PHYSICS AND MATHEMATICS

РЕАЛІЗАЦІЯ МЕТОДИКИ ПОБУДОВИ МОДЕЛЕЙ НЕЛІНІЙНИХ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ В ЕКОНОМІЦІ ТА ФІНАНСАХ ТА ЗАСТОСУВАННЯ МЕРЕЖІ БАЙЄСА ДЛЯ ПОКРАЩЕННЯ РІВНЯ АДЕКВАТНОСТІ МОДЕЛЕЙ ТА ЯКОСТІ ОЦІНОК КОРОТКОСТРОКОВИХ ПРОГНОЗІВ

***Дудка Б. Р.**, магістр*

Україна, Київ, Інститут прикладного системного аналізу НТУУ «КПИ»

DOI: https://doi.org/ 10.31435/rsglobal_ws/30082018/6046

ARTICLE INFO

Received: 24 July 2018
Accepted: 25 August 2018
Published: 30 August 2018

KEYWORDS

Bayes networks,
estimation of forecasts,
nonlinear process,
regression models.

ABSTRACT

In this article was represented implementation of methodology of nonlinear non stationary processes modeling in economy and finances which based on different conceptions of system analysis, regression analysis and applying of bayesian network to improve estimation of short term forecasts.

Citation: Dudka B. R. (2018) Realizatsiia Metodyky Pobudovy Modelei Neliniinykh Nestatsionarnykh Protsesiv V Ekonomitsi Ta Finansakh Ta Zastosuvannia Merezhi Baiiesa Dlia Pokrashchennia Rivnia Adekvatnosti Modelei Ta Yakosti Otsinok Korotkostrokovykh Prohnoziv. *World Science.* 8(36), Vol.1. doi: 10.31435/rsglobal_ws/30082018/6046

Вступ. Економічне прогнозування є одним з найважливіших факторів успішного розвитку усіх держав світу. Саме від визначення раціональних прогностичних аспектів залежить вдалий соціально-економічний розвиток будь-якого суспільства.

Економічні та фінансові системи, що вивчаються сучасною наукою, з великими труднощами піддаються дослідженню звичайними (вербальними) теоретичними методами. Прямий експеримент над ними неможливий. Ціна помилок і прорахунків велика, тому математичне моделювання досліджуваних процесів є неминучою складовою науково-технічного прогресу.

На сьогодні більшість процесів в економіці та фінансах є нелінійними та нестаціонарними. Такі процеси характеризуються значною кількістю складностей та особливостей, що необхідно враховувати, при моделюванні та прогнозуванні відповідних процесів. Такого роду процеси містять тренд або змінну дисперсію. Під трендом будемо розуміти загальну тенденцію при різнонаправленому русі, яка визначена загальною спрямованістю змін показників часового ряду. Виділяють два типи тренду: детермінований та стохастичний. Процеси з трендами та змінною дисперсією особливо характерні для фінансово-економічних процесів, які будуть використані у якості об’єкту дослідження.

Наступною проблемою, при побудові фінансово-економічних процесів є наявність в них нелінійностей. Нелінійність означає наявність непередбачуваних змін у напрямі розвитку процесів. Вона може проявлятись як підвищеною реакцію на зміну одних факторів, так і повною нечутливістю до інших.

Перша книга, в якій були частково описані принципи та етапи побудови моделей нестаціонарних процесів, була опублікована у кінці 1986 року [Дрейпер Н., Смит Г. Прикладной регрессионный анализ (т.2). - М.: Финансы и статистика, 1986. - 366 с.]. Потім була опублікована книга [Бідюк, П. І. Аналіз часових рядів. Навчальний посібник [Текст] / П. І. Бідюк, В. Д. Романенко, О.Л. Тимощук.- К.: Політехніка, 2010. -317 с] у якій була наведена методика побудови моделей нелінійних нестаціонарних процесів буда доповнена декількома етапами, які полягали у застосування комбінованих методів регресійного та статистичного аналізу, також була наведена прогнозуюча функція для деяких лінійних моделей , застосування якої дає змогу покращити якість точкового прогнозу. Статей, в яких були б подальші дослідження щодо модифікації сформованої методики за допомогою методів інтелектуального аналізу даних для покращення точкових оцінок прогнозу, отриманих шляхом використання моделей процесів, які були побудовані з використання даної методики та формування ймовірнісних оцінок прогнозу не було знайдено. Внаслідок цього, можна зробити висновок, що запропонована тематика є доволі новою.

