

А.Е. Матушкіна, К.Е. Золотько

Дніпровський національний університет ім. Олеся Гончара

ПРОГНОЗУВАННЯ ЦІН ПОЛІЕТИЛЕНУ З ВИКОРИСТАННЯМ МЕТОДІВ КОРРЕЛЯЦІНО-РЕГРЕСІЙНОГО АНАЛІЗУ В СЕРЕДОВИЩІ STATISTICA 8

Виконано аналіз залежності ціни поліетилену від вартості ресурсів (етилен, попит) з використанням прикладного програмного забезпечення «Statistica-8.0». Розроблений метод може бути представлений для прогнозування ціни поліетилену на європейському ринку з ймовірністю більше, ніж 85%. Використання методів кореляційно-регресійного аналізу дозволяє одержати близькі фактичним значенням ціни поліетилену.

Выполнен анализ зависимости цены полиэтилена от стоимости ресурсов (этилен, спрос) с использованием прикладного программного обеспечения «Statistica 8.0». Разработанный метод может быть представлен для прогнозирования цены полиэтилена на европейском рынке с вероятностью более 85%. Использование методов корреляционно-регрессионного анализа позволяет получить близкие фактическим значением цены полиэтилена.

The analysis of the dependence of the price of polyethylene on the cost of resources (ethylene, demand) using the software application "Statistica 8.0" was performed. The developed method can be presented for predicting the price of polyethylene in the European market with a probability of more than 85%. Using the methods of correlation-regression analysis allows you to get close to the actual value of the price of polyethylene.

Ключові слова: математичні методи аналізу, кореляційно-регресійний аналіз, прогнозування цін, поліетилен європейського ринку.

Вступ. В роботі спостерігається аналіз залежності ціни поліетилену від вартості ресурсів (етилен, попит) з використанням прикладного програмного забезпечення «Statistica 8.0». Завчасно отримана інформація про тенденції на ринку збуту та попиту відіграє велику роль у підтриманні конкурентоспроможності об'єктів. Розроблена модель з достатньою точністю показує цінову ситуацію на ринку, однак на практиці можна побачити розбіжності між розрахунковими і опублікованими прогнозними ціноутвореннями. Використання методів кореляційно-регресійного аналізу дозволяє одержати близькі значення до фактичних значень ціни поліетилену. Це підтверджує можливість використання розрахункових методів прогнозування.

Постановка проблеми. Метою роботи є удосконалення методів прогнозування ціни поліетилену в залежності від цін етилену для забезпечення конкурентоспроможності підприємств на європейському ринку.

Метод розв'язування та аналіз роботи. В роботі досліджується метод підвищення точності прогнозування цін поліетилену на європейському ринку з урахуванням вже існуючих даних: ціни етилену на попит.

Запропонована множинна регресійна модель залежності цін поліетилену від цін на етилен, використаних при виробництві. Розроблена модель з достатньо високою точністю відображає цінову ситуацію на ринку, однак на практиці спостерігається розбіжність між розрахунковими та дійсними цінами, що призвело до розробки підвищення точності прогнозу.

Ціль. Розробка метода підвищення точності прогнозування цін поліетилену на європейському ринку з урахуванням прогнозування цін на сирові матеріали.

Основна частина. В роботі було застосовано регресійний аналіз.

Регресійний аналіз - статистичний метод, за допомогою якого можна побудувати модель з однієї залежної змінної (відгуком) і однієї або декількома незалежними змінними (факторами).

Регресійний аналіз можна використовувати для визначення наявності зв'язку між змінними, оскільки наявність такого зв'язку і є передумовою для застосування аналізу та для створення прогнозу [1].

За допомогою бази даних цін поліетилену та «Statistica-8.0» було:

- 1) виявлено, які з факторів діють на відгук, а які – ні;
- 2) розранжовано фактори за ступенем впливу на відгук;
- 3) прогнозовано значення відгуку при певних значеннях факторів.

