ПРАКТИЧНА РОБОТА № 5

Тема: Тренд-аналіз випадкових процесів

Мета роботи: Зробити тренд-аналіз випадкового процесу зміни концентрації стронцію у Південному Бузі

Стислі теоретичні відомості

Головні визначення

Випадкова величина – величина, яка може прийняти те, чи інше значення, заздалегідь невідоме яке саме.

Випадкова функція – функція, яка унаслідок експерименту може прийняти той чи інший вигляд (заздалегідь невідомо який).

Випадковий процес – процес, який триває у часі і може прийняти той чи інший вигляд (заздалегідь невідомо який).

Конкретна реалізація випадкового процесу – це конкретний вигляд випадкового процесу. Зафіксуємо момент t_k, в цей момент випадковий процес перетворюється у випадкову величину $x_i(t_k)$.

Множина випадкових величини $\{x_j(t_k)\}$ для реалізацій (j – номер реалізації) у конкретний момент часу – **перетин** випадкового процесу у момент t_k.

Приклад випадкового процесу – життя:

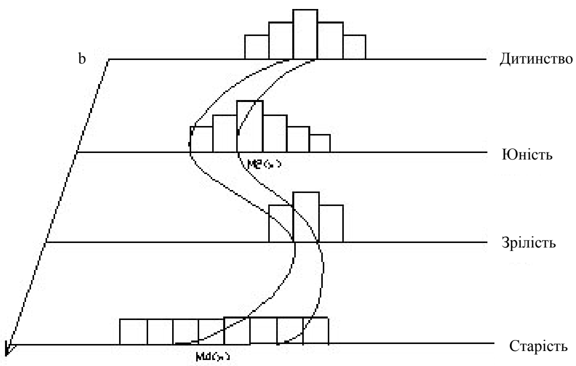

Рис. 12. Прожите життя – реалізація випадкового процесу

Часовий ряд – набір випадкових величин $x1, x2,..., x_n$ у момент часу $t1, t2,..., t_n$.

Часовий ряд – це точки реалізованого процесу.

Мета аналізу часових рядів:

– опис особливостей;
– вибір моделі;
– прогноз майбутнього;
– керування випадковим процесом.

Етап аналізу часових рядів:

– графічне зображення;
– вибір математичної моделі процесу;
– визначення і відокремлення складових, які залежать від часу;
– визначення відокремлення коливань;
– визначення випадкової компоненти
– знаходження математичних моделей складових процесу;
– прогноз на майбутнє.

Характеристики випадкового процесу

1. Математичне сподівання випадкового процесу у момент t_k:

– для дискретних випадкових величин

$$M(X(t_k)) = \sum_{i=1}^{N} p_i * X_i(t_k),$$

де i – номер реалізації випадкового процесу;

– для безперервних випадкових величин

$$M(X(t_k)) = \int_0^{\infty} f(x, t_k) * x(t_k) * dx.$$

Точкова оцінка математичного сподівання

$$M^*(X(t_k)) = \frac{1}{N} \sum_{i=1}^{N} X_i(t_k),$$

де i – номер реалізації випадкового процесу.

2. Дисперсія випадкового процесу у момент t_k:

$$D(X(t_k)) = \sum_{i=1}^{N} p_i * (X_i(t_k) - M(X(t_k)))^2,$$

де i – номер реалізації випадкового процесу.

Для безперервних випадкових величин

$$D(X(t_k)) = \int_0^\infty f(x, t_k) * (x(t_k) - M(X(t_k)))^2 * dx.$$

3. Середнє квадратичне відхилення випадкового процесу у момент t_k:

$$\sigma(t_k) = \sqrt{D(X(t_k))}.$$

4. Кореляційна функція 2-х випадкових процесів відображає зв'язок (залежність) значень випадкового процесу $x(t_i)$ та значень випадкового процесу $y(t_i)$ один від одного:

$$Cor(X(t_i), Y(t_i)) = \frac{M\big((X(t_i) - M(X(t_i)) * (Y(t_i) - M(Y(t_i))\big)}{\sigma(X(t_i)) * \sigma(Y(t_i))}.$$