Беручи до уваги вищевказане, можна зробити висновок, що доповнення методики побудови нелінійних нестаціонарних процесів методами інтелектуального аналізу даних дало б змогу розширити спектр економічних та фінансових процесів, які підлягають моделюванню та підвищити адекватність побудованих моделей.

Дослідження присвячене аналізу використання мережі Байєса для покращення точкових оцінок прогнозів моделей, які побудовані з використанням даної методики.

Постановка задачі.

Мета роботи: реалізувати методику побудови моделей нелінійних нестаціонарних процесів та проаналізувати ефективність застосування мережі Байєса для покращення оцінок прогнозів моделі, яка була побудована з використанням даної методики.

Методика побудови моделей часових рядів

На рисунку 1 зображена методика побудови моделей часових рядів, яка буде використана для побудови моделей досліджуваного процесу.

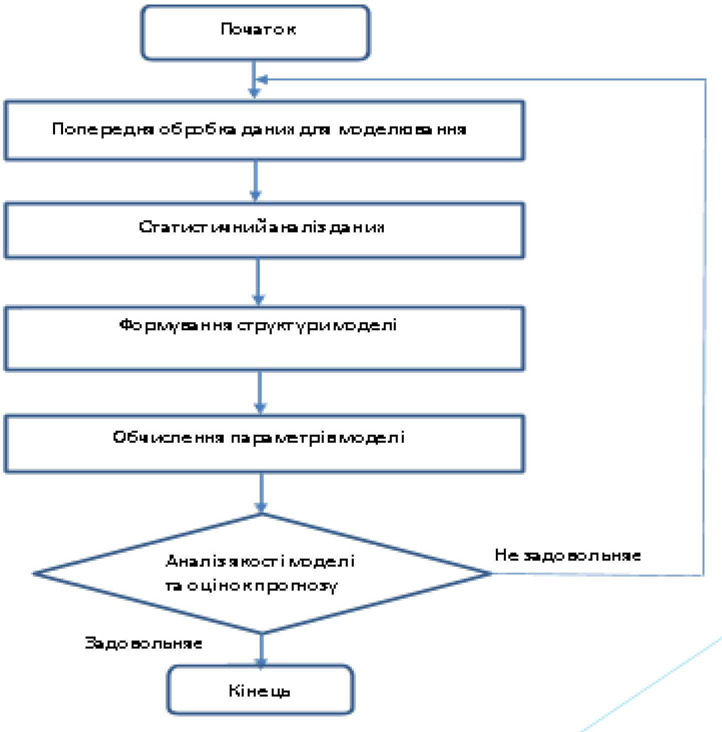

Рис. 1. Методика побудови моделей часових рядів

Детальний процесс побудови методики , яка забражена на рисунку 1, можна знайти у роботі [5]. Детальний огляд методів статистичного аналізу, методів формування структури моделей часових рядів, методів попередньої обробки даних та аналізу якості моделі та оцінок прогнозів можна знайти у роботах [1-4].

Методика побудови мережі Байєса для прогнозування

На рисунку 2 зображена методика побудови мережі Байєса для прогнозування.