Побудована за допомогою регресійного аналізу модель являє собою рівняння виду:

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k$$

де X_1, X_2, \dots, X_k — фактори;

Y – відгук;

b_0, b_1, \dots, b_k — параметри (коефіцієнти) регресії.

Математично, суть регресійного аналізу зводиться до знаходження параметрів регресії, перевірки їх значимості і оцінці прийнятності всієї побудованої лінійної моделі в цілому.

Основними цілями регресійного аналізу є:

- визначення ступеня детермінованості варіації залежної змінної незалежними змінними;
- передбачення значення залежної змінної за допомогою незалежних;
- визначення вкладу окремих незалежних змінних в варіацію залежної.

В процесі реалізації отриманої моделі було: знайдено і усунено мультиколінеарності факторів; проаналізовано кореляційну таблицю, знайдено мультиколінеарності чинники; інтерпретовано коефіцієнти Beta (з ряду факторів залишили лише ті, які суттєво впливають на ціноутворення поліетилену); оцінено нормальність розподілу залишків по гістограмі; оцінено залежність залишків від передбачених за рівнянням регресії значень відгуку(знайдено фактори, які не впливають на відгук); зроблено висновок про нормальність

розподілу залишків; обґрунтовано регресійні рівняння; проаналізовано коефіцієнт детермінації; виконано порівняння прогнозу з фактичним результатом.

На рис.1 приведено базу даних з ціновими показниками поліетилену.

	1	2	3	4	5
	Prices	Ethylene	Demand	Capacity	Naptha
1	1320	800	7	800	788
2	1320	800	7	800	752
3	1310	780	6	850	763
4	1290	770	4	900	757
5	1250	750	2	1000	797
6	1250	750	3	1000	775
7	1250	760	3	950	785
8	1260	770	4	950	774
9	1270	770	5	950	751
10	1270	760	5	950	765
11	1250	760	4	1000	781
12	1250	760	4	1000	768

Рис. 1. База даних ціновими показниками поліетилену

Ціни поліетилену являються залежною змінною, а такі показники, як ціни етилену, попит, об'єм виробництва та ціна нафти, — незалежними змінними.

За допомогою «Statistica-8.0» було побудовано кореляційну матрицю (рис. 2), виявлено, які фактори діють на відгук, а які – ні.

Variable	Correlations (Spreadsheet12)				
	Ethylene	Demand	Capacity	Naptha	Prices
Ethylene	1,000000	-0,051565	0,097533	0,068105	-0,113693
Demand	-0,051565	1,000000	-0,222867	-0,168863	0,836706
Capacity	0,097533	-0,222867	1,000000	0,180412	-0,112283
Naptha	0,068105	-0,168863	0,180412	1,000000	-0,040182
Prices	-0,113693	0,836706	-0,112283	-0,040182	1,000000

Рис. 2. Кореляційна матриця база даних з 4 незалежними змінними.

Наступним кроком було побудувати регресійну модель та виявити коефіцієнт Beta (рис. 3).

Один з мультиколінеарних факторів (попит) не підходить, тому він був видалений з незалежних змінних (рис. 4).

Побудована таблиця рівня значущості, що впливає на залежність змінної (рис. 5). Зі стовпця «p-level» можна зробити висновок, що показники ціни нафти мають значення більше, ніж 0,05, тому слід видалити і цей фактор також. Якщо існує більше таких факторів, то потрібно було би видалити всі незалежні змінні.

Regression Summary for Dependent Variable: Prices (Spreadsheet12)						
R= ,84920084 R ² = ,72114207 Adjusted R ² = ,66243514						
F(4, 19)=12,284 p<,00004 Std.Error of estimate: 14,804						
N=24	Beta	Std.Err. of Beta	B	Std.Err. of B	t(19)	p-level
Intercept			1079,616	181,8060	5,938287	0,000010
Ethylene	-0,082718	0,121926	-0,037	0,0542	-0,678434	0,505676
Demand	0,864875	0,125444	14,004	2,0312	6,894540	0,000001
Capacity	0,070724	0,126077	0,015	0,0272	0,560959	0,581382
Naptha	0,098737	0,124436	0,183	0,2309	0,793476	0,437298

Рис. 3. Модель аналізу та коефіцієнт Beta.

