

УДК 519.226, 330.322

О.М. Терентьев, П.І. Бідюк, Л.О. Коршевнюк, Т.І. Просянкін-Жарова

**АНАЛІЗ ІНВЕСТИЦІЙНИХ І СОЦІАЛЬНО-ЕКОНОМІЧНИХ ПРОЦЕСІВ МЕТОДАМИ
МОДЕЛЮВАННЯ ОБМЕЖЕНИХ МНОЖИН БАГАТОВИМІРНИХ ДАНИХ**

In this paper the analysis of development of economic region of Ukraine is performed with the use of statistical data. The special features of mathematical models construction techniques for analysis and short- and medium term forecasting of regional macroeconomic processes are considered. The approaches are studied to constructing mathematical models with short time series and with the use of intellectual data analysis techniques such as principal component method, Bayesian networks, regression with lagged variables, extended autoregression and trend polynomials. Examples for constructing forecasting models for a regional net product and investments in Cherkasy region are given. Using the models constructed three scenarios are defined for the regional net product development: optimistic, nominal and pessimistic. It is shown that even with the nominal scenario for the period till 2015 the region will get enough investments into basic capital that will provide a simple renovation of production; and in 2020 it is possible to reach the level of extended renovation. The descriptions of the methods used are also given.

Вступ

Фінансова криза продемонструвала чутливість економічної системи країни до змін, що відбуваються у світовій економіці, відзначила стійкість регіонів, особливо аграрних, у протидії кризовим явищам. Тому об'єктивною необхідністю є зосередження уваги на регіональній складовій реформування економіки, підвищенні ролі її інвестиційної складової. Обсяги інвестицій, зокрема для аграрного регіону, не є беззаперечно визначальним фактором економічного зростання. Не менш суттєвими факторами є організація інвестиційного процесу, прогнозування та планування інвестицій, їх територіальне розміщення та спрямування.

Коректне планування майбутнього розвитку соціально-економічних і фінансових процесів ґрунтується на оцінках прогнозів, обчислених за допомогою відповідних (адекватних) математичних і статистичних моделей [1, 2]. Для розв'язання задач коротко- і середньострокового прогнозування сьогодні застосовують методи, що ґрунтуються на різних ідеях стосовно побудови прогнозуючих моделей [2, 3]. Найбільш популярними є відносно прості у побудові та подальшому використанні моделі множинної регресії, авторегресії з ковзним середнім, нейронні мережі, нейронечіткі структури, моделі на основі нечіткої логіки, регресія на опорних векторах, ймовірнісні моделі байєсівського типу та ін. [3]. У випадку наявності в модельованих процесах невизначеностей статистичного і структурного характеру прийнятні результати стосовно оцінок прогнозів можна отримати за допомогою нечіткої логіки та ймовірнісних моделей, зокрема байєсівських мереж [4]. Байєсівська мережа – ймовірнісна модель у

виділі спрямованого ациклічного графа, вершинами якого є вхідні та вихідні змінні процесу і дуги якого вказують на наявні причинно-наслідкові зв'язки між вибраними змінними. Кожна змінна характеризується таблицею умовних ймовірностей, яка формується на основі апріорної інформації про ці змінні у вигляді експертних оцінок і статистичних даних.

У статті подано результати аналізу та моделювання важливих показників розвитку системи інвестиційної діяльності Черкаської області, яка є лідером з виробництва продукції сільськогосподарства (7 % валової продукції сільськогосподарства України), але за рейтингом інвестиційної привабливості тривалий час займає лише 13 місце (в 2004 р.) серед інших областей. Аналогічна ситуація спостерігається і в таких традиційно аграрних регіонах, як Вінницька, Полтавська, Сумська, Хмельницька, Херсонська області.