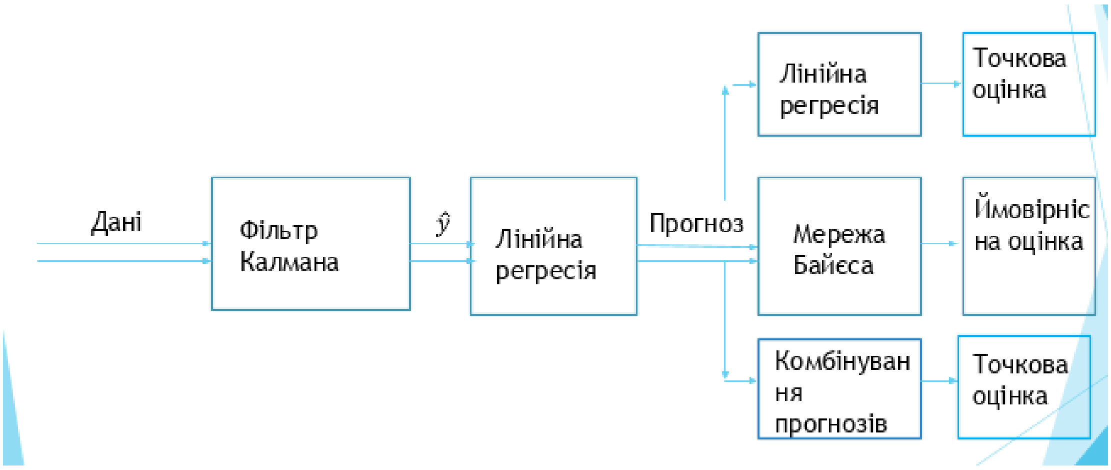

Рис. 2. Методика побудови мережі Байєса для прогнозування

Застосування мережі Байєса для покращення результатів прогнозування

У якості об’єкта дослідження було обрано часовий ряд значення якого значення курсу акцій компанії “Enterprise solutions”. Потужність часового ряду становить 429 значень.

До даного часового ряду був застосований фільтр Калмана для попередньої обробки даних.

Далі будемо будувати наступні типи моделей: ARIMA(p,i,q), ARMA(p,q) та AR(p) та визначемо найкращу модель враховуючи такі критерії адекватності моделі як: коефіцієнт детермінації, скоригований коефіцієнт детермінації, сума квадратів залишків. Для виконання обчислень будемо використовувати статистичний пакет Eviews. На рисунку 3 наведені статистичні характеристики побудованих моделей.

Типи моделей	Адекватність			Якість прогнозу		
	R^2	∑e^2	DW	RMSE	MAPE	V
Модель автороегресії(AR)						
AR(1)	0,9802	73,4998	2,2423	0,5987	2,3520	0,0154
AR(2)	0,9804	72,2074	2,0195	2,4185	2,4801	0,0636
AR(3)	0,9808	69,7785	1,9929	0,5877	2,3519	0,0152
AR(4)	0,9822	62,8651	2,0114	0,5663	2,2958	0,0147
Модель автороегресії з ковзним середнім (ARMA)						
ARMA(1,1)	0,9805	72,1677	1,9747	0,5933	2,3714	0,0153
ARMA(1,3)	0,9813	69,2111	1,9810	0,5810	2,3553	0,0150
ARMA(1,4)	0,9813	68,8663	1,9987	1,9514	2,4688	0,0481

Рис. 3. Статистичні характеристики побудованих моделей

Модель ARMA(1,3) є більш точною, порівнюючи характеристики прогнозу та адекватності моделі, у порівняння з моделями ARMA(1,1), AR(3), AR(2) та AR(1), але , у порівнянні з моделю AR(4), має менший ступінь адекватності. Тобто найкращою за характеристиками прогнозу та адекватності серед побудованих моделей є модель AR(4). Відповідну модель будемо використовувати для побудови ймовірнісної оцінки прогнозу за допомогою мережі Байєса.

Для побудови мережі Байєса був знайдений відповідний прогноз за вибраною моделлю авторегресії четвертого порядку AR(4).

Відповідно значення прогнозу ми будемо використовувати у якості залежної змінної у мережі Байєса. Опишемо структуру мережі.

Мережа буде складатися з 6 вершин, які будуть являти собою відповідні ендогенні та екзогенні змінні, які відповідають структурі вибраної для побудови прогнозу моделі AR(4). У таблиці 1 представлені назви відповідних вершин мережі Байєса та їх опис.

Таблиця 1. Назви відповідних вершин мережі Байєса та їх опис

Назва вершини	Опис
Lag1	Відповідає значенню змінної ar(1)
Lag2	Відповідає значенню змінної ar(2)
Lag4	Відповідає значенню змінної ar(4)
Lag10	Відповідає значенню змінної ar(10)
Curr	Відповідає значенню змінної y(k)
Predict	Відповідає значенню змінної y(k+1)

У якості вхідних параметрів представлені відповідні змінні, які ми отримали за вибраної регресійної моделі. Також для навчання та перевірки роботи мережі вхідна вибірка була розділена на тестову на перевірочну у наступному співвідношенні – 60/40. Для методу навчання мережі був обраний Наївний баєсів класифікатор. У якості залежної змінної було обрано зміну Predict, яка відповідає набору значень прогнозу, які були отримані за допомогою обраної моделі авторегресії AR(4). На рисунку 4 зображена структура побудованої мережі.