5. Автокореляційна функція випадкового процесу відображає зв'язок (залежність) значень одного випадкового процесу $x(t)$ в різні моменти часу t_i та $t_i + k$ (відстань у часі k має назву лаг):

$$Cor(X(t_i), X(t_i + k)) = \frac{M\big((X(t_i) - M(X(t_i)) * (X(t_i) - M(X(t_i + k))\big)}{\sigma(X(t_i)) * \sigma(X(t_i + k))}.$$

Часові ряди, тобто реалізації випадкового процесу, включають в себе 2 головні частини: невипадкову, детерміновану частину, яка обчислюється, як функція часу $d(t)$, та випадкову частину $X_в(t)$.

Тому визначають такі основні види моделей випадкового процесу:

$X(t) = d(t) + X_в(t)$ – адитивна модель,

$X(t) = d(t) \cdot X_в(t)$ – мультиплікативна модель.

У детермінованій компоненті виділяють 3 складові: тренд $tr(t)$, сезонну компоненту $S(t)$, циклічну компоненту $G(t)$.

При адитивній моделі $d(t) = tr(t) + S(t) + G(t)$, при мультиплікативній моделі $d(t)= tr(t) \cdot S(t) \cdot G(t)$.

Тренд $tr(t)$ – плавна, нециклічна компонента, яка враховує вплив довготривалих факторів і відображає суттєву природу процесу.

Моделі трендів:

$tr(t) = b_0 + b_1 \cdot t$ – лінійна;

$tr(t) = b_0 + b_1 \cdot t + b_1 \cdot t^2$ – поліном;

$tr(t) = e^{b0 + b1 \cdot t}$ – експоненціальна, де процес має помтійну швидкість приросту;

$tr(t) = b_0 \cdot x^{b1}$ – ступенева;

$tr(t) = b_0 \cdot b_1^x$ – показникові;

$tr(t) = e^{b_0 * b_1^t}$ та $tr(t) = \frac{1}{b_0 * b_1^x + \gamma}$, де $0 < b_1 < 1$, $\gamma > 0$, якщо тренд має S форму.

Сезонна компонента – описує поведінку факторів, які змінюються циклічно, але з відомим періодом. Сезонна компонента може мати період 12 місяців, 6 місяців і так далі.

Циклічна компонента – описує тривалі періодичні підйоми і спади, котрі можуть змінюватись по амплітуді і періоду, але період не відомо заздалегідь:

$$G(t) = \sum_{i=1}^{\infty} A_i \cdot \sin(i\varpi T + \theta_i).$$

Щоб визначити випадкову компоненту, бажано звести процес до стаціонарного та ергодичного, і зробити так, щоб:

$$M(t) = 0,\ Cor(X(t), X(t+k)) \to 0, \text{ при } k \to \infty.$$

Для цього визначають $d(t)$ і рахують $X_в(t) = x(t) - d(t)$ або $X_в(t) = x(t)/d(t)$. Після цього перевіряють, чи є решта $X_в(t)$ стаціонарним та ергодичним процесом. Якщо $X_в(t)$ – стаціонарний та ергодичний процес, то по виду автокореляційної функції $X_в(t)$ визначають її модель, роблять висновок про те, що тренд-аналіз зроблено вірно.

Тренд-аналіз випадкового процесу

1. *Визначення тренду*

Спочатку будується графік випадкового процесу. Потім вибирається модель: адитивна, або мультиплікативна, або змішана. Коефіцієнти тренду знаходяться методом найменших квадратів. Оберемо b_0, b_1 так, щоб

$$F(b_0, b_1) = \sum_{i=1}^{n} (X(t_i) - tr(t_i))^2 \to \min .$$

Для цього розв’язуємо систему рівнянь

$$\frac{\partial F(b_0, b_1)}{\partial b_0} = 0, \quad \frac{\partial F(b_0, b_1)}{\partial b_1} = 0.$$

2. *Визначення сезонної компоненти*

Середньосезонну компоненту визначають, усереднюючи значення сезонних компонент по кожному з місяців протягом декількох років спостереження:

– для адитивної моделі

$$\text{sez}_{\text{сз}}(t_i) = \frac{1}{m+1} * \sum_{k=0}^{m} \left[X(t_i + k * T) - tr(t_i + k * T)\right];$$

– для мультиплікативної моделі

$$\text{sez}_{\text{сз}}(t_i) = \frac{1}{m+1} * \sum_{k=0}^{m} \frac{X(t_i + k * T)}{tr(t_i + k * T)},$$

де T – період сезонної компоненти, i – номер періоду, що спостерігається (для поточного періоду $i = 0$), $m + 1$ – число періодів.

