

Кузьмін В.М., Лапач С.М. «Полігональна регресія при наявності гетероскедастичності», «Економіка і управління» –2007, –№1. –С.81–86.

Описывается применение полигональной регрессии при наличии гетероскедастичности в данных. На примере показано, что учет гетероскедастичности повышает прогностические свойства модели

Кл. сл..

Полигональная регрессия, гетероскедастичность

Heteroskedasticity, Polygonal Regression

Кузьмін В.М., Лапач С.М.

Полігональна регресія при наявності гетероскедастичності в економічних задачах

Привид бродить по матстату – привид гетероскедастичності

При побудові регресійних моделей аналіз наявності гетероскедастичності та усуненню її розглядається як дуже важливий момент. Це зв'язано з тим, що об'єкти економічного дослідження в більшості випадків неоднорідні. Всі сучасні підручники з економетрики приділяють велику увагу цій проблемі [1–4, 8, 9]. Таким чином, гетероскедастичність економічних даних стала загальновідомим фактом і її необхідно враховувати в регресійному аналізі. При наявності гетероскедастичності оцінки коефіцієнтів рівняння регресії не будуть ефективними. Дисперсій оцінок будуть зміщеними, а висновки, які зроблені на основі t- та F-статистик – ненадійними [2]. Для виправлення ситуації використовують метод зважених найменших квадратів [2–4, 8, 9] або узагальнений метод найменших квадратів [1]. Але в економічних задачах дисперсії окремих точок спостережень, як правило, невідомі і виникає проблема, як знайти відповідні вагові коефіцієнти, які обернені до стандартних відхилень. Узагальнений метод найменших квадратів вимагає точного знання структури рівняння регресії [1]. Рекомендовані процедури підбору вагових коефіцієнтів неформалізовані [2–4, 8]. Крім того, не описано використання полігональної регресії, яка має суттєві переваги при інтерпретації та прогнозуванні економічних процесів [5, 6].

В статті розглядаються нові підходи до статистичної обробки економічних даних, що включають побудову рівняння гетероскедастичності та полігональну регресію при гетероскедастичних даних. Запропонований підхід дозволяє формалізовано будувати рівняння гетероскедастичності. Також вперше описується використання корегування гетероскедастичності по відношенню до полігональної регресії.

Вихідні дані

Для наочності використано приклад з книги [7]. Дані приведені в табл.1 і проілюстровані рис. 1.

Таблиця 1. Доходи та заощадження за Лізером

Рік	Доход	Заощадження
1946	8,8	0,36
1947	9,4	0,21
1948	10	0,08
1949	10,6	0,2
1950	11	0,1
1951	11,9	0,12
1952	12,7	0,41
1953	13,5	0,5
1954	14,3	0,43

1955	15,5	0,59
1956	16,7	0,9
1957	17,7	0,95
1958	18,6	0,82
1959	19,7	1,04
1960	21,1	1,53
1961	22,8	1,94
1962	23,9	1,75
1963	25,2	1,99

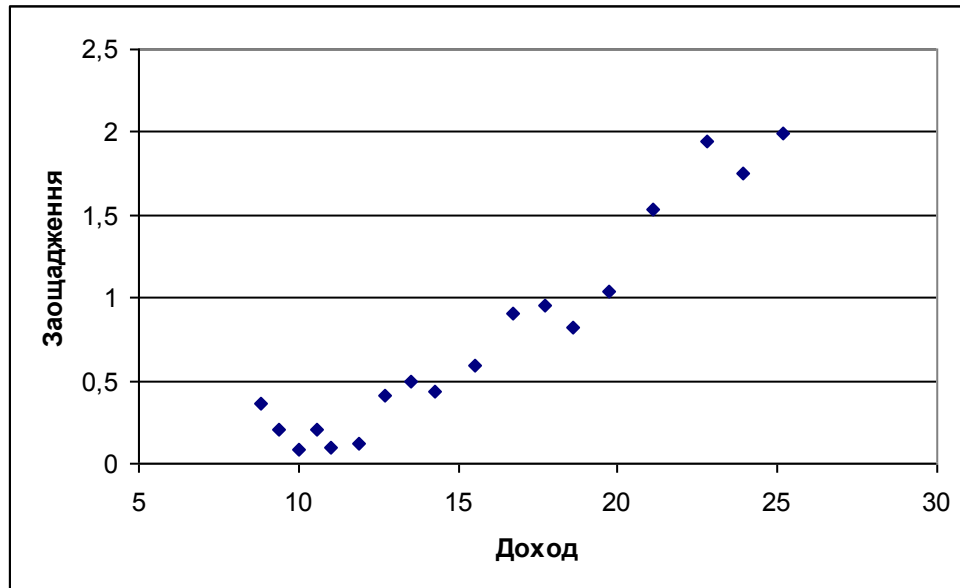


Рис. 1 Залежність між доходами та заощадженнями за табл. 1

В [7] для апроксимації використана пряма лінія, хоча з рис.1 легко побачити, що залежність нелінійна. Перевірка лінійності залежності за тестом [10] з довірчою ймовірністю 0,95 підтверджує візуальні спостереження.