Постановка задачі

Мета роботи – виконати аналіз статистичних даних стосовно розвитку вибраного регіону України з метою побудови математичних моделей з урахуванням коротких вибірок; для короткострокового прогнозування побудувати моделі множинної регресії, в т.ч. моделі редукованого типу, отримані за методом головних компонент; побудувати ймовірнісну модель у вигляді байєсівської мережі для визначення значимих змінних; створити модель у формі розширеної авторегресії для довгострокового прогнозування; застосувати побудовані моделі для оцінювання короткострокових прогнозів і виконати порівняльний аналіз отриманих результатів.

Особливості математичного моделювання на коротких вибірках

Одна з головних проблем при побудові математичних моделей економічних процесів — це необхідність наявності достатнього обсягу статистичних даних для виявлення закономірностей, оцінювання адекватної структури та обчислення оцінок параметрів моделі.

Існуюча система збору та накопичення статистичної інформації, обмеження щодо конфіденційності окремих показників, адаптація статистичної методології до міжнародних стандартів ускладнюють формування досить великих наборів рядів часових даних відповідних показників. Це призводить до того, що при побудові адекватних математичних моделей для регіонального рівня необхідно ширше застосувати сучасні інтелектуальні засоби, здатні працювати та виявляти закономірності на коротких часових вибірках даних.

Проблема побудови математичних моделей на коротких наборах даних у статистиці більш відома під назвою “прокляття розмірності”. Наприклад, вісьмома точками можна щільно заповнити одновимірний або двовимірний простори, але ці ж вісім точок будуть досить розріджено заповнювати тривимірний простір, а для випадку побудови моделі за десятьма вимірами на сімнадцяти змінних ці ж точки будуть біль-

ше схожі на віддалені одна від одної галактики у Всесвіті (рис. 1).

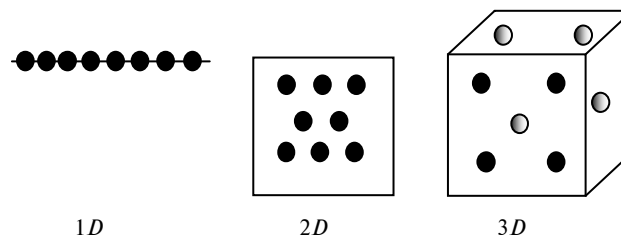


Рис. 1. Ілюстрація проблеми “прокляття розмірності”

Для вирішення цієї проблеми запропоновано скористатись статистичними параметрами і методами попереднього відбору найбільш значимих змінних процесу, а саме: критерієм R -квадрат, байєсівською мережею та методом головних компонент.

Опис статистичних даних

Для побудови математичних моделей використано статистичні дані стосовно Черкаської області з 2000 по 2011 рр. [5, 6].

Вибрані змінні наведено у табл. 1.

У подальшому також буде використовуватися змінна y_1 , що являє собою валовий регіональний продукт (ВРП) за попередній рік, тобто $y_1(k) = y(k - 1)$.

Таблиця 1. Змінні, що використовувалися для моделювання, і їх опис

Змінна	Опис змінної	Одиниця виміру
y	Валовий регіональний продукт (у фактичних цінах)	млн грн
x_{01}	Інвестиції в основний капітал (у фактичних цінах)	млн грн
x_{02}	Основні засоби (у фактичних цінах, на кінець року)	млн грн
x_{03}	Обсяг інноваційних витрат у промисловості	тис. грн
x_{04}	Індекс споживчих цін (грудень до грудня попереднього року)	відсоток
x_{05}	Обсяг реалізованої промислової продукції (робіт, послуг) (у фактичних цінах)	млн грн
x_{06}	Продукція сільського господарства (у порівняних цінах 2005 р.)	млн грн
x_{07}	Обсяг реалізованої будівельної продукції (у фактичних цінах)	млн грн
x_{08}	Доходи населення	млн грн
x_{09}	Рівень зареєстрованого безробіття (на кінець року)	відсоток
x_{10}	Середньорічна кількість найманих працівників	тис.
x_{11}	Середньорічна номінальна заробітна плата	грн
x_{12}	Зовнішня торгівля товарами, експорт	млн дол США
x_{13}	Зовнішня торгівля товарами, імпорт	млн дол США
x_{14}	Прямі іноземні інвестиції в область	тис. дол США
x_{15}	Роздрібний товарообіг підприємств (у фактичних цінах)	млн грн
x_{16}	Відправлення (перевезення) вантажів усіма видами транспорту	млн т
x_{17}	Відправлення (перевезення) пасажирів транспортом загального користування	млн осіб