3. *Визначення випадкової компоненти*

Випадкова компонента обчислюється так:

– для адитивної моделі $x_{\text{вип}}(t_i) = X(t_i) - tr(t_i) - sez_{\text{ср}}(t_i)$;

– для мультиплікативної моделі

$$x_{\text{вип}}(t_i) = \frac{X(t_i)}{tr(t_i) * sez_{\text{ср}}(t_i)}.$$

Після цього перевіряють, чи є $X_{\text{в}}(t)$ стаціонарним процесом, і якщо $X_{\text{в}}(t)$ – стаціонарний процес, то по виду автокореляційної функції $X_{\text{в}}(t)$ визначають її модель.

Автокореляційна функція випадкової компоненти відображає залежність значення випадкової компоненти $X_{\text{в}}(t_i)$ від значення тієї ж випадкової величини $X_{\text{в}}(t_i + k)$, де k – відстань у часі:

$$Cor(X_в(t_i), X_в(t_i+k)) = \frac{M\big((X_в(t_i) - M(X_в(t_i)) * (X_в(t_i) - M(X_в(t_i+k))\big)}{\sigma(X_в(t_i)) * \sigma(X_в(t_i+k))}.$$

Оцінка автокореляційної функції

$$Cor(X_в(t_i), X_в(t_i+k)) = \frac{\sum_{i=1}^{N}(X(t_i) - M^*(X(t_i)))(X(t_i+k) - M^*(X(t_i+k)))}{\sqrt{\sum_{i=1}^{N}[X(t_i) - M^*(X(t_i))]^2 \cdot \sum_{i=1}^{N}[X(t_i+k) - M^*(X(t_i+k))]^2}}.$$

Стаціонарні випадкові процеси – випадкові процеси, в яких математичне сподівання та дисперсія постійні протягом часу, а автокореляційна функція $Cor(X(t_i), X(t_i + k))$ залежить від відстані у часі k, а не від того, де розташовано t.

Якщо при $k \to \infty$ $Cor(X(t_i), X(t_i + k)) \to 0$, то стаціонарний процес набуває ергодичної властивості і статистику можно набирати не по різних реалізаціях процесу, а по одній реалізації, яка достатньо довго триває у часі і яку можливо розглядати як декілька різних реалізацій процесу

Якщо при тренд-аналізі випадкового процесу правильно визначено тренд і вилучено з процесу, а також визначена і видалена сезонна компонента, то автокореляційна функція випадкової компоненти $Cor(X_в(t_i), X_в(t_i + k)) \to 0$ при $k \to \infty$.

Якщо при тренд-аналізі випадкового процесу неправильно визначено тренд або не вилучено з процесу, то $Cor(X_в(t_i), X_в(t_i + k))$ не згасає при $k \to \infty$, тобто довготривалий зв’язок залишився у тренді.

Якщо при тренд-аналізі випадкового процесу неправильно визначена або вилучена сезонна компонента, то $Cor(X_в(t_i), X_в(t_i + k))$ зростає скачками при k, що дорівнює періоду коливань, оскільки через період зв’язок поновлюється.

