях життєвого циклу виробів [4, 5].

Вибір моделі прогнозування залучив чимало уваги протягом останніх 30 років. Представлено багато моделей на вибір – різні форми експоненціального згладжування, моделі авторегресії та інтегрованого ковзного середнього (ARIMA) нейронні мережі сітки тощо – і на практиці доводиться вибирати, яку саме використовувати [6].

У багатьох теоретичних та практичних дослідженнях вивчалися різні методи виявлення найкращої моделі. Якщо можна визначити найбільш відповідну модель для кожного часового ряду, точність прогнозування може бути значно покращена, як правило, на 25–30 % [7].

Взагалі програмне забезпечення для прогнозування рекомендує або вибирає модель на основі статистичного алгоритму. Ефективність моделей-кандидатів оцінюється або за даними вибірки, зазвичай за допомогою відповідних інформаційних критеріїв, або шляхом вилучення набору точок даних для створення вибірки перевірки (оцінка поза вибіркою, [8]), також відомий як помилка перехресної перевірки). Однак легко розробити приклади, коли вибір статистичної моделі (заснований на оцінці у вибірці чи поза вибіркою) не вдається. Такі випадки поширені в реальних програмах прогнозування і тому робить вибір моделі прогнозування нетривіальним завданням на практиці.

Отже, вибір моделей відповідних моделей прогнозування є актуальною проблемою. Таким чином, *об'єктом дослідження* є модель виробничої системи товарів військового призначення з нестационарними процесами. *А мета роботи* полягає в підвищенні ефективності прогнозування поведінки виробничої системи шляхом розробки комбінованого методу прогнозу.

## **2. Методика проведення дослідження**

Практикуючі можуть застосовувати судження до різних завдань у процесі прогнозування, а саме:

- 1) визначення набору моделей-кандидатів;
- 2) вибір моделі;
- 3) параметризація моделей;
- 4) складання прогнозів;
- 5) уточнення/коригування прогнозу.

Вельми цікавим є випадок, коли два (або більше) прогнозів зроблені на одну і ту ж подію. Як правило, реакція більшості статистиків та бізнесменів, коли це відбувається, – це спроба виявити, який є кращим (або найкращим) прогнозом; тоді кращий прогноз приймається та використовується, а другий відкидається. Хоча це може мати певні заслуги, коли аналіз є основною метою

вправи, це не мудра процедура, якщо мета полягає в тому, щоб зробити максимально якісний прогноз, оскільки відкинутий прогноз майже завжди містить корисну незалежну інформацію.

Ця незалежна інформація може бути двох видів:

1. Один прогноз базується на змінних чи інформації, які не враховуються при іншому прогнозі.

2. При прогнозуванні передбачаються інші припущення щодо форми співвідношення змінних.

Зокрема, другий випадок не обов'язково призводить до ситуації, коли комбінований прогноз покращується при покращенні індивідуального прогнозу, хоча це цілком можливо. Слід зазначити, що вводиться одна важлива умова щодо характеру окремих прогнозів, а саме – незміщеність. Сукупність прогнозів, за якими істинні значення послідовно переоцінюються, у поєднанні з набором незміщених прогнозів призведе до зміщених прогнозів; швидше за все, комбіновані прогнози даватимуть додаткові помилки зміщення взамін об'єктивних прогнозів. Першим кроком є перевірка того, що окремі набори прогнозів є незміщеними; чи можлива корекція при наявності зміщення оцінок.

Перш ніж обговорити різні способи комбінування прогнозів, дамо емпіричне обґрунтування, склавши первинну комбінацію двох прогнозів. Вибрані прогнози стосуються даних дискретних часових рядів, для яких щомісячно робляться прогнози на один період вперед [9]. Прогнози, опубліковані у роботі [10], підтверджують, що методи прогнозування, розроблені у дослідженнях, виявилися настільки успішними, що необхідно шукати процеси, для яких можна знайти альтернативні методи, який прогноз краще.

Існує багато альтернатив експоненціального згладжування для складання бізнес-прогнозів, таких як нейронні мережі та інші методи машинного навчання. Тим не менш, методи комбінованого прогнозу часових рядів залишаються дуже привабливими.