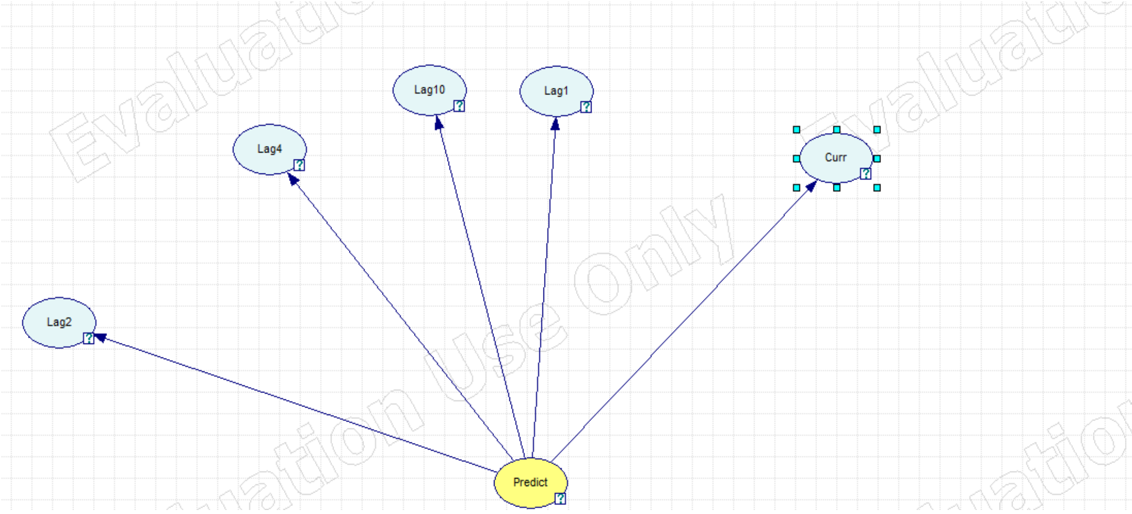

Рис. 4. Структура побудованої мережі

Побудована мережа має 6 вершин на 40 станів на які розбиті усі вхідні дані внаслідок процедури дискретизації.

Побудована мережа працює наступним чином:

а) на залежних вершинах вибирається відповідний стан;

б) перераховуються відповідній ймовірності;

в) відображаються відповідні ймовірності настання того чи іншого стану у залежній змінній.

Наведемо приклад роботи побудованої мережі

У таблиці 2 наведені значення регресорів(незалежних вершин мережі) для побудови мережі Байєса.

Таблиця 2. Значення регресорів (незалежних вершин мережі)

Назва вершини	Значення
Lag1	22,94
Lag2	23,91
Lag4	21,66
Lag10	25,05
Curr	22,15

Тоді стани цих вершин, які відповідають вищевказаним значенням будуть наступні: у таблиці 3 наведені назви вершин мережі та їх відповідні стани. На рисунку 5 наведений приклад роботи побудованої мережі Байєса.