По вигляду автокореляційної функції можна визначити такі моделі випадкової компоненти випадкового процесу (рис. 13, 14, 15):

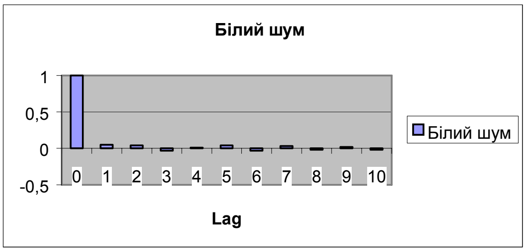

Рис. 13. Модель: $x_{вип}(t) = \varepsilon(t)$

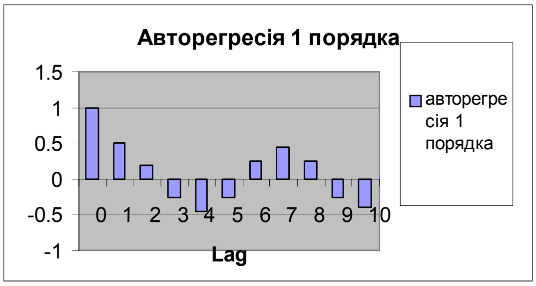

Рис. 14. Модель: $x_{вип}(t) = \varphi_1 * x_{вип}(t-1) + \varepsilon(t)$;
$\varphi_1 = Cor(X_в(t_i), (X_в(t_i + 1)$

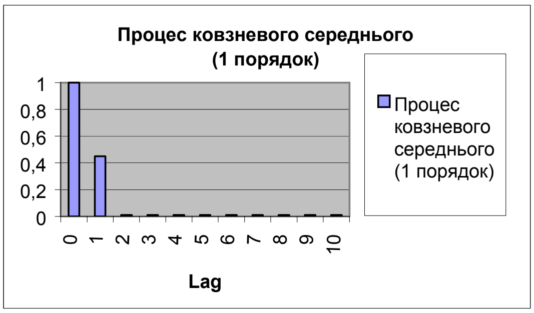

Рис. 15. Модель:

$$x_{вип}(t) = \theta_1 * \varepsilon(t-1) + \varepsilon(t); \qquad Cor(X_в(t_i), X_в(t_i+1)) = \frac{\theta_1}{1+\theta_1^2}.$$

Далі, маючи модель випадкового процесу і моделі усіх складових випадкового процесу, можна спрогнозувати випадковий процес для майбутніх значень *t*.

Приклад виконання тренд-аналізу у пакеті SPSS

– Побудова графіка: **Graphs** → **Scatter** → **Simple**.
– Обрати модель (адитивну або мультиплікативну). У даному прикладі модель мультиплікативна, т.я. розмах коливань збільшується при збільшенні тренду (рис. 16).

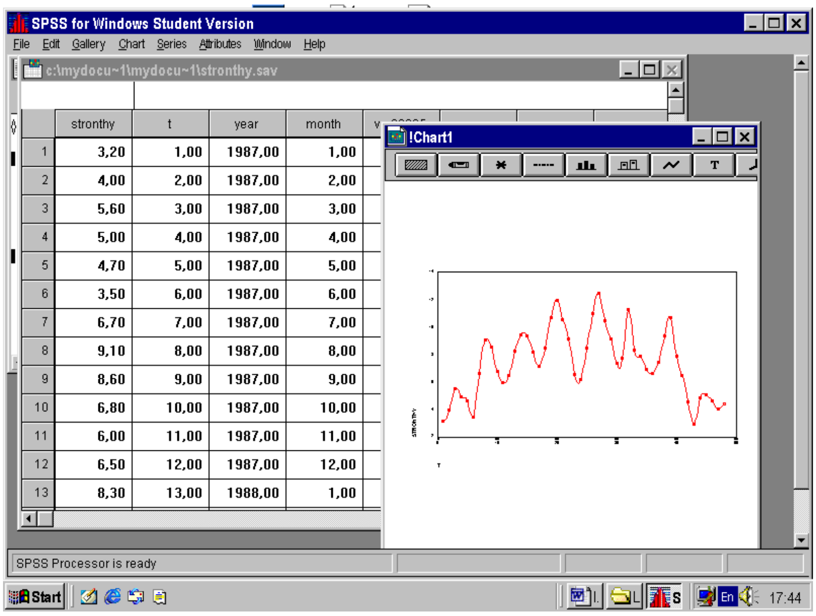

Рис. 16

– Знайти коефіцієнти тренду: **Statistic** → **Regression** → **Curve Estimation**. Обрати вид регресійної залежності (*x*-dependent, *t*-independent). Визначити коефіцієнти b_0, b_1, b_2 з протоколу (рис. 17, 18).

