

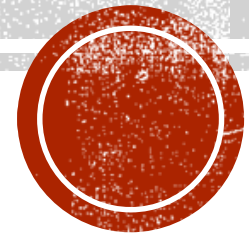
АВТОМАТИЗОВАНИЙ ВИБІР ПРОГНОЗУЮЧИХ МОДЕЛЕЙ В СИСТЕМАХ ПІДТРИМКИ ПРИЙНЯТТЯ РІШЕНЬ

Виконала студентка 4 курсу групи КА-35

Левенчук Людмила Борисівна

Керівник д.т.н., професор кафедри ММСА

Бідюк Петро Іванович



АКТУАЛЬНІСТЬ

- Комп'ютерні технології дозволяють обробляти великі об'єми даних за прийнятні витрати часу, будувати адекватні математичні моделі досліджуваних процесів, прогнозувати розвиток процесів та планувати управлінські дії.
- Прогнозування – один з вирішальних елементів організації управління та прийняття обґрунтованих рішень у різних сферах життєдіяльності.
- Системи підтримки прийняття рішень дозволяють швидко обрати кращий варіант розв'язання задач з множини існуючих за допомогою відповідних критеріїв якості.



ОБ'ЄКТ І ПРЕДМЕТ ДОСЛІДЖЕННЯ

Об'єкт: Лінійні та нелінійні нестационарні процеси в економіці та фінансах.

Предмет: Математичні моделі і методи формального опису процесів в економіці та фінансах; система підтримки прийняття рішень

Мета: Створення системи підтримки прийняття рішень для побудови математичних моделей на основі статистичних даних і автоматизованого вибору прогнозуючих моделей.



ПОСТАНОВКА ЗАДАЧІ

- Виконати огляд відомих систем для побудови прогнозуючих моделей. Запропонувати комплексні критерії вибору моделі.
- Розробити архітектуру і функціональну схему системи для побудови та автоматизації вибору прогнозуючих моделей.
- Розробити ПЗ для розв'язання задач побудови математичної моделі вибраних ФЕП з функцією автоматизованого вибору кращої моделі.
- Виконати обчислювальні експерименти за допомогою розробленого ПЗ з використанням фактичних статистичних даних.
- Проаналізувати отримані результати, визначити перспективи подальших досліджень.



СИСТЕМИ ДЛЯ СТАТИСТИЧНОЇ ОБРОБКИ ДАНИХ

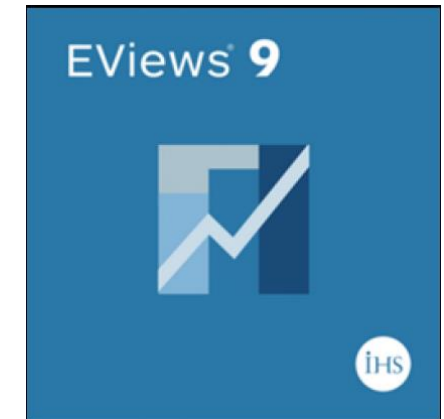
■ SPSS

- + Гнучкість
- + Допоміжні матеріали
- Велика вартість



■ EViews

- + Компактність
- + Доступність
- Візуалізація



■ STATISTICA

- + Візуалізація
- + Продуктивність
- Відсутність тестів на стаціонарність



■ SAS

- + Потужність
- Велика вартість
- Великий об'єм



КРИТЕРІЇ АДЕКВАТНОСТІ МАТЕМАТИЧНОЇ МОДЕЛІ

- Статистика Стьюдента: $t_a = \frac{\hat{a} - a_0}{SE_{\hat{a}}}$
- Коефіцієнт множинної детермінації: $R^2 = \frac{\text{var}(\hat{y})}{\text{var}(y)}$
- Сума квадратів похибок: $\sum_{k=1}^N e^2(k) = \sum_{k=1}^N [\hat{y}(k) - y(k)]^2$
- Інформаційний критерій Акайке: $AIC = N \cdot \ln(\sum_{k=1}^N e^2(k)) + 2n$
- Критерій Байєса-Шварца: $BSC = N \cdot \ln(\sum_{k=1}^N e^2(k)) + n \cdot \ln(N)$
- Статистика Дарбіна-Уотсона: $DW = 2 - 2 \frac{E[e(k)e(k-1)]}{\sigma_e^2}$
- Статистика Фішера: $F \sim \frac{R^2}{1-R^2}$
- Коефіцієнт Тейла: $U = \frac{\sqrt{\frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2}}{\sqrt{\frac{1}{N} \sum_{k=1}^N y^2(k) + \frac{1}{N} \sum_{k=1}^N \hat{y}^2(k)}}$