Проілюстрована комбінація – середнє арифметичне для двох індивідуальних прогнозів. У табл. 1 наведено деталі для пояснення.

**Таблиця 1**

Помилки пророкування

Місяць	Модель		
	Експоненціальне згладжування	Бокса-Дженкінса	Комбінована модель
1	2	3	4
Січень	1	-3	-1
Лютий	6	-10	-2
Березень	18	24	21
Квітень	18	22	20
Травень	3	-9	-3
Червень	-17	-22	-19,5

## Продовження таблиці 1

1	2	3	4
Липень	-24	10	-7
Серпень	-16	2	-7
Вересень	-12	-11	-11,5
Жовтень	-9	-10	-9,5
Листопад	-12	-12	-12
Грудень	-13	-7	-10
Дисперсія помилок	196	188	150

Перерахування цих та інших прогнозів даних робиться на більш пізньому етапі. На даний момент можна просто зазначити, що різниця помилок у трьох прогнозах становила 177,7 [6], 148,6 [1] та 130,2 (комбінація з рівними вагами до кожного окремого прогнозу). Таким чином, прогнози, очевидно, мали деяку цінність.

Для раціонального вибору параметрів методу комбінованого прогнозу треба, перш за все, обрати та обґрунтувати метод визначення вагових коефіцієнтів.

### 3. Результати дослідження та обговорення

Хоча комбінований прогноз, сформований шляхом надання однакової ваги кожному з окремих прогнозів, є прийнятним для ілюстративних цілей, однак при накопиченні даних більша вага надається набору прогнозів, які містять менші (середньоквадратичні) помилки. Проблема полягає у виборі найкращого підходу до зважування набору прогнозів. Існує багато способів визначення цих ваг, і метою було обрати метод, який, ймовірно, дасть низькі помилки для комбінованих прогнозів. Передбачається, що результативність окремих прогнозів буде узгодженою з часом у тому сенсі, що дисперсії помилок для двох прогнозів, позначені  $\sigma_1^2$  та  $\sigma_2^2$ , можна мінімізувати зі зростанням інтервалу спостереження для всіх значень часу  $t$ . Крім того, передбачалося, що обидва прогнози будуть незсунутими (природним шляхом або після внесення корекції). Комбінований прогноз отримується шляхом лінійного комбінування двох наборів прогнозів, з ваговим коефіцієнтом  $0 \leq k \leq 1$  для першого набору прогнозів і ваговим коефіцієнтом  $(1-k)$  для другого набору, завдяки чому комбінований прогноз стає незсунутим. Варіант помилок у комбінованому прогнозі  $\sigma_c^2$  може бути записаний:

$$\sigma_c^2 = k^2 \sigma_1^2 + (1-k)^2 \sigma_2^2 + 2\rho k \sigma_1 (1-k) \sigma_2, \quad (1)$$

де  $k$  – ваговий коефіцієнт, що дається першому набору прогнозів, а  $\rho$  – коефіцієнт кореляції між помилками у першому наборі прогнозів та помилками у другому наборі. Вибір  $k$  слід зробити так, щоб похибки комбінованих прогнозів були найменшими, для чого треба знайти мінімум загальної дисперсії

$\sigma_c^2$ . Диференціюючи по  $k$  і дорівнюючи похідну нулю, отримуємо мінімум  $\sigma_c^2$ , що має місце за наступної умови:

$$k = \frac{\sigma_2^2 - \rho\sigma_1\sigma_2}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \quad (2)$$

