



EKONOMETRIA

Temat wykładu:

Liniowy model ekonometryczny (regresji) z jedną zmienną objaśniającą

Prowadzący: dr inż. Zbigniew TARAPATA
e-mail: Zbigniew.Tarapata@isi.wat.edu.pl
http:// zbigniew.tarapata.akcja.pl/p_ekonometria/
tel.: 0-606-45-54-80

Plan wykładu

- Klasyczny model regresji liniowej – przypadek jednej zmiennej objaśniającej.
- Założenia schematu Gaussa-Markowa
- Metoda najmniejszych kwadratów.
- Przykład estymacji parametrów modelu

Model ekonometryczny - przypomnienie

Szukamy związku zmiennej objaśnianej Y_t z wartościami zmiennych objaśniających X_{it} , $i=1,...,m$, w chwili t .

Szukamy zatem związku f postaci:

$$Y_t = f(X_{1t}, X_{2t}, \dots, X_{mt})$$

Będziemy zakładać, że związek ten jest liniowy, zatem interesować nas będzie następująca zależność (dla jednej zmiennej objaśniającej tzn. gdy $m=1$):

$$Y_t = \alpha_0 + \alpha_1 X_{1t} + \xi_t \quad (*)$$

gdzie ξ_t obrazuje nam losowe odkształcenie modelu liniowego.

Jeżeli X_t oraz Y_t są zmiennymi losowymi (tak przyjmujemy w modelowaniu ekonometrycznym), to wtedy zależność (*) nosi nazwę tzw. **równania regresji**.

Liniowy model regresji - wprowadzenie

Procedura ekonometryczna:

Specyfikacja zmiennych → **Konstrukcja modelu** →
Estymacja parametrów → Weryfikacja modelu →
Zastosowanie modelu

Szukamy odpowiedzi na pytania:

- Co to jest regresja?
- Co to jest prosta regresji?
- Co to jest linia regresji?
- Czym różni się linia regresji I rodzaju od linii regresji II rodzaju?

Liniowy model regresji - wprowadzenie

Twierdzenie 1.

Współczynnik korelacji zmiennych losowych X i Y spełnia warunek:

$$|\rho_{XY}| = 1 \Leftrightarrow \exists_{\text{stała } a, b} P(Y = aX + b) = 1 \quad (1)$$

gdzie ρ_{XY} oznacza współczynnik korelacji między zmiennymi X i Y (oznaczany również przez r_{XY}).

Interpretacja: dla prawie wszystkich zdarzeń elementarnych zmienną losową Y można przedstawić jako funkcję liniową zmiennej losowej X .

Co to jest regresja?

- Jeżeli składowe wektora (X, Y) spełniają warunek (1), to prostą:

$$y = ax + b \text{ nazywamy } \underline{\text{prostą regresji}}$$

- Staje się ona linią regresji pierwszego rodzaju – linia regresji zmiennej losowej Y względem X :

$$y = h(x) = E(Y/X=x),$$

- Linia regresji zmiennej losowej X względem Y :

$$x = g(y) = E(X/Y=y)$$

- Jeżeli liniami są proste, to

$$h(EX) = EY, g(EY) = EX$$

Co to jest regresja?

- Jeżeli $|\rho| \neq 1$, to znaczy nie zachodzi równość: $P(Y=aX+b)=1$ i wtedy linia regresji pierwszego rodzaju nie jest prosta.
- Szukamy takiej funkcji liniowej, aby prawdopodobieństwo $P(Y=aX+b)$ było możliwie duże.
- Przyjmuje się kryterium: oczekiwany kwadratowy błąd aproksymacji:
$$e = E[(Y - aX - b)^2]$$
- Wartości a i b , dla których e jest minimalne wyznaczają prostą regresji II rodzaju.
- Jeśli $|\rho|=1$, to linie regresji I i II rodzaju pokrywają się.

Założenia schematu regresji (Gaussa-Markowa)

Założenie1. Model jest niezmienniczy względem obserwacji

$$f_1 = f_2 = \dots = f_n = f$$

Czyli wzór (2) można zapisać:

$$y_t = f_t(x_t, \xi_t) = f(x_t, \xi_t), \\ t = 1, \dots, n$$

Założenie2. Model jest liniowy względem parametrów

$$y_t = \alpha_0 + \alpha_1 x_t + \xi_t, t = 1, \dots, n$$

Założenia schematu regresji (Gaussa-Markowa), c.d.