Побудова моделі множинної лагової регресії з використанням критерію інформативності для прогнозування ВРП

У зв'язку з тим, що навчальна вибірка даних складається лише з десяти точок, а змінних аналізу, що можуть брати участь у побудові моделі, сімнадцять, для відбору найбільш значимих пропонується вибирати ті, які дають кращі значення за критерієм R -квадрат [7]:

$$R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)},$$

де $\text{var}(\hat{y})$ – дисперсія залежної змінної, оціненої за допомогою побудованої математичної моделі, $\text{var}(y)$ – вибіркова дисперсія залежної змінної, оціненої за допомогою її фактичних значень; для адекватної моделі $R^2 \rightarrow 1$.

Для простоти розглянемо приклад, коли для побудови моделі можна скористатись трьома регресорами: x_1, x_2, x_3 (табл. 2, рис. 2).

Таблиця 2. Приклад побудови моделі на основі критерію R -квадрат

Модель	Змінні	R -квадрат
Моделі з однієї змінної	x_3	0,31
	x_2	0,27
	x_1	0,01
Моделі з двох змінних	x_2, x_3	0,6
	x_1, x_3	0,33
	x_1, x_2	0,32
Повна модель	x_1, x_2, x_3	0,57

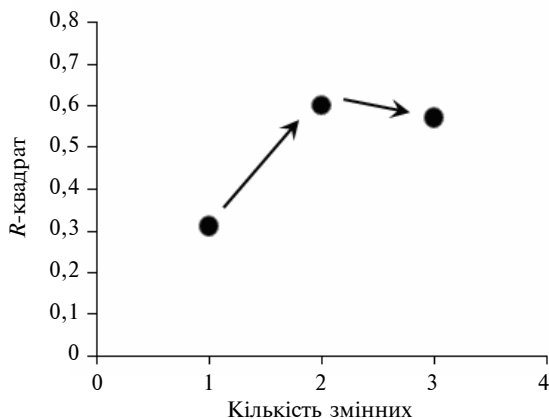


Рис. 2. Зміна значення R -квадрат залежно від включення регресорів у модель

Побудована за цим підходом модель множинної регресії має вигляд

$$y(k) = 1215 - 2,8428 \cdot x_{06}(k-1) + 3,8722 \cdot x_{07}(k-1) - 0,000923 \cdot x_{08}(k-1) + 23,2503 \cdot x_{10}(k-1) + 52,8486 \cdot x_{11}(k-1) - 12,1045 \cdot x_{15}(k-1).$$

Для побудованої моделі середня абсолютна похибка в процентах (САПП) дорівнює 3,23 %. Формула для обчислення САПП має вигляд:

$$\text{САПП} = \frac{100}{n} \cdot \sum_{i=1}^n \left| \frac{y - \hat{y}}{y} \right|,$$

де n – кількість вимірів, y – реальне значення, \hat{y} – прогнозне значення.

Зменшення розмірності простору змінних за методом головних компонент

Одним із досить корисних підходів є використання методу головних компонент (МГК), що був запропонований К. Пірсоном у 1901 р. Суть методу полягає в тому, що визначається набір ортогональних векторів, які краще всього описують напрям зміни в даних порівняно з вхідними даними у початковому вигляді.

У загальному випадку фахівці радять використовувати для аналізу таку кількість компонент, щоб сума їх власних чисел була більшою за 80 або 90 % кількості вхідних змінних. Тобто для заданої множини даних достатньо перших трьох компонент тому, що їх сума дорівнює 15,02 і пояснює 83,7 % змін у даних.