Variable	Correlations (Spreadsheet12)			
	Ethylene	Capacity	Naptha	Prices
Ethylene	1,000000	0,097533	0,068105	-0,113693
Capacity	0,097533	1,000000	0,180412	-0,112283
Naptha	0,068105	0,180412	1,000000	-0,040182
Prices	-0,113693	-0,112283	-0,040182	1,000000

Рис. 4. Кореляційна матриця база даних з 3 незалежними змінними.

Regression Summary for Dependent Variable: Prices (Spreadsheet12)						
R= ,85325902 R ² = ,85348833 Adjusted R ² = ----						
F(3,20)=,16036 p<,92174 Std.Error of estimate: 27,002						
N=24	Beta	Std.Err. of Beta	B	Std.Err. of B	t(20)	p-level
Intercept			1358,809	323,2718	4,203300	0,000437
Ethylene	-0,102953	0,222320	-0,046	0,0988	-0,463084	0,048304
Capacity	-0,099496	0,225504	-0,021	0,0487	-0,441215	0,063790
Naptha	-0,015221	0,224951	-0,028	0,4175	-0,067662	0,946727

Рис. 5. Таблицю рівня значущості факторів.

Була виконана перевірка на мультиколінеарність: після ряду перетворень у даній базі мультиколінеарність не спостерігається (рис 6).

Variable	Correlations (Spreadsheet12)		
	Ethylene	Capacity	Prices
Ethylene	1,000000	0,097533	-0,113693
Capacity	0,097533	1,000000	-0,112283
Prices	-0,113693	-0,112283	1,000000

Рис. 6. Оновлена кореляційна матриця база даних з 2 незалежними змінними.

Наступний крок — це аналіз залишків. По-перше, побудуємо гістограму залишків та нормально-ймовірнісний графік залишків (рис. 7). Чим більше симетрична гістограма, тим точніше дані; в даному випадку гіпотеза норма-

льності не відхиляється. Нормально-ймовірнісний графік показує, що залишки розподілені нормально.

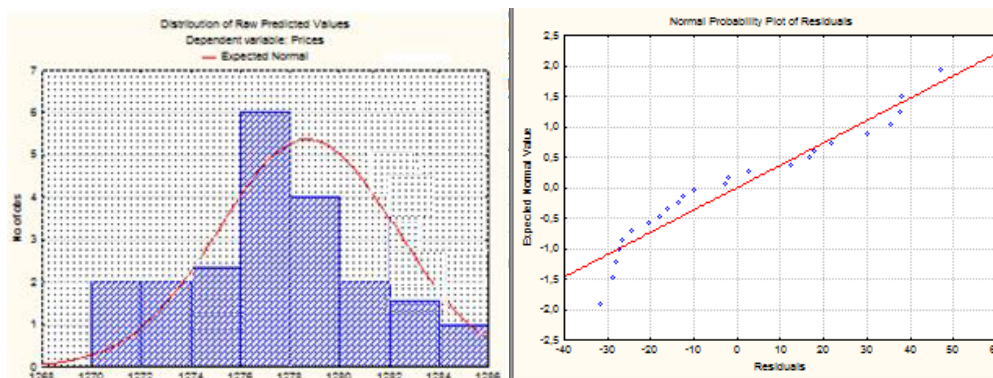


Рис. 7. Гістограма та нормально-ймовірнісний графік залишків.

Перевірка на наявність залежності залишків від прогнозованих значень наведена на рис 8. Можна побачити, що на діаграмі розсіювання не має ніякої системності і розташовані хаотично, саме тому залишки не залежать від прогнозованих значень. Аналіз виявив, що модель підходить для використання.