Таблиця 3. Назви вершин мережі та їх відповідні стани

Назва вершини	Стан
Lag1	S29
Lag2	S33
Lag4	S25
Lag10	S38
Curr	S28

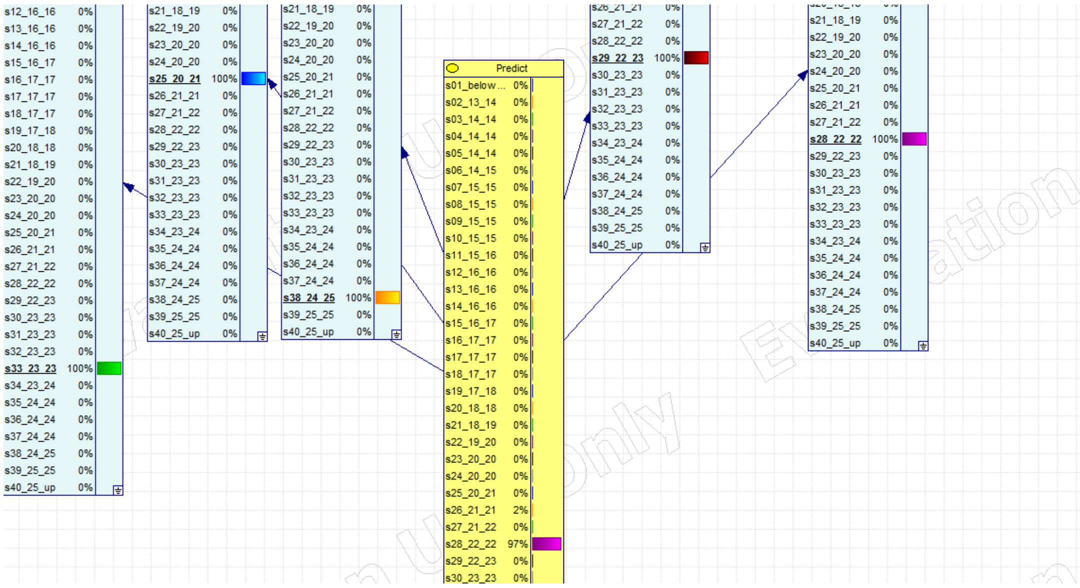

Рис. 5. Приклад роботи побудованої мережі Байєса

На рисунку 6 можна побачити, що залежна змінна Pred приймає значення з інтервалу s28 з ймовірністю 97%. Інтервал s28 містить наступні значення [22,10773; 22,205]. Для обрахунку прогнозного значення, враховуючи значення прогнозу за моделлю AR(4) та побудованою мережею Байєса, будемо використовувати наступну формулу:

$$Pred_{New} = \frac{Forecasted\_networl\_value + Forecasted\_model\_value}{2}, \qquad (1)$$

де $Forecasted\_network\_value$ – відповідне значення прогнозу за мережею Байєса; $Forecasted\_model\_value$ – відповідне значення прогнозу за прогнозуючою моделлю. $Forecasted\_network\_value$ – може бути обраховане як середина інтервалу, якому відповідає найбільша ймовірність, тобто

$$Forecasted\_network\_value = \frac{22{,}107 + 22{,}205}{2} = 22{,}1965. \qquad (2)$$

Знайдемо значення $Pred_{New}$ за нашим прикладом:

$$Pred_{New} = \frac{22{,}107 + 22{,}205}{2} = 22{,}1965 \qquad (3)$$

Порівнюючи значення прогнозу за мережею та за моделлю AR(4) з істинним значенням, тобто:

$Forecasted\_network\_value = 22{,}1965;$

$Forecasted\_model\_value = 22{,}237;$

$True\_value = 21{,}93$

можна зробити висновок, що прогноз за мережею Байєса кращий на 0,04. Тобто мережа, побудована на основі моделі AR(4), дає покращений результат прогнозу за моделлю.

Знайдемо похибку прогнозу у відсотках з наступною формулою

$$\text{APE} = \frac{|Pred_{new} - True\_value|}{True\_value} \times 100\% \qquad (4)$$

Враховуючи значення змінних, які вказані вище, можна знайти чому дорівнює помилка відповідного прогнозу

$$\text{APE} = \frac{|22{,}1965 - 22{,}15|}{22{,}15} \times 100\% = 0.2\% \qquad (5)$$

Значення помилки досить незначне, що дає прогноз є досить точним.

Аналіз результатів. На тестових даних були знайдені значення прогнозу за моделлю AR(4), за побудованою мережею Байєса та порівняні з очікуваним значенням. На рисунку 6 зображені графіки значень прогнозу за моделлю, мережею Байєса та істинних значень.