У рамках виконаних досліджень пропонується використовувати таку кількість компонент, що мінімізує середньоквадратичну похибку (СКП) моделі у вигляді регресії при включенні відповідного компонента.

В табл. 3 наведено побудовані компоненти. Як можна побачити, найменша похибка прогнозування ВРП за критерієм СКП досягається при використанні лише першого компонента.

Побудована за цим підходом модель має вигляд

$$y(k) = 11832 + 1833,3 \cdot \text{PCA}_1(k).$$

Для побудованої моделі САПП дорівнює 7,21 %.

Таблиця 3. Список компонентів для методу головних компонент

Номер компоненти	Власне значення компоненти	Процент пояснення змін у даних відповідним компонентом	Кумулятивний процент пояснення змін у даних	Значення СКП при включенні компонента в регресійну модель
PCA_1	11,27	62,61	62,61	1677
PCA_2	2,14	11,9	74,52	1809
PCA_3	1,65	9,18	83,7	1842
PCA_4	0,98	5,48	89,17	1915
PCA_5	0,86	4,82	94	2192
PCA_6	0,71	3,98	97,98	2660
PCA_7	0,26	1,47	99,44	3247
PCA_8	0,1003	0,56	100	3247

Комбінована модель на основі байєсівської мережі і множинної регресії для прогнозування ВРП

Байєсівські мережі (БМ) були запропоновані Д. Перлом у 1985 р, але незважаючи на їх новизну, вони вже досить широко використовуються для визначення причинно-наслідкових зв'язків при моделюванні процесів, що описуються великою кількістю факторів [8]. Саме тому було використано БМ для пошуку зв'язків між ВРП та іншими економічними показниками. Для подальшого використання при прогнозуванні в рамках виконаного дослідження запропоновано методику, що поєднує у собі такі математичні підходи – БМ і регресійний аналіз.

Запропонована двохетапна методика складається з таких кроків.

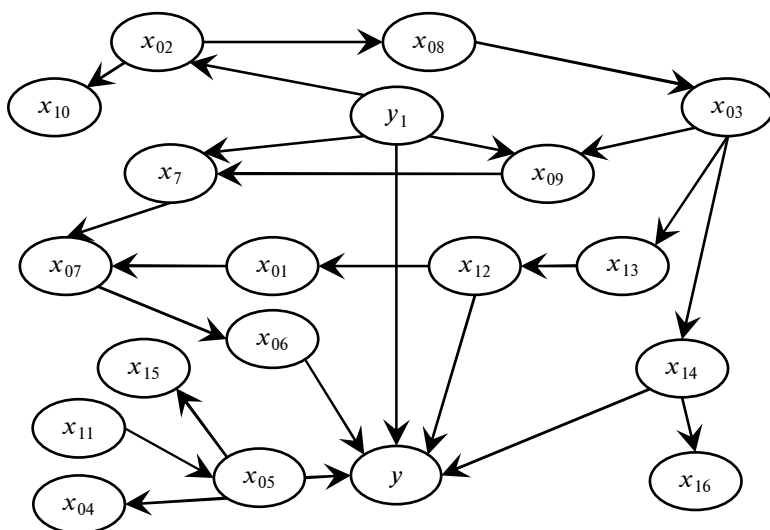


Рис. 3. Топологія побудованої БМ для дослідження розвитку системи інвестиційної діяльності Черкаської області

Етап 1. Будується топологія БМ, що надає інформацію, необхідну для виявлення причинно-наслідкових зв'язків між змінними та силу зв'язків між ними.

Крок 1. Обчислюються відхилення між поточним і попереднім значеннями для всіх змінних. Це робиться для того, щоб позбутися трендової складової в змінних, оскільки ця операція в грубому наближенні еквівалентна знаходженню першої похідної.

Крок 2. Знайдені на попередньому кроці відхилення перетворюються на номінальні змінні завдяки використанню квантильного перетворення. Для даної задачі було задано чотири квантилі.