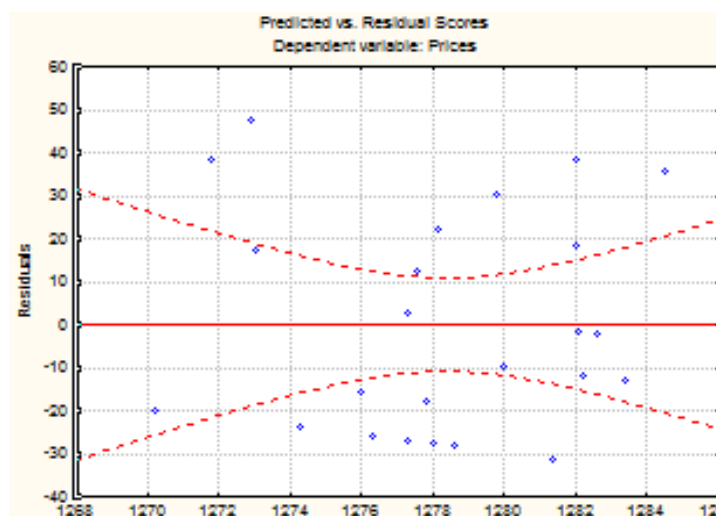


Рис. 8. Діаграма розсіювання залишків.

Оцінка моделі в цілому показує, що рівень значущості менше 0,05 (рис. 9), тому можна стверджувати, що модель прийнятна і буде працювати краще, ніж прогнозування по середнім значенням. Аналіз коефіцієнта детермінації показує, що 85% факторів враховано в моделі (рис. 3).

Перевірка побудови прогнозу за допомогою вводу параметрів ціни етилену та об'єму виробництва показує достовірність методу (рис. 10). Порівнюючи з дійсними даними, знаходимо: отримане значення несуттєво відрізняється від показників цін поліетилену.

Analysis of Variance; DV: Prices (Spreadsheet12)					
Effect	Sums of Squares	df	Mean Squares	F	p-level
Regress.	347,42	2	173,7105	0,250099	0,031010
Residual	14585,91	21	694,5673		
Total	14933,33				

Рис. 9. Діаграма розсіювання залишків.

Specify values for indep. vars		Predicting Values for (Spreadsheet12) variable: Prices		
Variable	B-Weight	Value	B-Weight * Value	
Ethylene	-0,046103	850,000	-39,188	
Capacity	-0,022076	1000,000	-22,076	
Intercept			1337,817	
Predicted			1276,554	
-95,0%CL			1263,356	
+95,0%CL			1289,752	

Рис. 10. Результати прогнозування.

Висновки . Отримані результати дозволили встановити, що методика, яка використовує кореляційно-регресійний аналіз, забезпечує високу точність розрахунків та може використовуватись в прогнозуванні цін на поліетилен. Розроблена методика може бути використана для підвищення точності розрахунків ціноутворення на європейському ринку поліетилену на основі даних про етилен та об'єму виробництва.

Бібліографічні посилання

1. **Draper, N.** Applied Regression Analysis, 3 Edition [Text] / N. Draper, H. Smith – In: Regression Analysis, 2016. – P. 912.
2. **Биткін, О.С.** Статистичне прогнозування цін на поліетилен з використанням макроекономічних тенденцій [Текст] / О.С. Биткін. – М., 2010. – С.175–189.
3. **Кривошеева, Е.С.** Економіко-математичні методи і прикладні моделі [Текст] / Е.С. Кривошеева. – 2007. – С. 451.
4. **Маланічев, А.Г.** Прогнозування світових цін на поліетилен [Текст] / А.Г. Маланічев. – М., 2012. – С. 108–135.
5. **Руденко, В.М.** Математична статистика [Текст] / В.М. Руденко. – Д. 273 с.
6. <https://chem-courier.ru/> – Бази даних аналітичного агентства «Хім-Кур'єр» на основі європейського ринку.

Надійшла до редколегії 13.04.2017