Порівняльний аналіз моделей

За представленими даними побудовано три регресійні моделі

$$\text{Лінійна } Y = -1,082 + 0,118X \quad (1)$$

$$\text{Парабола } Y = 0,202 - 0,118X + 0,005X^2 \quad (2)$$

$$\text{Полігон } Y = 1,168 - 0,098X + 0,236(X-11,2)_+ \quad (3)$$

Детальний опис побудова полігональної моделі приведено в [5]. 11,2 – це абсциса точки

$$\text{перелому, а } (X-11,2)_+ = \frac{(X-11,2) + |X-11,2|}{2}.$$

В таблиці 2 приведені множинний коефіцієнт кореляції та залишкова дисперсія для вказаних моделей. На рис.2 представлені вихідні дані та графіки отриманих моделей.

Таблиця 2. Статистичні характеристики моделей

Вид моделі	Множинний коефіцієнт кореляції	Залишкова дисперсія
Лінійна	0,958	0,036

Парабола	0,976	0,022
Полігон	0,961	0,018

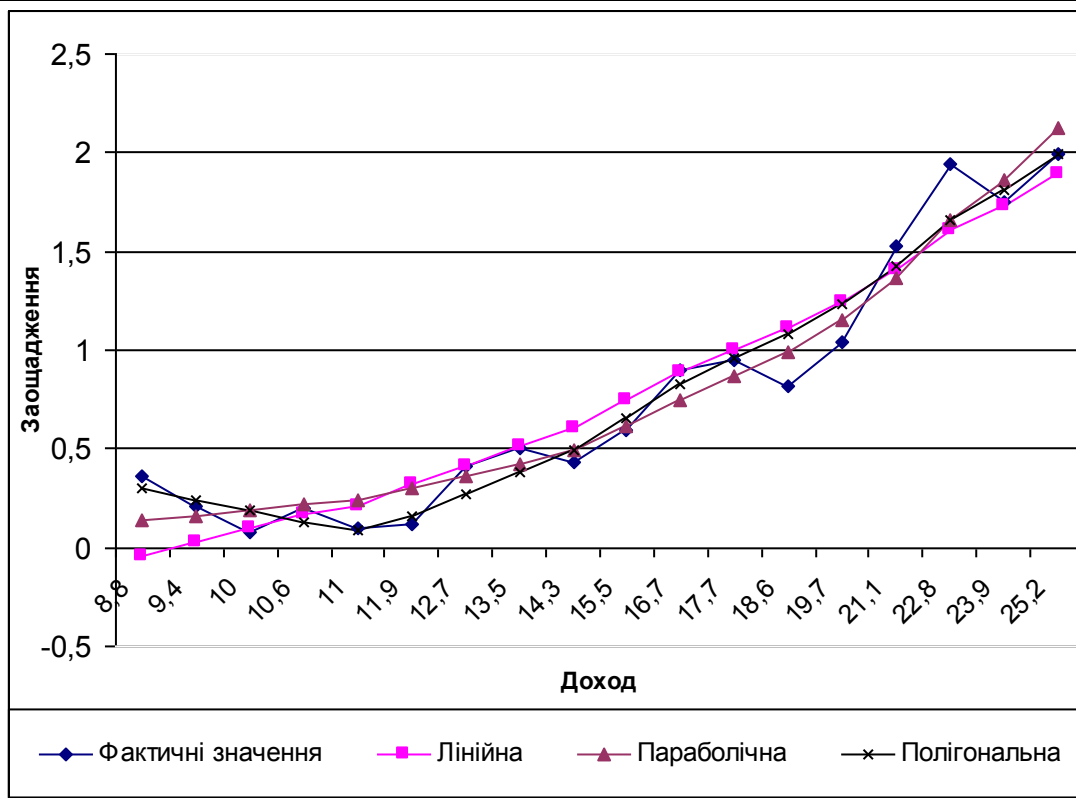


Рис.2 Залежність між доходом та заощадженнями (фактичні дані і моделі)

Параболічна та полігональна регресія апроксимують вихідні дані краще, ніж лінія, хоч формальний вибір зробити неможливо.