КРИТЕРІЇ ЯКОСТІ ОЦІНОК ПРОГНОЗІВ

- Середньоквадратична похибка:

$$\text{СКП} = \frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)]^2 = \frac{1}{N} \sum_{k=1}^N e^2(k)$$

- Середня похибка:

$$\text{СП} = \frac{1}{N} \sum_{k=1}^N [y(k) - \hat{y}(k)] = \frac{1}{N} \sum_{k=1}^N e(k)$$

- Середня похибка у процентах:

$$\text{СПП} = \frac{1}{N} \sum_{k=1}^N \frac{[y(k) - \hat{y}(k)]}{y(k)} \times 100\%$$

- Середнє абсолютне значення похибки у процентах:

$$\text{САПП} = \frac{1}{N} \sum_{k=1}^N |y(k) - \hat{y}(k)| \times 100\%$$

- Максимальна абсолютна похибка: $\text{МАП} = \max_k \{|y(k) - \hat{y}(k)|\}, 1 \leq k \leq N$

- Мінімальна абсолютна похибка: $\text{МіАП} = \min_k \{|y(k) - \hat{y}(k)|\}, 1 \leq k \leq N$



РЕГРЕСІЙНІ МОДЕЛІ НЕСТАЦІОНАРНИХ ПРОЦЕСІВ

- Авторегресія
$$y(k) = a_0 + a_1 y(k-1) + \dots + a_p y(k-p) = a_0 + \sum_{i=1}^p a_i y(k-i) + \varepsilon(k)$$
- Парна регресія
$$y(k) = a_0 + a_1 x(k) + \varepsilon(k)$$
- Множинна регресія
$$y(k) = a_0 + a_1 x_1(k) + a_2 x_2(k) + \dots + a_p x_p(k) + \varepsilon(k)$$
- Авторегресія з ковзним середнім
$$y(k) = a_0 + \sum_{i=1}^p a_i y(k-i) + \sum_{j=1}^q b_j \varepsilon(k-j) + \varepsilon(k)$$
- Змішана регресія
$$y(k) = a_0 + \sum_{i=1}^l a_i y(k-i) + b_1 x_1(k) + b_2 x_2(k) + \dots + b_p x_p(k) + \varepsilon(k)$$



КОМПЛЕКСНІ КРИТЕРІЇ

- $y_1 = |1 - R^2| + \alpha \cdot \ln(\text{СКП}),$

- $y_2 = |1 - R^2| + \beta \cdot \ln(\text{САПП}),$

- $y_3 = |1 - R^2| + |2 - DW| + \gamma_1 \cdot \ln(\sum_{k=1}^N e^2(k)) + \gamma_2 \cdot \ln(\text{САПП}) + \gamma_3 \cdot U.$



ПРИКЛАДИ РОБОТИ ПРОГРАМИ

DSS by Levenchuk

Вхідний ряд

n	Y(n)
0	98,9
1	99,3
2	98,5
3	102,8
4	103,7
5	100,8
6	98,8
7	104,8
8	102,2
9	102,1
10	101,2
11	106,3
12	104,4

Оберіть модель

☒ АРКС ☐ Лог. регресії

Оберіть діапазони параметрів

Порядок АР: від 1 до 15 Порядок КС: від 0 до 15

Оберіть метод наближення

МНК

Оберіть критерій оптимальності моделі

Сумма квадратів похибок

Додатково

☒ Врахувати a_0

Старт ...

