УДК 519.21

ПОБУДОВА ЛІНІЙНОЇ РЕГРЕСІЙНОЇ МОДЕЛІ М-ТИПУ ДЛЯ ВИПАДКОВИХ ЧАСОВИХ РЯДІВ

Сіроштан Сергій

Науковий керівник: доктор.-ф.-м. наук, професор Авраменко О.В.

Центральноукраїнський державний педагогічний університет імені Володимира Винниченка, м. Кропивницький, Україна

В статті побудоване рівняння лінійної регресії для випадкових часових рядів. Акцент здійснюється на випадкові часові ряди, які представляють собою послідовні значення випадкової послідовності, що містить як попарну так і мішану кореляцію між своїми членами. Побудова регресійної моделі здійснюється на основі мінімізації суми квадратів відхилень між фактичними значеннями та значеннями, породженими лінійною прогностичною моделлю. Відповідні результати поглиблюють класичну лінійну регресійну модель в контексті появи останньої у випадку випадкової послідовності з виродженими за розподілом доданками.

Ключові слова: регресія, часовий ряд, випадковий ряд, функція розподілу, дисперсія, математичне сподівання.

Construction of a linear regression of M-type for random time series

S Siroshtan

Scientific supervisor: Doctor of Physics and Mathematics Science Avramenko O.V.

Volodymyr Vynnychenko Ukrainian State Pedagogical University, Kropyvnystky, Ukraine

The linear regression equation for random time series is constructed in the article. The emphasis is on random time series, which are sequential values of a random sequence that contains both pairwise and mixed correlations between its members. The construction of the regression model is based on minimizing the sum of squares of deviations between the actual values and the values generated by the linear predictive model. The corresponding results *deepen the classical linear regression model in the context of the appearance of the latter in the case of a random sequence with degenerate terms.*

Keywords: regression, time series, random series, distribution function, variance, mathematical expectation.

1. Постановка проблеми.

Одним з класичних понять регресійного аналізу є поняття часового ряду. Часовий ряд – це ряд точок даних, проіндексованих (або перелічених, або відкладених на графіку) в хронологічному порядку. Найчастіше часовий ряд є послідовністю, взятою на рівновіддалених точках в часі, які йдуть одна за одною. Таким чином, він є послідовністю даних дискретного часу.

Аналіз часових рядів включає методи аналізу даних часових рядів з метою витягування значимих статистик та інших характетистик даних. Прогнозування часових рядів – це застосування моделі для передбачування майбутніх значень на основі значень попередньо спостережених.

Цілком природним об'єктом, який є природним узагальненням часового ряду є випадковий ряд. Для випадкового ряду також можливо розглянути поняття рівняння лінійної регресії, чому присвячена відповідна робота.

Об'єкт дослідження: лінійна регресійна модель для випадкових часових рядів.

Предмет дослідження: лінійна регресійна модель М-типу для випадкових часових рядів.

Мета дослідження: побудова лінійної регресійної моделі М-типу для випадкових часових рядів.

Результати дослідження можуть бути використані в статистичному аналізі а також при викладанні спец курсу з теорії ймовірностей та математичної статистики.

2. Побудова ймовірнісної лінійної регресійної моделі.

Розглянемо випадкові послідовності, які представляють ймовірнісний аналог часових рядів.

Розглянемо випадкову послідовність:

$$\xi_1, \xi_2, \dots, \xi_n, \dots$$

та відповідно випадкову послідовність

$$\tau_1, \tau_2, \dots, \tau_n, \dots$$

Знайдемо рівняння лінійної регресії

$$y = kx + b,$$

де величини k та b є деякими дійсними числами. Потрібно відзначити, що можливо розглянути модель, в якій величини k та b можуть бути деякими випадковими величинами, на які накладаються певні обмеження (приналежність до певного класу випадкових величин, обмеження на числові характеристики тощо).

Для початку розглянемо випадок $k, b \in R$. Отже, побудуємо рівняння лінійної регресії випадкового набору значень

$$\xi_1, \xi_2, \dots, \xi_n$$

на випадковий набір значень

$$\tau_1, \tau_2, \dots, \tau_n.$$

Знайдемо відхилення між регресійними та фактичними значеннями за наступною формулою:

$$\varepsilon_j = k\xi_j + b - \tau_{j.}$$

Загальне сумарне відхилення, яке дорівнює сумі квадратів відхилень між регресійними та фактичними значеннями за наступною формулою:

$$\Delta = \varepsilon_1^2 + \dots + \varepsilon_n^2$$

Зрозуміло, що сумарне відхилення Δ є випадковою величино, тому потрібно формалізувати процес її мінімізації.

Розглянемо M-модель, в якій потрібно мінімізувати математичне сподівання від випадкової величини Δ.

Формально задачу мінімізації можливо записати у вигляді:

$$M(\Delta) \to min,$$

або в іншому вигляді:

$$f(k;b) = M\left(\sum_{j=1}^{n} \varepsilon_j^2\right) \to min.$$

Оскільки математичне сподівання скінченної або нескінченної суми випадкових величин дорівнює сумі математичних сподівань відповідних доданків (безумовно при умові, що математичні сподівання від відповідних випадкових величин існують), то маємо:

$$M(\mu_1 + \mu_2 + \cdots) = M(\mu_1) + M(\mu_2) + \cdots$$

Потрібно відзначити, що в останній рівності доданки можуть бути довільними в розумінні залежності чи незалежності.