Крок 3. Для побудови топології мережі використано програмне забезпечення GeNIe 2.0 з такими настройками:

– метод побудови K2;

– максимальна кількість предків – 5;

– критерій оптимізації – мінімальна ентропія.

Етап 2. Прогнозування на основі побудованої причинно-наслідкової структури БМ.

На основі побудованої топології БМ визначаються найбільш значимі змінні, що впливають на цільову змінну, після чого будується рівняння множинної регресії із примусовим включенням до моделі виявлених змінних. Оцінка параметрів моделі виконується на основі рекурсивного методу найменших квадратів.

Найбільш значимими для подальшого прогнозування ви-

Таблиця 4. Реальні та прогнозні значення, отримані за побудованими моделями

Рік	Реальні історичні значення	Множинна лагова регресія за критерієм R -квадрат	Лінійна регресія на основі МГК	Комбінована модель БМ і множинної регресії
2002	3852	3867,18	4182,49	4216,53
2003	4565	4544,54	4914,24	5357,96
2004	6623	6663,65	6771,29	6262,34
2005	9014	9046,33	9939,39	9339,17
2006	10957	10872,25	10595,96	11028,1
2007	13656	13656,16	13400,55	13336,96
2008	19101	19104,44	16591,6	16264,23
2009	18707	18668,16	21755,59	18725,04
2010	20016	20068,79	18340,37	21961,15
2011	21417	21361,75	21314,26	21304,72

Таблиця 5. Статистичні характеристики побудованих математичних моделей

Статистична характеристика	Множинна лагова регресія за критерієм R -квадрат	Лінійна регресія на основі МГК	Комбінована модель БМ і множинної регресії
Інформаційний критерій Акайке	80,64	135,4	110,34
Середньоквадратична похибка	85,98	1677,83	395,43
Критерій Шварца–Байєса	82,02	135,79	111,33
Сума квадратів похибок	14785,91	19705502,04	625446,49

явилися показники обсягу реалізованої промислової продукції, прямих іноземних інвестицій, відправлення (перевезення) вантажів, валової продукції сільського господарства та значення валового регіонального продукту за попередній рік.

Рівняння для прогнозування має вигляд

$$y(k) = 11429,9 + 2,7017 \cdot y(k-1) - 1,3027 \cdot x_{05}(k-1) - 3,9693 \cdot x_{06}(k-1) + 38,8244 \cdot x_{14}(k-1).$$

Для побудованої моделі САПП дорівнює 6,4 %.

Результати короткострокового прогнозування

У табл. 4 наведено прогнозні значення, оцінені на основі запропонованих математичних моделей, а в табл. 5 подані статистичні характеристики відповідних моделей.

Авторегресійна модель для довгострокового прогнозування ВРП

Аналіз часового ряду даних ВРП свідчить про наявність таких статистичних характеристик:

- 1) лагу першого порядку в даних (табл. 6);
- 2) нестационарності за математичним сподіванням та дисперсією.

Таблиця 6. Значення часткової автокореляційної функції для часового ряду ВРП

Порядок запізнення	Значення ЧАКФ
1	0,798
2	-0,136
3	-0,111
4	-0,147
5	0,006

Для врахування зазначених вище особливостей ряду побудована авторегресійна модель для довгострокового прогнозування, що складається з двох рівнянь:

рівняння тренду

$$\text{trend}(k) = -1059,24 + 1867,61 \cdot k,$$

де k – змінна, що набуває значення від 1 до 21 (значення $k = 1$ відповідає 2000 року, $k = 2$ відповідає 2001 і так далі до $k = 21$, що відповідає 2020 року);