Гетероскедастичність

Тест, приведений в [11] підтверджує гетероскедастичність представлених даних. Для отримання рівняння гетероскедастичності, тобто залежності стандартного відхилення від незалежної змінної, виконується наступна послідовність дій.

1. Побудувати лінії регресії послідовно через кожні 5 точок. Тобто, спочатку через 1–4, потім 2–5 і т.д.
2. Знайти залишкову дисперсію для кожної ковзної регресійної моделі.
3. Знайти середнє значення незалежної змінної для кожної регресії.
4. Побудувати регресійну модель залежності стандартного відхилення від незалежної змінної. Для нашого випадку вона має наступний вигляд

$$\sigma = 0,073 + 0,055 \times Y \quad (4).$$

5. Знайти вагові коефіцієнти для кожної точки експерименту $\varpi_i = \frac{\bar{\sigma}^2}{\sigma^2(y_i)}$

Після цього знаходимо коефіцієнти полігональної регресії за звичайною схемою зваженого методу найменших квадратів:

$$Y = 1,054 - 0,087X + 0,220 (X-11,2)_+ \quad (5)$$

Графіки рівнянь регресії представлені на рис. 3.

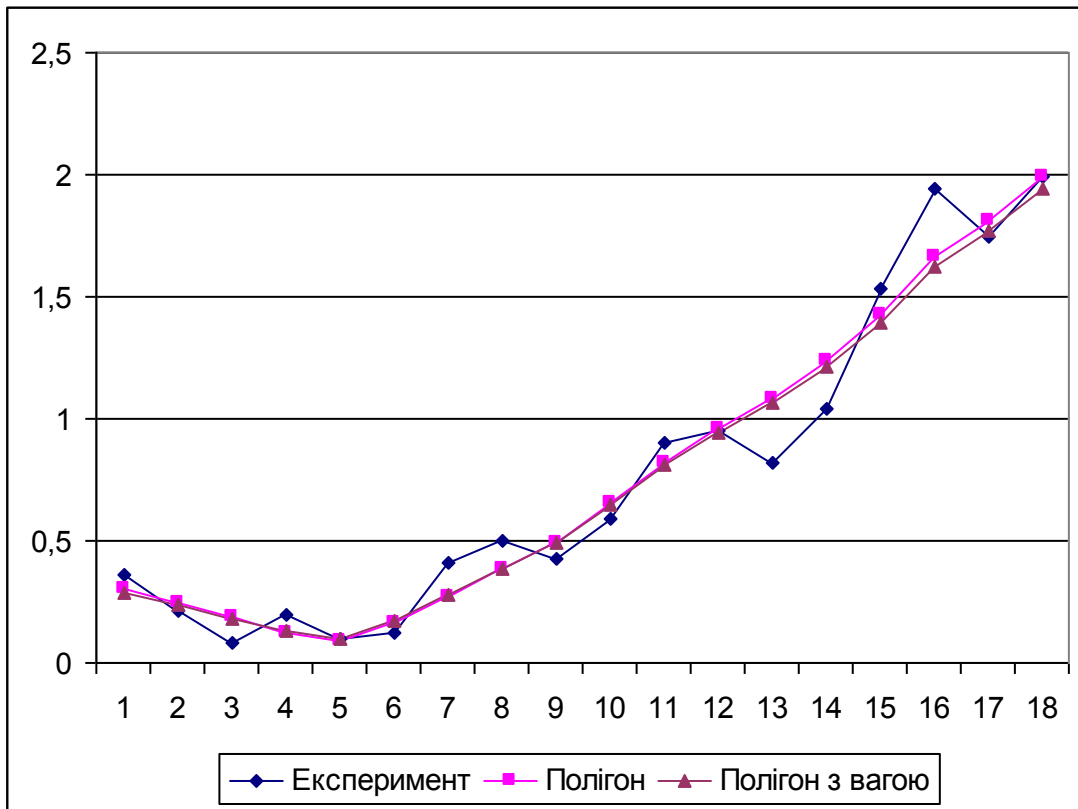


Рис.3. Графіки полігональних регресій

Скорегована з урахуванням гетероскедастичності модель більш точна. Це може не мати значення при апроксимації, але зіграти суттєву роль при прогнозуванні.

Послідовність побудови математичної моделі

- 1.Визначити можливість використання лінії для апроксимації.
2. Перевірити наявність гетероскедастичності в вихідних даних.
- 3.При виявленні гетероскедастичності побудувати рівняння гетероскедастичності і визначити вагові коефіцієнти для зваженого методу найменших квадратів.
- 4.Вибрати структуру моделі. Зауважимо, що це відповідальна фаза. І якщо для апроксимації параболічна і полігональна регресії можуть бути взаємозамінними, то для екстраполяції це невірно (див. рис.4).
- 5.Побудова завершальної регресійної моделі.