Також, в подальшому буде використано наступну формулу:

$$M(C\varphi) = CM(\varphi), C - const.$$

Отже, маємо:

$$f(k;b) = \sum_{j=1}^{n} M(\varepsilon_j^2).$$

Знайдемо величини

$$M(\varepsilon_j^2).$$

Використаємо класичну комбінаторну формулу:

$$\left(\sum_{j=1}^{r} b_j\right)^2 = \sum_{j=1}^{r} b_j^2 + 2\sum_{l<i} b_l b_i$$

яка випливає з поліноміальної теореми (формули):

$$(c_1 + \cdots + c_r)^n = \sum_{i_1+\cdots+i_r=n} P(i_1; \dots; i_r) c_1^{i_1} \cdot \dots \cdot c_r^{i_r}$$

де

$$P(i_1; \dots; i_r) = \frac{n!}{i_1! \dots i_r!}$$

при $n = 2$ маємо:

$$M(\varepsilon_j^2) = M\left(\left(k\xi_j + b - \tau_j\right)^2\right) =$$

$$= M\left(k^2\xi_j^2 + b^2 + \tau_j^2 + 2kb\xi_j - 2k\xi_j\tau_j - 2b\tau_j\right) =$$

$$= k^2M\left(\xi_j^2\right) + 2kbM\left(\xi_j\right) - 2kM\left(\xi_j\tau_j\right) - 2bM\left(\tau_j\right) + b^2 + M(\tau_j^2)$$

Отже, сумарне відхилення дорівнює.

$$f(k; b) = \sum_j (k^2M\left(\xi_j^2\right) + 2kbM\left(\xi_j\right) - 2kM\left(\xi_j\tau_j\right) - 2bM\left(\tau_j\right) + b^2 + M(\tau_j^2))$$

Зрозуміло, що функція

$$f(k; b) = \sum_j (k^2M\left(\xi_j^2\right) + 2kbM\left(\xi_j\right) - 2kM\left(\xi_j\tau_j\right) - 2bM\left(\tau_j\right) + b^2 + M(\tau_j^2))$$

є неперервною.

Також, функція $f(k; b)$ обмежена знизу, тобто існує таке M, що

$$f(k; b) > M, \forall k, b \in R.$$

Відповідний факт випливає з того, що виконується нерівність

$$t^2 \geq 0, \forall t \in R.$$

тобто в нашому випадку

$$M\left(\left(k\xi_j + b - \tau_j\right)^2\right) =$$

$$= k^2 M\left(\xi_j^2\right) + 2kbM\left(\xi_j\right) - 2kM\left(\xi_j\tau_j\right) - 2bM\left(\tau_j\right) + b^2 + M(\tau_j^2) \geq 0$$

Параметри k, b класично знаходимо з умови:

$$\begin{cases} \dfrac{d(f)}{d(k)} = 0 \\ \dfrac{d(f)}{d(b)} = 0 \end{cases}$$

Отже, маємо:

$$\frac{d(f(k;b))}{d(k)} = \sum_j \left(2kM\left(\xi_j^2\right) + 2bM\left(\xi_j\right) - 2M\left(\xi_j\tau_j\right)\right) = 0$$

$$\frac{d(f(k;b))}{d(b)} = \sum_j (2kM\left(\xi_j\right) - 2M\left(\tau_j\right) + 2b) = 0$$

Перетворимо відповідну систему в більш зручній формі:

$$\begin{cases} k\sum_j M\left(\xi_j^2\right) + b\sum_j M\left(\xi_j\right) = \sum_j M\left(\xi_j\tau_j\right) \\ k\sum_j M\left(\xi_j\right) + bn = \sum_j M\left(\tau_j\right) \end{cases}$$

Розв’яжемо відповідну систему використовуючи правило Крамера:

$$\begin{cases} ax + by = c \\ dx + ex = f \end{cases}$$

то розв’язки мають вигляд:

$$x = \frac{\begin{vmatrix} c & b \\ f & e \end{vmatrix}}{\begin{vmatrix} a & b \\ d & e \end{vmatrix}}$$

$$y = \frac{\begin{vmatrix} a & c \\ d & f \end{vmatrix}}{\begin{vmatrix} a & b \\ d & e \end{vmatrix}}$$

де

$$\begin{vmatrix} q & w \\ z & v \end{vmatrix} = qv - zw.$$

Маємо:

$$k = \frac{n \sum_j M(\xi_j \tau_j) - \sum_j M(\tau_j) \sum_j M(\xi_j)}{n \sum_j M(\xi_j^2) - (\sum_j M(\xi_j))^2}$$

$$b = \frac{\sum_j M(\tau_j) \sum_j M(\xi_j^2) - \sum_j M(\xi_j \tau_j) \sum_j M(\xi_j)}{n \sum_j M(\xi_j^2) - (\sum_j M(\xi_j))^2}$$

Висновки. Таким чином, для випадкових послідовностей можливо розглянути поняття лінійної регресійної моделі М-типу, за якої здійснюється мінімізація суми квадратів відхилень між фактичними значеннями та значеннями, породженими лінійною прогностичною моделлю. Цілком природними подальшими поглибленнями є поліноміальні регресійні моделі.

Список літератури

1. Валтер Я. Стохастические модели в экономике. − М.: Статистика, 1976. − 232 с.
2. Вентцель А. Д. Курс теории случайных процессов. – М.: Наука, 1975. − 320 с.
3. Гнеденко Б.В. Курс теории вероятностей. –М.: Наука, 1988. – 448 с.
4. Колмогоров А.Н. Основные понятия теории вероятностей. – М.:

Наука, 1974. –120 с.

5. Розанов Ю. А. Случайные процессы. Краткий курс. – М.: Наука, 1979. − 184 с .