рівняння прогнозування відхилень від тренду

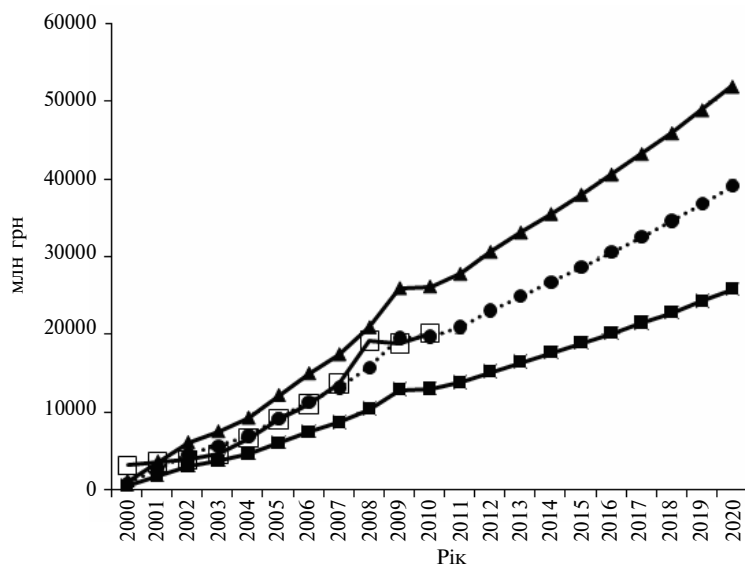


Рис. 4. Сценарії розвитку динаміки ВВП Черкаської області на період до 2020 р.: $\bullet\bullet$ – помірний прогноз; \blacksquare – песимістичний прогноз; \blacktriangle – оптимістичний прогноз; \square – реальні історичні значення

$$dy(k) = -523,6 + 0,537 \cdot dy(k-1) + 0,00016698 \cdot x_{01}(k-1) + 6,001318 \cdot 10^{-5} \cdot x_{01}(k-2) - 0,0001721 \cdot x_{01}(k-3),$$

де dy – відхилення від тренду, x_{01} – рівень інвестицій в основний капітал.

На основі наведених рівнянь остаточний прогноз обчислюється так:

$$y(k) = \text{trend}(k) + dy(k).$$

Із використанням описаних рівнянь побудовано три сценарії зростання ВРП Черкащини – оптимістичний, помірний і песимістичний. Результати наведено на рис. 4.

Як показали виконані розрахунки, побудована економіко-математична модель для довгострокового прогнозування обсягів інвестицій в основний капітал має прийнятні статистичні характеристики та забезпечує належну якість прогнозу і може бути використана під час прогнозування обсягів інвестицій в основний капітал на довгострокову перспективу.

Саме по собі зростання обсягів інвестицій не забезпечує економічного зростання, а тому, спрогнозувавши обсяги інвестицій в основний капітал, необхідно перевірити, чи перебувають прогнозовані значення на рівні, що відповідає необхідному для простого та розширеного відтворення. Крім того, при розробці державних програм і стратегій важливим є окреслення строків, коли даний результат може бути досяг-

нуто. Оскільки для досягнення простого відтворення достатньо, щоб інвестиції становили 30 % обсягу валового регіонального продукту, то для розширеного відтворення – до 40 % [5, с. 273].

За основу взято прогноз за помірним варіантом розвитку подій. Динаміка співвідношення валового регіонального продукту та інвестицій в основний капітал подана в табл. 7.

Як видно з табл. 7, навіть за помірним сценарієм на період до 2015 р. в область спрямовуватиметься обсяг інвестицій в основний капітал, достатній для забезпечення простого відтворення, а у 2020 р. є можливим досягнення рівня розширеного відтворення.

Таблиця 7. Динаміка співвідношення обсягу інвестицій та валового регіонального продукту Черкаської області у 2000–2009 р. та її прогноз до 2020 р.