При  $\rho=0$  вираз (2) зводиться до:

$$k = \frac{\sigma_2^2}{\sigma_1^2 + \sigma_2^2}. \quad (3)$$

Можна показати, що якщо  $k$  визначається рівнянням (2), значення  $\sigma_c^2$  не перевищує меншу з дисперсій  $[\sigma_1^2, \sigma_2^2]$ . Щоб довести це, розглянемо співвідношення між дисперсіями  $[\sigma_1^2, \sigma_2^2]$  при класичному проорокуванні з ваговими коефіцієнтами  $[k, (1-k)]$  та дисперсією  $\sigma_c^2$  при комбінованому проорокуванні. За умов стаціонарності та незміщеності оцінок дисперсія помилки проорокування  $\sigma_c^2$  дорівнює [1]:

$$\sigma_c^2 = \frac{\sigma_1^2\sigma_2^2(1-\rho^2)}{\sigma_1^2 + \sigma_2^2 - 2\rho\sigma_1\sigma_2}. \quad (4)$$

Тоді

$$\begin{cases} \sigma_c^2 - \sigma_1^2 = \frac{\sigma_1^2(\sigma_1 - \rho\sigma_2)^2}{(\sigma_1 - \rho\sigma_2)^2 + \sigma_2^2(1-\rho^2)} \leq 0; \\ \sigma_c^2 - \sigma_2^2 = \frac{\sigma_2^2(\sigma_2 - \rho\sigma_1)^2}{(\sigma_2 - \rho\sigma_1)^2 + \sigma_1^2(1-\rho^2)} \leq 0. \end{cases} \quad (5)$$

Значить  $\sigma_c^2 \leq \min \overline{\sigma_1^2, \sigma_2^2}$ .

Таким чином, комбінований прогност дає поліпшення у всіх випадках, за винятком того випадку, коли модуль різниці дисперсій окремих прогностів є величиною другого порядку малості, а коефіцієнт кореляції  $\rho$  близький до одиниці. Строго кажучи, це може бути цілком звичайним явищем.

До початку процесу комбінування прогностів ( $t=0$ ) оптимальне значення  $k$  невідоме. Величина вагового коефіцієнту  $k$  змінюється у міру накопичення поточної інформації щодо відносної ефективності двох початкових прогностів. Таким чином, комбінований прогност  $S_T$  для періоду спостереження  $T$  точніше

записується як:

$$C_T = k_T \phi_{1T} + (1 - k) \phi_{2T}, \quad (6)$$

де  $[\phi_{1T}, \phi_{2T}]$  – прогноз за період  $T$  від першого та другого набору даних, відповідно.

Рівняння (2) і (3) використовуються в якості основи для розробки тих чи інших методів прогнозу. Однак результативність якогось з методів прогнозу може змінюватися з часом (наприклад, покращуватися). Значить, метод, заснований на оцінці дисперсії помилок на усьому інтервалі спостереження від початку пророкування, може стати таким чином не оптимальним. Треба будувати методи, які надають більше ваги останнім помилкам, ніж тим, що зафіксовані раніше.

#### 4. Висновки

Оптимальна попередня ідентифікація найкращої моделі апостеріорі може принести значні переваги щодо прогнозування дискретних, в тому числі нестационарних часових рядів. Наразі дослідження зосереджені на автоматичному та статистичному підходах до вибору моделі.

Однак користувачі методів прогнозування часто вимагають системні рекомендації та застосування суджень при виборі моделі прогнозування.

У даній роботі представлено метод поєднання двох наборів прогнозів. Слід зазначити, що, коли кожен набір прогнозів містить деяку незалежну інформацію, комбіновані прогнози можуть дати покращення. Один із важливий висновок полягає в тому, що запропоновані методи комбінування прогнозів дозволяють змінювати вагові коефіцієнти, що часто може призвести до більш кращих прогнозів. Запропонований метод відрізняється цим від тих методів, які пропонують застосування постійних вагових коефіцієнтів, визначених після отримання всіх окремих помилок прогнозу. Хоча коментарі у цій роботі стосуються виключно поєднання двох прогнозів, є всі підстави поєднати більше двох прогнозів (там, де вони існують).

Підсумовуючи можна сказати, що може бути певна користь у порівнянні окремих прогнозів із комбінованим прогнозом. Комбінований прогноз зі значно меншою дисперсією помилок, ніж будь-який окремий прогноз, передбачає, що моделі, що використовуються для окремих прогнозів, здатні до деякого вдосконалення. Таким чином, можна отримати напрями пошуку більш вірної моделі, яка мала б аналітичну цінність. Слід зазначити однак, що ця модель є адекватною та досить точною, але вона може бути двох видів. Це може бути принципово інша модель, що включає нову змінну або інше трактування змінної, або це може бути просто підтвердженням ості параметрів моделі. Проблему уточнення цих нестационарних параметрів планується розглянути у майбутньому.