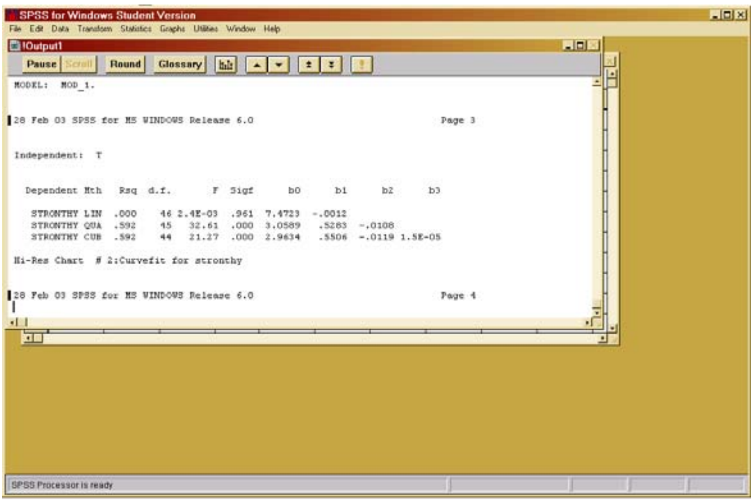

Рис. 17

Обираємо квадратичну регресійну модель: $b_0 = 3{,}0589$; $b_1 = 0{,}5283$; $b_2 = -0{,}0108$.

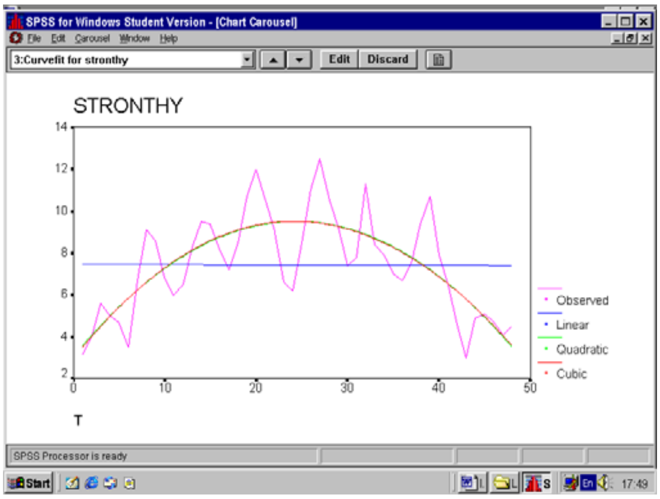

Рис. 18

Розрахувати тренд за коефіцієнтами та видом регресійної залежності: Transform → Compute (рис. 19).