Założenie3. Zmienna objaśniająca jest nielosowa, jej wartości są ustalonymi liczbami rzeczywistymi – *warunek identyfikacji* :

$$E\{Y_t | X_t\} = E\{Y_t\}$$

$$D^2\{Y_t | X_t\} = D^2\{Y_t\}$$

Założenie4. Składnik losowy ma rozkład normalny:

$$\xi_t : N(\mu_\xi, \sigma_t)$$

Założenia schematu regresji (Gaussa-Markowa), c.d.

Założenie5. Występujące zakłócenia, które reprezentuje składnik losowy mają tendencję do wzajemnej redukcji: $E\xi_t = 0$

Założenie6. Składnik losowy jest sferyczny:
nie występuje autokorelacja składnika losowego

$$\text{Cov}(\xi_t, \xi_\tau) = 0, t \neq \tau$$

składnik losowy jest homoskedastyczny

$$D^2\xi_t = \sigma^2$$

Założenia schematu regresji (Gaussa-Markowa), c.d.

Założenie 7. Informacje zawarte w próbie są jedynymi informacjami, na podstawie których dokonuje się estymacji parametrów modelu (2).

Powyższe założenia definiują tzw.

standardowy model liniowy

Metoda najmniejszych kwadratów (MNK, KMNK)

- Rozważamy model (2) i spełnione są założenia 3-7 schematu Gaussa-Markowa;
- Wartość oczekiwana zmiennej objaśnianej równa jest zatem:

$$E(Y_t) = \alpha_0 + \alpha_1 x_t \quad (3)$$

- Powyższe równanie wyznacza **linię regresji populacji generalnej**;
- Parametry α_0 , α_1 są nieznane i podlegają oszacowaniu na podstawie próby statystycznej - otrzymamy **linię regresji próby**:

$$\hat{y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_t$$

Metoda najmniejszych kwadratów (MNK, KMNK)

- Estymatory $\hat{\alpha}_0, \hat{\alpha}_1$ są funkcjami zmiennych losowych Y_t ;
- Różnice $e_t = Y_t - \hat{y}_t$ nazywamy resztami modelu;
- Mamy 4 funkcje:
 - linia regresji populacji generalnej LRPG;
 - linia regresji próby LRP;
 - wartości empiryczne (populacja generalna) WEPG;
 - wartości empiryczne (próba) WEP;

Metoda najmniejszych kwadratów (MNK, KMNK)

- LRPG $EY_t = \alpha_0 + \alpha_1 x_t$
- LRP $\hat{y}_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_t$
- WEPG $Y_t = \alpha_0 + \alpha_1 x_t + \xi_t$
- WEP $y_t = \hat{\alpha}_0 + \hat{\alpha}_1 x_t + e_t$
- WEPG=WEP

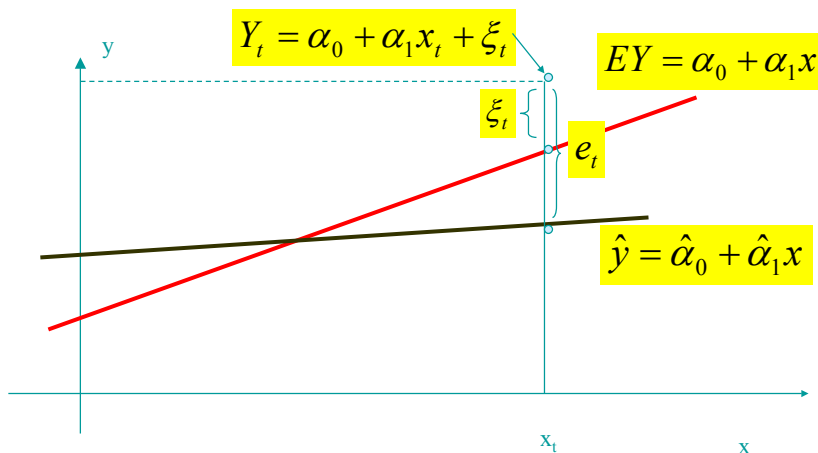
Metoda najmniejszych kwadratów (MNK, KMNK)