Висновки

В роботі пропонується алгоритм побудови регресійної моделі в умовах гетероскедастичності. Для кожного пункту алгоритму розроблені тести та процедури, що дозволяють формалізувати процес побудова та прийняття рішення. Використання полігональної регресії забезпечує підсилення прогностичних характеристики рівняння регресії при аналізі економічних даних.

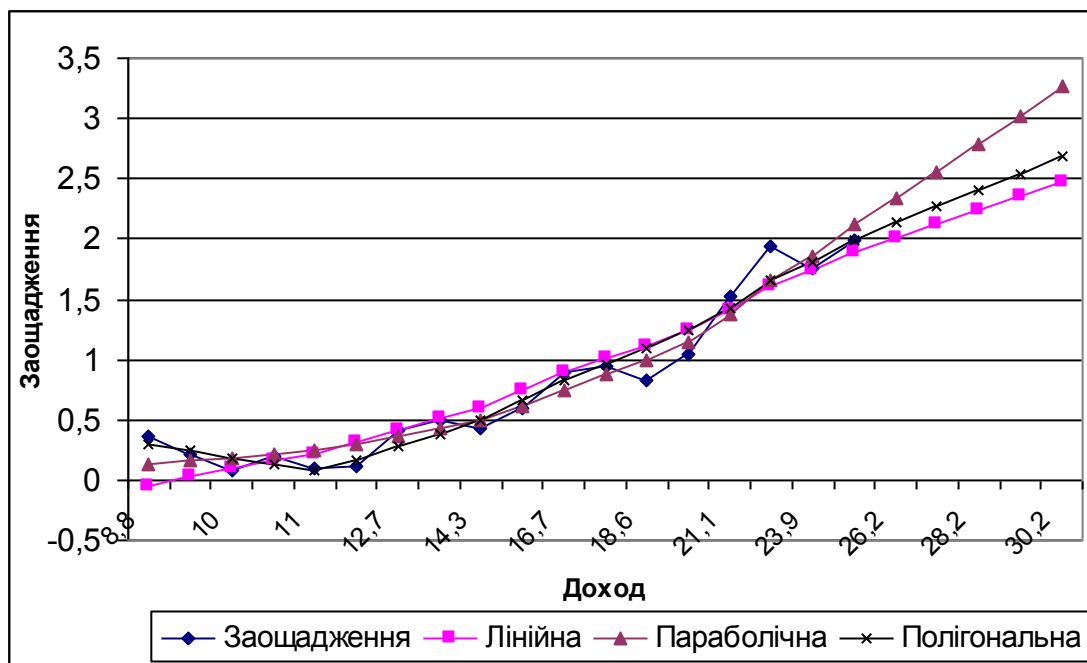


Рис. 4. Екстраполяція з використанням різних моделей

Література

1. Айвазян С.А. Прикладная статистика. Основы эконометрики: –Т.2, –М.: ЮНИТИ-ДАНА, 2001. –431с.
2. Бородич С.А. Эконометрика –Мн.: Новое знание, 2001. –408с.
3. Грубер Й. Эконометрия. Т.1 Введение в эконометрию –К.: Астарта, 1996. –398с.
4. Доугерти К. Введение в эконометрику –М.: ИНФРА-М, 2001. –402с.
5. Кузьмін В.М., Лапач С.М. Використання полігональної регресії в економічних дослідженнях // Економіка і управління –2004, –№3. –С. 79–84.
6. С.Н. Лапач, А.В. Чубенко, П.Н. Бабич Краткосрочное прогнозирование. Общий подход / Провизор, №8, 2004, С.16–18.
7. Лизер С. Эконометрические методы и задачи –М.: Статистика, –1971. –142с.
8. Магнус Я.Р., Катышев П.К., Пересецкий А.А. Эконометрика. Начальный курс –5-е изд., испр. –М.: Дело, 2001. –400с.
9. Назаренко О.М. Основы эконометрики –К.: Центр навчальної літератури, 2004. –392с.
10. Kuzmin V.N. New Statistical Method for Identification of Nonlinearity of Empirical Data / p.159-164 Computer data analysis and modeling v.1 – Minsk: 1998. /Proceeding of Firth International Conference (June 8-12, 1998, Minsk)
11. Kuzmin V.N. The Statistical Analysis of Econometric Data Under Heteroskedasticity / p.37-42 Computer data analysis and modeling v.2 – Minsk: 2001. /Proceeding of Sixth International Conference (September 10-14, 2001, Minsk)