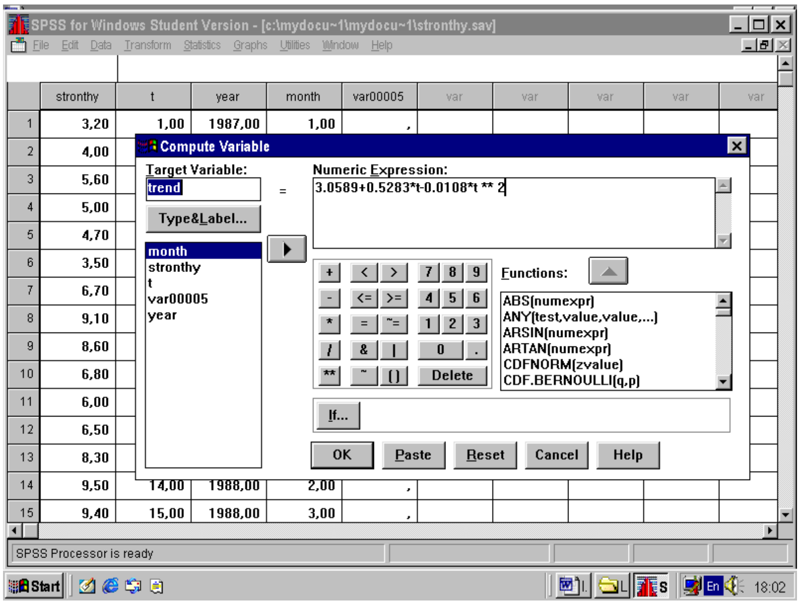

Рис. 19

– Розрахувати сезонну компоненту:
  Transform → Compute → Sezon = stronthy/trend.
– Розрахувати середньосезонну компоненту:
  Transform → Compute → Sezonsr = (sezon+LAG(sezon,12) +LAG(sezon,24)+LAG(sezon,36))/4.
– Розрахувати залишки (випадкова компонента):
  Transform → Compute → Random = stronthy/trend/sezonsr.
– Побудувати автокореляційну функцію для випадкової компоненти: **Graph → TimeSeries → Autocorrelation**. По графіку перевірити правільность тренд-аналізу та визначити модель сезонної компоненти (рис. 20, 21).

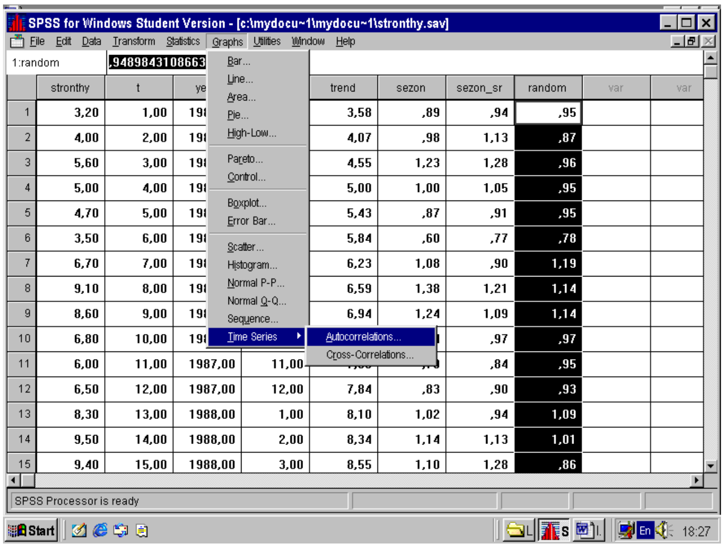

Рис. 20

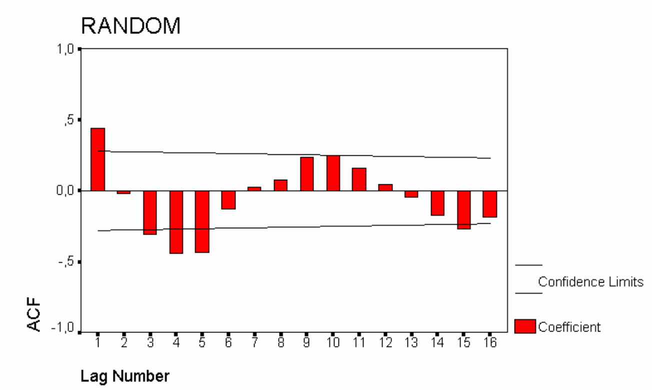

Рис. 21. Автокореляційна функція випадкової компоненти

Вигляд автокореляційної функції випадкової компоненти свідчить про те, що тренд-аналіз зроблено вірно.

Контрольні запитання

1. Що таке випадкова величина і випадковий процес?
2. Чисельні характеристики випадкового процесу: математичне сподівання та автокореляційна функція (визначення, графіки, на що вони вказують).
3. Що таке стаціонарні та ергодичні випадкові процеси?
4. Які складові визначаються у моделі випадкового процесу?
5. Яким чином визначається тренд у випадковому процесі, і що ми отримуємо в результаті тренд-аналізу?
6. Яким чином визначається сезонна компонента у випадковому процесі при мультиплікативній моделі процесу? Яким чином визначається сезонна компонента у випадковому процесі при адитивній моделі процесу?
7. Який вигляд мають графіки випадкового процесу при адитивній моделі і при мультиплікативній моделі процесу?
8. Що таке модель випадкової компоненти випадкового процесу, і для чого будується її автокореляційна функція?
9. Який вигляд має автокореляційна функція випадкового процесу, якщо не видалено тренду або сезонної компоненти?