MNK

$$\left(\sum_{t=1}^n e_t^2\right) \rightarrow \min$$

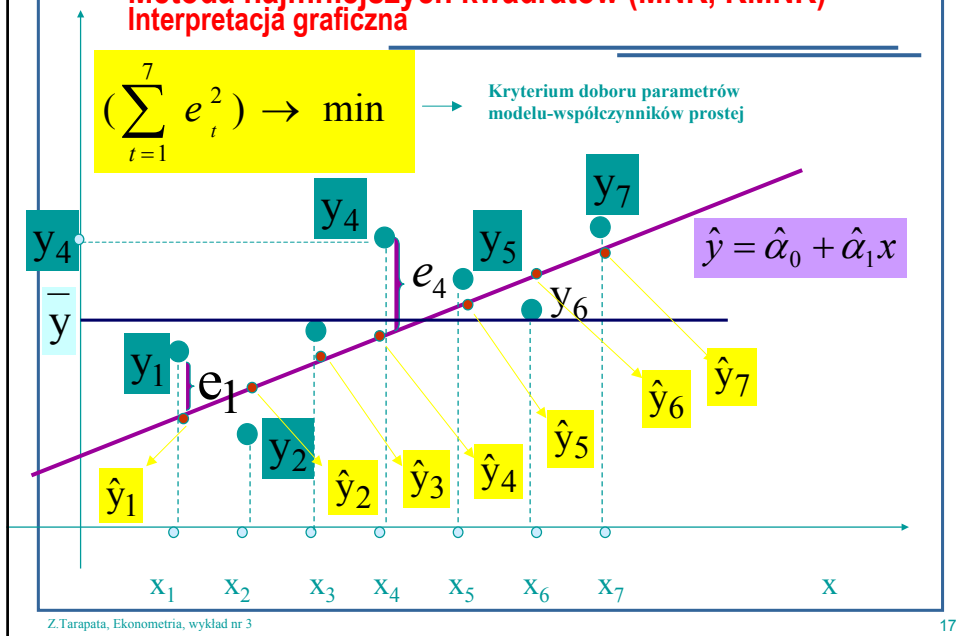
- Reszty e_t można traktować jako realizacje składnika losowego;
- Zadanie estymacji parametrów α_0, α_1 jest równoważne estymacji wartości oczekiwanej zmiennej objaśnianej na podstawie danych empirycznych ;
- Szukamy linii, która jest najlepsza z punktu widzenia sumy kwadratów reszt e_t (min) .

Metoda najmniejszych kwadratów (MNK, KMNK) Interpretacja graficzna



Metoda najmniejszych kwadratów (MNK, KMNK)

Interpretacja graficzna



17

Metoda najmniejszych kwadratów (MNK, KMNK)

Oszacowanie parametrów modelu

$\left(\sum_{t=1}^n e_t^2 \right) \rightarrow \min$ ← **Definicja problemu**
(Kryterium doboru parametrów)

$$\frac{\partial \left(\sum_{t=1}^n e_t^2 \right)}{\partial \hat{\alpha}_0} = \frac{\partial \left(\sum_{t=1}^n (y_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_t)^2 \right)}{\partial \hat{\alpha}_0} = 0,$$

$$\frac{\partial \left(\sum_{t=1}^n e_t^2 \right)}{\partial \hat{\alpha}_1} = \frac{\partial \left(\sum_{t=1}^n (y_t - \hat{\alpha}_0 - \hat{\alpha}_1 x_t)^2 \right)}{\partial \hat{\alpha}_1} = 0$$

Sposób
rozwiązania
problemu

18

Metoda najmniejszych kwadratów (MНК, KMNK)

Oszacowanie parametrów modelu

- Rozwiązując układ równań przedstawiony na poprzednim slajdzie otrzymujemy:

$$\hat{\alpha}_1 = \frac{\sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^n (x_t - \bar{x})^2}$$

$$\hat{\alpha}_0 = \bar{y} - \hat{\alpha}_1 \cdot \bar{x}$$

Liniowy model ekonometryczny z jedną zmienną objaśniającą

Przykład

Przykład

Weźmy pod uwagę następujący model:

$$y_t = \alpha_0 + \alpha_1 \cdot x_{1t} + \zeta_t$$

gdzie:

y_t – koszt produkcji pewnego wyrobu w chwili t ;

x_{1t} – wielkość produkcji w chwili t ;

Wartości zmiennych w kolejnych 10-ciu latach przedstawia tabela.