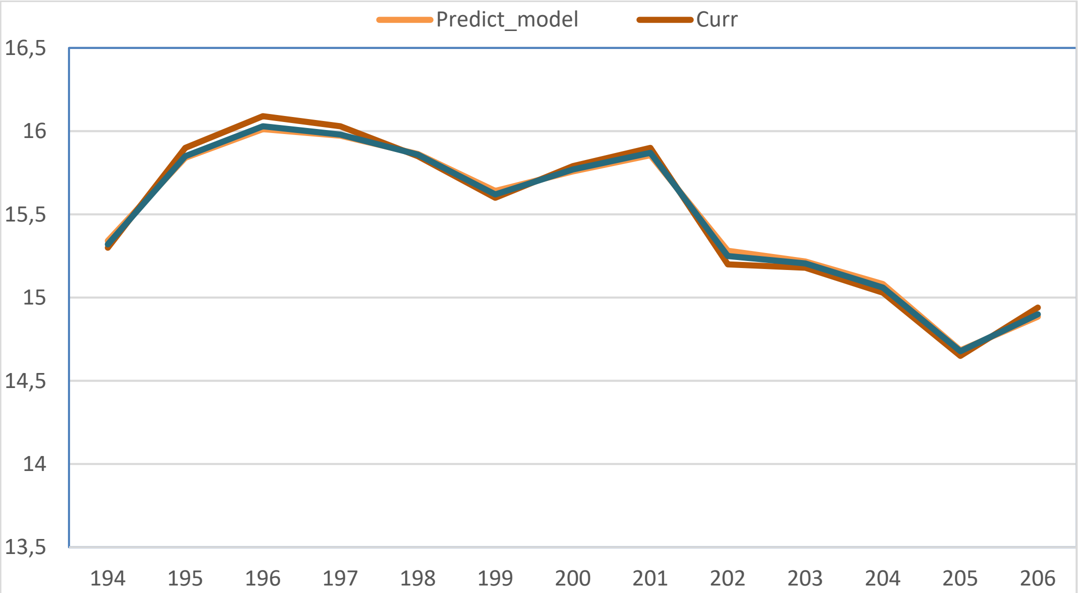

Рис. 6. Графіки значень прогнозу за моделлю, мережею Байєса та істинних значень

З вищевказаних графіків можна зробити висновок, що значення прогнозу, які знайдені за допомогою побудованої моделі авторегресії 4-го порядку є досить близькими до істинних значень. Також можна побачити, що використання мережі Байєса сприяє уточненню значень прогнозу. У таблиці 4 представлені статистичні характеристики моделі авторегресії та моделі, яка побудована за допомогою мережі Байєса.

Таблиця 4 – Статистичні характеристики моделей авторегресії та мережі Байєса.

Статистичні характеристики / Назва моделі	RMSE	MAPE
Авторегресія AR(4)	0,051044	0,41558
Мережа Байєса	0,036102	0,26820

Висновки. Таким чином, реалізована методика моделювання забезпечує отримання адекватних моделей за умови відповідності даних вимогам інформативності та повноти. Застосування мережі Байєса як методу інтелектуального аналізу дає можливість підвищити рівень адекватності моделей та отримати оцінки короткострокового прогнозу більш високої точності. Мережа Байєса дає змогу отримати ймовірнісну оцінку короткострокових прогнозів.

Розглянуту методику моделювання доцільно доповнити тестами на аналіз нелінійностей та типу не стаціонарності досліджуваних процесів. Доцільно увести комбінований критерій аналізу адекватності моделей для реалізації автоматизованого режиму вибору кращої моделі. Доповнити наведену методику моделювання альтернативними методами заповнення пропусків даних і методами оцінювання параметрів нелінійних моделей. Реалізувати автоматичне формування комбінованого прогнозу, отриманого за допомогою використання мережі Байєса та регресійної моделі.

ЛІТЕРАТУРА

1. Демківський Є. О., Бідюк П. І. Система підтримки прийняття рішень при прогнозуванні нестаціонарних процесів // Наукові праці Миколаївського державного гуманітарного університету ім. Петра Могили, 2008, Вип. 77, с. 137-159.
2. Згуровский М. З., Бидюк П. И., Терентьев А. Н. Методы построения байесовских сетей на основе оценочных функций // Кибернетика и системный анализ, 2008, № 2, с. 81-88.
3. Статистика: Підручник/ А. В. Головач, А. М. Єріна, В. О. Козирєв та ін.; За ред. А. В. Головача, А.М. Єріної, О.В. Козирєва.: – К.:Вища школа, 1993. – 623 с.
4. Box, G. E. P. and Jenkins, G. M. (1976). Time Series Analysis: Forecasting and Control, Revised Edition, Holden-Day, San Francisco.
5. Дудка Б. Р. Реалізація методики побудови моделей часових рядів // Київ: ІПСА НТУУ «КПІ», «Системні науки та кібернетика», 2016.