Завантажено ряд ISC
Кількість елементів: 67
Ряд є нелінійним



СУМА КВАДРАТІВ ПОХИБОК

DSS by Levenchuk

Вхідний ряд		Оберіть модель		Параметри моделі		Вихідний ряд																																																																																			
f_x <table><thead><tr><th>n</th><th>Y(n)</th></tr></thead><tbody><tr><td>0</td><td>98,9</td></tr><tr><td>1</td><td>99,3</td></tr><tr><td>2</td><td>98,5</td></tr><tr><td>3</td><td>102,8</td></tr><tr><td>4</td><td>103,7</td></tr><tr><td>5</td><td>100,8</td></tr><tr><td>6</td><td>98,8</td></tr><tr><td>7</td><td>104,8</td></tr><tr><td>8</td><td>102,2</td></tr><tr><td>9</td><td>102,1</td></tr><tr><td>10</td><td>101,2</td></tr><tr><td>11</td><td>106,3</td></tr><tr><td>12</td><td>104,4</td></tr></tbody></table>		n	Y(n)	0	98,9	1	99,3	2	98,5	3	102,8	4	103,7	5	100,8	6	98,8	7	104,8	8	102,2	9	102,1	10	101,2	11	106,3	12	104,4	☒ АРКС ☐ Лог.регресії Оберіть діапазони параметрів Порядок АР: від 1 до 15 Порядок КС: від 0 до 15 Оберіть метод наближення МНК Оберіть критерій оптимальності моделі Сумма квадратів похибок Додатково ☒ Врахувати a0 Старт ...		<table><thead><tr><th>Параметр</th><th>Значення</th></tr></thead><tbody><tr><td>a0</td><td>92,323263</td></tr><tr><td>a1</td><td>0,6836362</td></tr><tr><td>a2</td><td>-0,3861864</td></tr><tr><td>a3</td><td>0,3001980</td></tr><tr><td>a4</td><td>-0,0959863</td></tr><tr><td>a5</td><td>-0,0499451</td></tr><tr><td>a6</td><td>-0,0454594</td></tr><tr><td>a7</td><td>-0,0943367</td></tr><tr><td>a8</td><td>0,0535390</td></tr><tr><td>a9</td><td>-0,0948126</td></tr><tr><td>a10</td><td>0,0245288</td></tr><tr><td>a11</td><td>-0,0931767</td></tr></tbody></table> Прогноз ...		Параметр	Значення	a0	92,323263	a1	0,6836362	a2	-0,3861864	a3	0,3001980	a4	-0,0959863	a5	-0,0499451	a6	-0,0454594	a7	-0,0943367	a8	0,0535390	a9	-0,0948126	a10	0,0245288	a11	-0,0931767	<table><thead><tr><th>n</th><th>Y(n)</th></tr></thead><tbody><tr><td>0</td><td>103,273566699</td></tr><tr><td>1</td><td>103,286809953</td></tr><tr><td>2</td><td>104,314789692</td></tr><tr><td>3</td><td>102,585472603</td></tr><tr><td>4</td><td>100,513709712</td></tr><tr><td>5</td><td>104,992561073</td></tr><tr><td>6</td><td>104,710569445</td></tr><tr><td>7</td><td>101,125971814</td></tr><tr><td>8</td><td>104,208869798</td></tr><tr><td>9</td><td>103,980617902</td></tr><tr><td>10</td><td>106,364844124</td></tr><tr><td>11</td><td>105,773172439</td></tr><tr><td>12</td><td>102,938393409</td></tr></tbody></table>		n	Y(n)	0	103,273566699	1	103,286809953	2	104,314789692	3	102,585472603	4	100,513709712	5	104,992561073	6	104,710569445	7	101,125971814	8	104,208869798	9	103,980617902	10	106,364844124	11	105,773172439	12	102,938393409
n	Y(n)																																																																																								
0	98,9																																																																																								
1	99,3																																																																																								
2	98,5																																																																																								
3	102,8																																																																																								
4	103,7																																																																																								
5	100,8																																																																																								
6	98,8																																																																																								
7	104,8																																																																																								
8	102,2																																																																																								
9	102,1																																																																																								
10	101,2																																																																																								
11	106,3																																																																																								
12	104,4																																																																																								
Параметр	Значення																																																																																								
a0	92,323263																																																																																								
a1	0,6836362																																																																																								
a2	-0,3861864																																																																																								
a3	0,3001980																																																																																								
a4	-0,0959863																																																																																								
a5	-0,0499451																																																																																								
a6	-0,0454594																																																																																								
a7	-0,0943367																																																																																								
a8	0,0535390																																																																																								
a9	-0,0948126																																																																																								
a10	0,0245288																																																																																								
a11	-0,0931767																																																																																								
n	Y(n)																																																																																								
0	103,273566699																																																																																								
1	103,286809953																																																																																								
2	104,314789692																																																																																								
3	102,585472603																																																																																								
4	100,513709712																																																																																								
5	104,992561073																																																																																								
6	104,710569445																																																																																								
7	101,125971814																																																																																								
8	104,208869798																																																																																								
9	103,980617902																																																																																								
10	106,364844124																																																																																								
11	105,773172439																																																																																								
12	102,938393409																																																																																								