## Література

1. Box, G. E. P., Jenkins, G. M., Siegel, A. F. (2016). *Time series analysis: forecasting and control*. London: Elsevier Inc., 622.
2. Petropoulos, F., Makridakis, S., Assimakopoulos, V., Nikolopoulos, K. (2014). “Horses for Courses” in demand forecasting. *European Journal of Operational Research*, 237 (1), 152–163. doi: <http://doi.org/10.1016/j.ejor.2014.02.036>
3. Priestley, M. B., Tong, H. (1973). On the Analysis of Bivariate Non-Stationary Processes. *Journal of the Royal Statistical Society: Series B (Methodological)*, 35 (2), 153–166. doi: <http://doi.org/10.1111/j.2517-6161.1973.tb00949.x>
4. Petropoulos, F., Goodwin, P., Fildes, R. (2017). Using a rolling training approach to improve judgmental extrapolations elicited from forecasters with technical knowledge. *International Journal of Forecasting*, 33 (1), 314–324. doi: <http://doi.org/10.1016/j.ijforecast.2015.12.006>
5. Seifert, M., Siemsen, E., Hadida, A. L., Eisingerich, A. B. (2015). Effective judgmental forecasting in the context of fashion products\*. *Journal of Operations Management*, 36 (1), 33–45. doi: <http://doi.org/10.1016/j.jom.2015.02.001>
6. Kubiv, S. (2019). Approximations and forecasting quasi-stationary processes with sudden runs. *Technology Audit and Production Reserves*, 4 (4 (48)), 37–39. doi: <http://doi.org/10.15587/2312-8372.2019.179265>
7. Fildes, R., Petropoulos, F. (2015). Simple versus complex selection rules for forecasting many time series. *Journal of Business Research*, 68 (8), 1692–1701. doi: <http://doi.org/10.1016/j.jbusres.2015.03.028>
8. Verbeek, M. (2017). *A Guide to Modern Econometrics*. Hoboken: John Wiley & Sons, Inc., 508.
9. Greene, W. H. (2012). *Econometric analysis*. Upper Saddle River: Pearson Education, Inc., 1189.
10. Barnard, G. A. (1963). New Methods of Quality Control. *Journal of the Royal Statistical Society. Series A (General)*, 126 (2), 255–259. doi: <http://doi.org/10.2307/2982365>

*The object of research is a model of the production system of military goods with non-stationary processes. In the study of the time series of the characteristics of the production system, various competing models, as a rule, are obtained under production conditions with stochastic data on the output of products due to bottleneck problems. So, the choice of the best model that describes the production system becomes difficult and critical, because some models that most closely correspond to the observed data may not foresee future values in accordance with the complexity of the model. This study seeks to demonstrate the procedure for selecting a model in a random data system using adjusted weights. This paper presents a method for combining two sets of forecasts. The obtained measurements serve as input with an autocorrelation function and a partial autocorrelation function to obtain the order of predictive models. The model parameters are evaluated and used for forecasting and compared with the original and converted data to obtain the sum*

*of squared errors in (SSE). Models are evaluated for adequacy and subsequently tested against Akaike and Schwarz criteria. Two separate sets of forecasts of time series data are combined to form a combined set of forecasts. It should be noted that when each set of forecasts contains some independent information, combined forecasts can provide an improvement. The proposed method for combining forecasts allows to change weights, can lead to better forecasts. The main conclusion is that a set of forecasts can lead to a lower standard error than any of the initial forecasts. Past errors of each of the initial forecasts are used to determine the weight for joining two original forecasts in the formation of combined forecasts. However, the effectiveness of the forecast may change over time.*

**Keywords:** *forecasting model, discrete time series, random output data, combined forecasting method.*