WYKŁAD II 13.10.2014r.

LINIOWY MODEL Z JEDNĄ ZMIENNĄ OBJAŚNIAJĄCĄ

1. Regresja prosta.

Załóżmy, że jest liniowa zależność zmiennej objaśnianej Y od zmiennej objaśniającej X i od składnika losowego ε, tzn

$$Y = \alpha_0 + \alpha_1 X + \varepsilon$$

α_0, α_1 - nieznane parametry strukturalne modelu.

Celem jest oszacowanie nieznanych parametrów α_0, α_1 na podstawie posiadanych obserwacji zmiennej Y oraz zmiennej X.

Zakłada się, że dysponuje się n-elementowymi szeregami czasowymi obiektów w tym samym czasie (dla danych przekrojowych).

2. Oznaczenie obserwacji zmiennych w modelu

y_t - wartość zmiennej Y, w okresie t, albo wartości zmiennej Y dla obiektu o numerze t, $t = 1, 2, \ldots, n$.

x_t - wartość zmiennej X, w okresie t, albo wartości zmiennej X dla obiektów o numerze t, $t = 1, 2, \ldots, n$.

$$y = [y_1, y_2, \ldots, y_n]'$$

- wektor obserwacji zmiennej objaśnianej Y

$$\mathbf{X} = \begin{bmatrix} 1 & x_1 \\ 1 & x_2 \\ \ldots & \\ 1 & x_n \end{bmatrix}$$

- macierz zaobserwowanych wartości zmiennej objaśniającej.

3. Zapis macierzowy modelu

$$\varepsilon = [\varepsilon_1, \varepsilon_2, \dots \varepsilon_n]'$$

- wektor składników losowych

$$\alpha = [\alpha_0, \alpha_1]'$$

- wektor nieznanych parametrów modelu

Liniowy model ekonometryczny (1) w zapisie macierzowym ma postać:

$$y = X\alpha + \varepsilon \qquad (2)$$

4. Układ równań liniowych

W zapisie $y = X\alpha + \varepsilon$ kryje się n równań liniowych

$$\begin{cases} y_1 = \alpha_0 + \alpha_1 x_1 + \varepsilon_1 \\ y_2 = \alpha_0 + \alpha_1 x_2 + \varepsilon_2 \\ \dots \\ y_n = \alpha_0 + \alpha_1 x_n + \varepsilon_n \end{cases}$$

w których parametry strukturalne α_0, α_1 oraz składniki losowe $\varepsilon_1, \varepsilon_2, \dots, \varepsilon_n$ są nieznane i należy je oszacować.

Aby otrzymać dobre oszacowania nieznanych parametrów, trzeba przyjąć pewne założenia.

5. Założenia Gaussa - Markowa estymacji MNK

ZI. Zmienna objaśniająca X jest nieskorelowana ze składnikiem losowym ε.

ZII. Rząd macierzy $r(X) = 2 \leq n$.
Oznacza to, że nie wszystkie x_i są takie same.

ZIII. Przeciętna wartość każdego składnika losowego ε_i jest równa 0.

ZIV. Składniki losowe są nieskorelowane oraz wariancja każdego składnika losowego ε_i jest taka sama i wynosi δ^2.

ZV. Informacje zawarte w próbie $(y_1, x_1), (y_2, x_2), \ldots, (y_n, x_n)$ są jedynymi, na podstawie których estymuje się parametry strukturalne modelu $\mathcal{L}_0$ i $\mathcal{L}_1$.

ZVI. Składniki losowe ε_i mają rozkład normalny $N(0, \delta^2)$

Założenie ZVI nie jest konieczne. Przyjęcie go, powoduje, że otrzymane estymatory nieznanych parametrów modelu mają rozkład normalny.

Jednakowa wariancja wszystkich składników losowych oznacza, że model jest homoskedastyczny. Jeżeli wariancje składników losowych nie są takie same, to model nazywa się heteroskedastyczny.

Szacowanie nieznanych parametrów strukturalnych modelu metodą najmniejszych kwadratów przy założeniach ZI - ZV.

6. Regresja prosta z próby.

Prostą $Y = \mathcal{L}_0 + \mathcal{L}_1 X$ nazywa się prostą regresji lub prostą regresji z populacji.
Niech:

$\hat{\mathcal{L}}_0$ - oszacowanie parametru $\mathcal{L}_0$

$\hat{\mathcal{L}}_1$ - oszacowanie parametru $\mathcal{L}_1$

$\hat{\mathcal{L}} = \begin{bmatrix} \hat{\mathcal{L}}_0 \\ \hat{\mathcal{L}}_1 \end{bmatrix}$ - estymatory parametrów $\mathcal{L} = \begin{bmatrix} \mathcal{L}_0 \\ \mathcal{L}_1 \end{bmatrix}$

Prosta $\hat{y} = \hat{\alpha}_0 + \hat{\alpha}_1 x$ nazywa się prostą regresji z próby.

Wstawiając obserwacje $x_1, x_2, \ldots, x_n$ zmiennej x do równania prostej regresji z próby, otrzymujemy wartości $\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_n$, czyli

$$\hat{y} = \hat{\alpha}_0 + \hat{\alpha}_1 x_i, \qquad i = 1, 2, \ldots, n$$

Wartości $\hat{y}$ nazywamy wartościami teoretycznymi zmiennej objaśnianej Y.

7. Reszty modelu

Wartości $y_1, y_2, \ldots, y_n$ nazywamy wartościami zaobserwowanymi lub wartościami empirycznymi zmiennej Y

e_i – różnica między i-tą obserwacją zmiennej y i wartością regresji liniowej z próby w punkcie $x = x_i$, tzn

$$e_i = y_i - (\hat{\alpha}_0 + \hat{\alpha}_1 x_i)$$

gdzie $i = 1, \ldots, n$

Wielkość e_i nazywa się resztą modelu.

8. Interpretacja reszt i prostej regresji z próby

Wykres

9. Opis MNK

MNK – polega na wyznaczeniu oszacowań $\hat{\alpha}_0$ $\hat{\alpha}_1$ parametrów odpowiednio α_0, α_1 tak, aby suma kwadratów reszt była najmniejsza, tzn aby

$$\sum_{i=1}^{n} e_i^2 \text{ była najmniejsza.}$$

Czyli szukamy $\hat{\alpha}_0$ i $\hat{\alpha}_1$ takich, aby

$$\min_{(\alpha_0, \alpha_1)} \sum_{i=1}^{n} \left(y_i - (\alpha_0 + \alpha_1 x_i)\right)^2 = \sum_{i=1}^{n} \left(y_i - (\hat{\alpha}_0 + \alpha_1 x_i)\right)^2$$

Tak wyznaczone $\hat{\alpha}_0$, $\hat{\alpha}_1$ nazywają się estymatorami parametrów α_0 α_1 wyznaczonymi metodą najmniejszych kwadratów (EMNK)