Завантажено ряд ISC
Кількість елементів: 67
Ряд є нелінійним
Найкраща MSE при
Авторегресія: 14
Ковзне середнє: 0
MSE: 2,94947093478376
DW: 1,95271536257843
RQ: 0,424262260642124

КОЕФІЦІЄНТ ДЕТЕРМІНАЦІЇ

DSS by Levenchuk

Вхідний ряд

n	Y(n)
0	98,9
1	99,3
2	98,5
3	102,8
4	103,7
5	100,8
6	98,8
7	104,8
8	102,2
9	102,1
10	101,2
11	106,3
12	104,4

Оберіть модель

☒ АРКС ☐ Лог.регресії

Оберіть діапазони параметрів

Порядок АР: від 1 до 15 Порядок КС: від 0 до 15

Оберіть метод наближення

МНК

Оберіть критерій оптимальності моделі

Коефіцієнт детермінації

Додатково

☒ Врахувати a0

Старт ...

Параметри моделі

Параметр	Значення
a0	15,1092418...
a1	1,09993941...
a2	-0,1573981...
a3	0,09893085...
a4	0,07652321...
a5	-0,2639204...
b0	-0,4735445...
b1	-0,4695879...

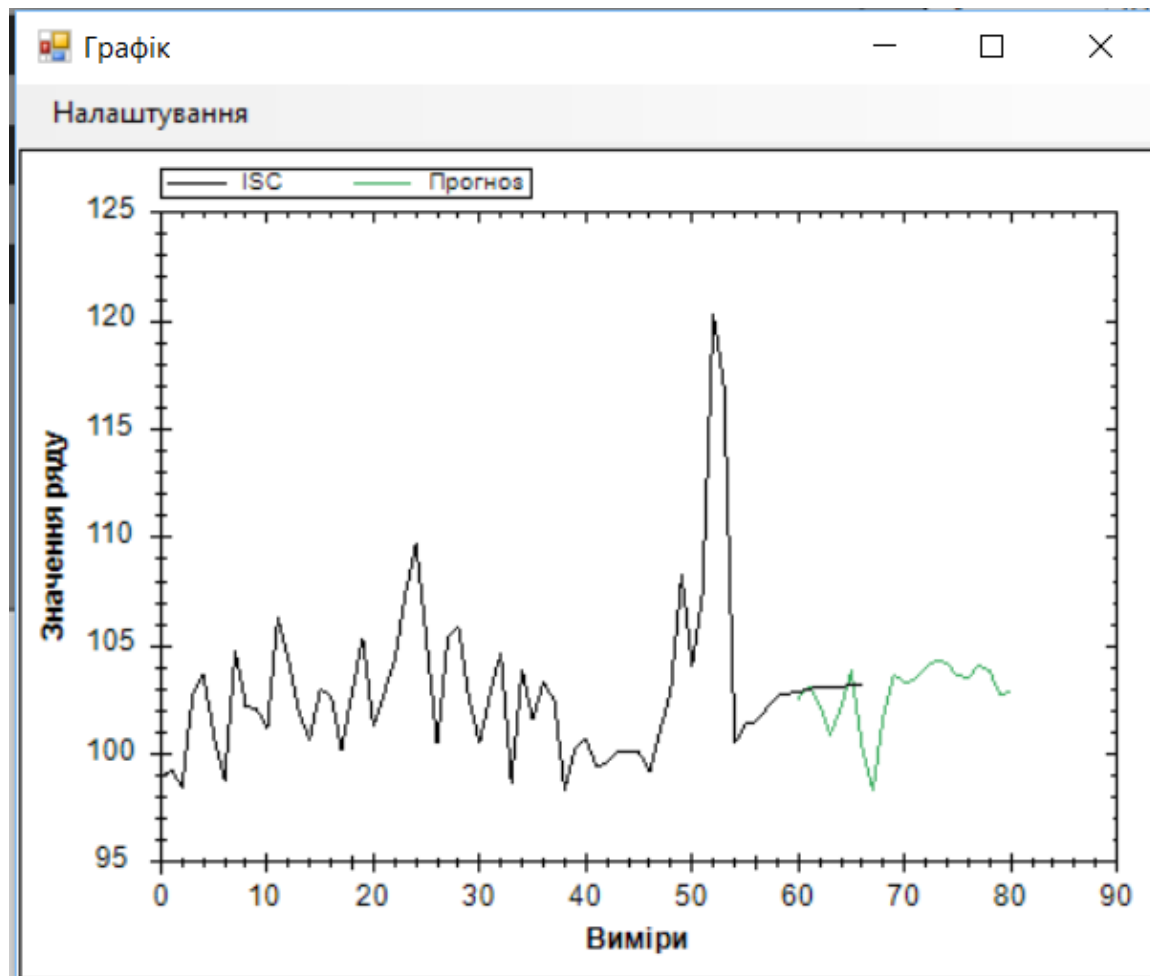
Прогноз ...