Рік	Валовий регіональний продукт, млн грн	Інвестиції в основний капітал (розраховані за моделлю), млн грн	Співвідношення інвестицій та валового регіонального продукту, %
2000	3 179	330	10
2001	3 590	431	12
2002	3 852	706,8	18
2003	4 565	1142,7	25
2004	6 623	2690,6	41
2005	9 014	2408,2	27
2006	10 957	3734,8	34
2007	13 656	4808,3	35
2008	19 101	5814,6	30
2009	18 707	3616,3	19
2010	19 672*	3385,1	17
2011	20 896*	4283,5	21
2012	23 028*	5041,7	22
2013	24 918*	6130,3	25
2014	26 716*	7084,1	27
2015	28 597*	8422,2	30
2016	30 541*	9312,2	31
2017	32 526*	10606,9	33
2018	34 559*	12037,2	35
2019	36 796*	13615,8	37
2020	39 080*	15750,2	40

* – прогноз.

Висновки

Вибрано змінні для аналізу регіонального розвитку Черкаської області України і побудовано прогнозуючі математичні моделі деяких процесів на основі зібраних статистичних даних. Для редукції кількості вхідних змінних використано метод головних компонент і байєсівську мережу. Зокрема, побудовано множинну лагову регресію, яка забезпечує отримання оцінок короткострокового прогнозу з САПП = 3,23 %.

З використанням методу головних компонент побудовано регресійну модель і показано, що найменша похибка прогнозування валового регіонального продукту за критерієм СКП досягається при використанні лише першої компоненти (СКП = 1677; САПП = 7,21 %). За допомогою створеної байєсівської мережі встановлено, що найбільш значимими для подальшого прогнозування є показники обсягу реалізованої промислової продукції, прямих іноземних інвестицій, відправлення (перевезення) вантажів, валової продукції сільського господарства та значення валового регіонального продукту за попередній рік. Модель, побудована з використанням для редукції кількості змінних

байєсівської мережі, має хороші прогнозуючі характеристики, зокрема САПП = 6,4%.

Побудовано авторегресійну модель для довгострокового прогнозування, що складається з двох рівнянь: рівняння тренду і рівняння для прогнозування відхилень від тренду. Отримана модель має прийнятні прогнозуючі характеристики. З використанням створених моделей побудовано три сценарії зростання ВРП Черкащини – оптимістичний, помірний і песимістичний. Показано, що навіть за помірним сценарієм на період до 2015 р. в область спрямовуватиметься обсяг інвестицій в основний капітал, достатній для забезпечення простого відтворення, а в 2020 р. є можливим досягнення рівня розширеного відтворення.

У подальших дослідженнях планується виконання дослідження соціально-економічних процесів на рівні адміністративно-територіальних одиниць регіону з метою побудови уточнених математичних моделей, призначених для побудови прогнозів, які можна використати в органах місцевого самоврядування та державного управління для прийняття рішень з урахуванням специфіки відповідних населених пунктів.

1. Кейн Э. Экономическая статистика и эконометрия. – М.: Статистика, 1977. – 230 с.
2. Грін У. Економетричний аналіз. – К.: Основи, 2003. – 1110 с.
3. Бідюк П.І., Меньїленко О.С., Половцев О.В. Методи прогнозування. – Луганськ: Альма-Матер, 2008. – 608 с.
4. G. Cowell et al., "Probabilistic networks and expert systems". New York: Springer, 1999, 322 pp.
5. Регіональний розвиток Черкаської області за 2010 р.: 36 статей. – Черкаси, 2011. – С. 343.
6. Статистичний щорічник Черкаської області за 2010 рік. – Черкаси, 2011. – С. 546.
7. Кобзарь А.И. Прикладная математическая статистика. Для инженеров и научных работников. – М.: ФИЗМАТЛИТ, 2006. – 816 с.
8. Бідюк П.І., Кузнєцова Н.В., Терентьєв О.М. Система підтримки прийняття рішень для аналізу фінансових даних // Наукові вісті НТУУ "КПІ". – 2011. – № 1. – С. 48–61.

Рекомендована Радою
Навчально-наукового комплексу
"Інститут прикладного системного
аналізу" НТУУ "КПІ"

Надійшла до редакції
25 січня 2012 року