Oszacować parametry modelu na podstawie danych historycznych.

t	y_t	x_{1t}
1	1,4	1
2	1,7	1,5
3	1,5	1
4	2	1,5
5	2,2	2
6	2,3	2,5
7	2,5	2
8	2,8	3
9	2,6	3
10	2,9	3,5

Liniowy model ekonometryczny z jedną zmienną objaśniającą

Przykład, c.d.

t	y_t	x_t	$y_t - \bar{y}$	$x_t - \bar{x}$	$(x_t - \bar{x})^2$	$(x_t - \bar{x}) \cdot (y_t - \bar{y})$
1	1,4	1	-0,79	-1,1	1,21	0,869
2	1,7	1,5	-0,49	-0,6	0,36	0,294
3	1,5	1	-0,69	-1,1	1,21	0,759
4	2	1,5	-0,19	-0,6	0,36	0,114
5	2,2	2	0,01	-0,1	0,01	-0,001
6	2,3	2,5	0,11	0,4	0,16	0,044
7	2,5	2	0,31	-0,1	0,01	-0,031
8	2,8	3	0,61	0,9	0,81	0,549
9	2,6	3	0,41	0,9	0,81	0,369
10	2,9	3,5	0,71	1,4	1,96	0,994
	$\bar{y} = 2,19$	$\bar{x} = 2,1$		suma	6,9	3,96

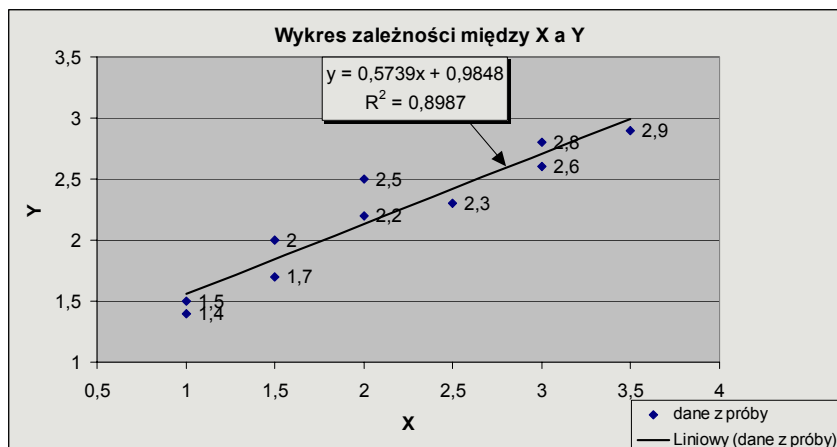
$$\hat{\alpha}_1 = \frac{\sum_{t=1}^n (x_t - \bar{x})(y_t - \bar{y})}{\sum_{t=1}^n (x_t - \bar{x})^2} = \frac{3,96}{6,9} = 0,57$$

$$\hat{\alpha}_0 = \bar{y} - \hat{\alpha}_1 \cdot \bar{x} = 2,19 - 0,57 \cdot 2,1 = 0,99$$

Zatem: $\hat{y}_t = 0,99 + 0,57 \cdot x_t$

Liniowy model ekonometryczny z jedną zmienną objaśniającą

Przykład, c.d.



Liniowy model ekonometryczny z jedną zmienną objaśniającą

Szacowanie stopnia dopasowania modelu do danych z próby (współczynnik R^2)

- Aby oszacować stopień dopasowania modelu regresji do danych z próby wyznacza się tzw. **współczynnik determinacji R^2** , który przyjmuje wartości z przedziału (0,1) (im R^2 bliższe 1 tym model lepiej dopasowany do próby):

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2}$$

Liniowy model ekonometryczny z jedną zmienną objaśniającą

Szacowanie stopnia dopasowania modelu do danych z próby (współczynnik R^2) - przykład

t	y_t	x_t	$\hat{y}_t$	$y_t - \bar{y}$	$e_t = y_t - \hat{y}_t$	$(y_t - \bar{y})^2$	$e_t^2 = (y_t - \hat{y}_t)^2$
1	1,4	1	1,56	-0,79	-0,16	0,6241	0,0256
2	1,7	1,5	1,845	-0,49	-0,145	0,2401	0,021025
3	1,5	1	1,56	-0,69	-0,06	0,4761	0,0036
4	2	1,5	1,845	-0,19	0,155	0,0361	0,024025
5	2,2	2	2,13	0,01	0,07	0,0001	0,0049
6	2,3	2,5	2,415	0,11	-0,115	0,0121	0,013225
7	2,5	2	2,13	0,31	0,37	0,0961	0,1369
8	2,8	3	2,7	0,61	0,1	0,3721	0,01
9	2,6	3	2,7	0,41	-0,1	0,1681	0,01
10	2,9	3,5	2,985	0,71	-0,085	0,5041	0,007225
$\bar{y} = 2.19$					suma	2,529	0,2565

$$R^2 = 1 - \frac{\sum_{t=1}^n (y_t - \hat{y}_t)^2}{\sum_{t=1}^n (y_t - \bar{y})^2} = 1 - \frac{0,2565}{2,529} \approx 0,9$$

Model bardzo dobrze dopasowany do danych, bo $R^2 \approx 0,9$