Вихідний ряд

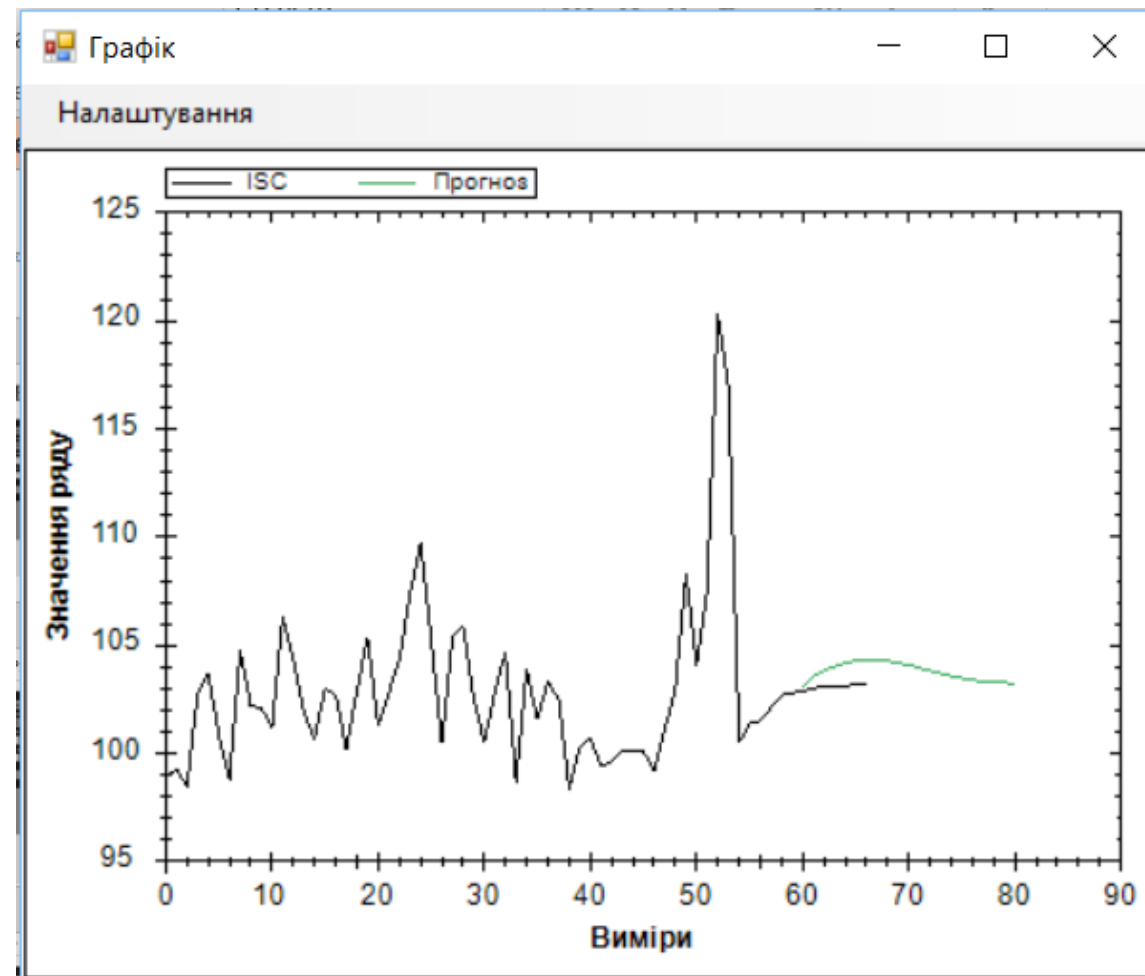
n	Y(n)
0	104,234133612
1	101,161266871
2	100,047066731
3	105,608612836
4	101,147073814
5	102,652223310
6	102,407777680
7	106,366750636
8	104,063638242
9	102,074914779
10	101,588311116
11	102,568333798
12	102,067760056

Завантажено ряд ISC
Кількість елементів: 67
Ряд є нелінійним
Найкраща RQ при
Авторегресія: 5
Ковзне середнє: 2
MSE: 3,58407413497951
DW: 2,31813257990857
RQ: 1,07527486211408

ПРОГНОЗИ



САПП=2,65%



САПП=2,93%



АНАЛІЗ РЕЗУЛЬТАТІВ

Ряд	Тип моделі	Адекватність моделі			Характеристика прогнозу		
		R^2	$\sum e^2$	DW	СКП	СПП %	САПП %
IPV	АРІКС(11, 0)	0,296	3,619	2,014	0,1869	-3,94	3,94
	АРІКС(13, 0)	0,2999	3,671	2,0002	0,5707	-1,2	1,2
	АРІКС(6, 6)	1,0265	5,4427	1,4478	1,6915	-3,558	3,558

Ряд	Тип моделі	Адекватність моделі			Характеристика прогнозу		
		R^2	$\sum e^2$	DW	СКП	СПП %	САПП %
ISC	АРІКС (14,0)	0,4243	2,9495	1,9527	1,3871	2,935	2,935
	АРІКС (9,0)	0,376	2,969	2,001	0,278	5,8	5,8
	АРІКС (5, 2)	1,075	3,584	2,318	1,254	-2,65	2,65

Ряд	Тип моделі	Адекватність моделі			Характеристика прогнозу		
		R^2	$\sum e^2$	DW	СКП	СПП %	САПП %
VVP	AP (1)	0,8939	52041	1,848	1,3871	2,935	2,935



НАУКОВА НОВИЗНА

- Спроектовано і реалізовано оригінальну інформаційну систему для моделювання процесів довільної природи за експериментальними (статистичними) даними.
- Побудовані нові моделі вибраних ФЕП, які забезпечують обчислення високоякісних оцінок прогнозів.
- Розроблено комплексний критерій якості моделі і прогнозу, який забезпечує вибір кращої моделі в автоматичному режимі.
- Запропоновано модифіковану методику моделювання процесів довільної природи за статистичними даними, яка передбачає автоматичний вибір структури моделі.



ВИСНОВКИ



- Виконано огляд моделей, які можуть бути використані для формального опису лінійних та нелінійних процесів в економіці та фінансах.
- Розроблена та програмно реалізована СППР для обробки статистичних даних, побудови моделей ФЕП та оцінювання прогнозів.
- Для вибраних ФЕП побудовані регресійні моделі АР, **АР(І)КС** і логістичну регресію, які характеризуються високим ступенем адекватності і забезпечують можливість обчислення високоякісних оцінок прогнозів.



ПОДАЛЬШЕ ДОСЛІДЖЕННЯ

- Розширення номенклатури типів математичних моделей для опису ФЕП.
- Використання методів інтелектуального аналізу даних.
- Застосування методів комбінування оцінок прогнозів, обчислених за допомогою альтернативних методів.
- Розробка та реалізація комерційної СППР.



ДЯКУЮ ЗА УВАГУ!

