Krótkie wprowadzenie do ekonometrii

23 listopada 2011

Modele ekonometryczne są narzędziem umożliwiającym **ilościową** analizę zjawisk ekonomicznych. Modele te mają postać równań opisujących powiązania pomiędzy badanymi zmiennymi.

Wśród najważniejszych celów budowy modeli ekonometrycznych można wyróżnić:

- badanie zależności ekonomicznych (ocenę siły oddziaływania jednych kategorii ekonomicznych na drugie)
- prognozowanie

Konstrukcja najprostszego liniowego modelu ekonometrycznego z 1 zmienną:

$$y_t = \alpha_0 + \alpha_1 x_t + \epsilon_t$$

y_t - zmienna objaśniana, zwana również endogeniczną,
x_t - zmienna objaśniająca, zwana również egzogeniczną,
α_0, α_1- nieznane parametry (α_0- wyraz wolny, α_1- współczynnik kierunkowy)
ϵ_t- składnik losowy.

W modelu tym badamy wpływ zmiennej **objaśniającej** (**egozgenicznej**) na zmienną **objaśnianą** (**enodgeniczną**). Zmienna objaśniana jest więc skutkiem, a objaśniająca - przyczyną. Przykładowo może to być wpływ (1) zarobków na konsumpcję albo (2) wpływ ceny na popyt. Składnik losowy reprezentuje losowe, niewyjaśnione przez model zaburzenia, w pewnym sensie odstępstwa od normalnych praw ekonomicznych (zwykle nieobserwowalne - dlatego nie wiemy o nim nic przed badaniem). Jeśli myślimy o przykładzie konsumpcji (1), to może on reprezentować zmianę nawyków konsumpcyjnych, a w przykładzie popytu (2) może to być np. zmiana "mody" na kupno danego towaru. Oprócz takich czysto losowych i nieobserwowalnch zjawisk składnik losowy będzie zawierał również wszystkie inne przyczyny, których nie ma w modelu (np. wpływ wielkości posiadanego majątku na konsumpcję w (1), wpływ dochodów na popyt w (2) itp.). Parametry czyli siła z jaką zmienna objaśniająca wpływa na zmienną objaśnianą nie są znane. Model ekonometryczny jest narzędziem umożliwiającym oszacowanie tych parametrów (słowo „oszacowane" jest użyte, gdyż dokładne wartości nie zostaną poznane, możemy tylko z pewnym przybliżeniem je poznać).

W rzeczywistości żadne zjawisko ekonomiczne nie ma wyłącznie jednej przyczyny. Rozszerzając model przedstawiony powyżej o dodatkowe przyczyny otrzymujemy model z wieloma zmiennymi objaśniającymi.

$$y_t = \alpha_0 + \alpha_1 x_{1t} + \alpha_2 x_{2t} + ... + \alpha_K x_{Kt} + \epsilon_t$$

y_t - zmienna objaśniana (endogeniczna),
$x_{1t}, x_{2t}, ...$ - zmienne objaśniające (egzogeniczne),
$\alpha_0, \alpha_1, \alpha_2, ...$- nieznane parametry,
ϵ_t- składnik losowy.

W celu oszacowania parametrów modelu ekonometrycznego stosujemy **metodę najmniejszych kwadratów** (MNK). Zasada działania MNK jest następująca: wybieramy parametry, tak aby suma różnic pomiędzy rzeczywistą, obserowaną wartością y_t (tzw. wartością empiryczną) a wartością y_t wyznaczoną przez model - $\hat{y_t}$ (tzw. wartością teoretyczną) była jak najmniejsza (różnice te nazywamy **resztami**). Można powiedzieć, że w ten sposób wybieramy takie wartości parametrów, aby model był „najbliżej rzeczywistości“:

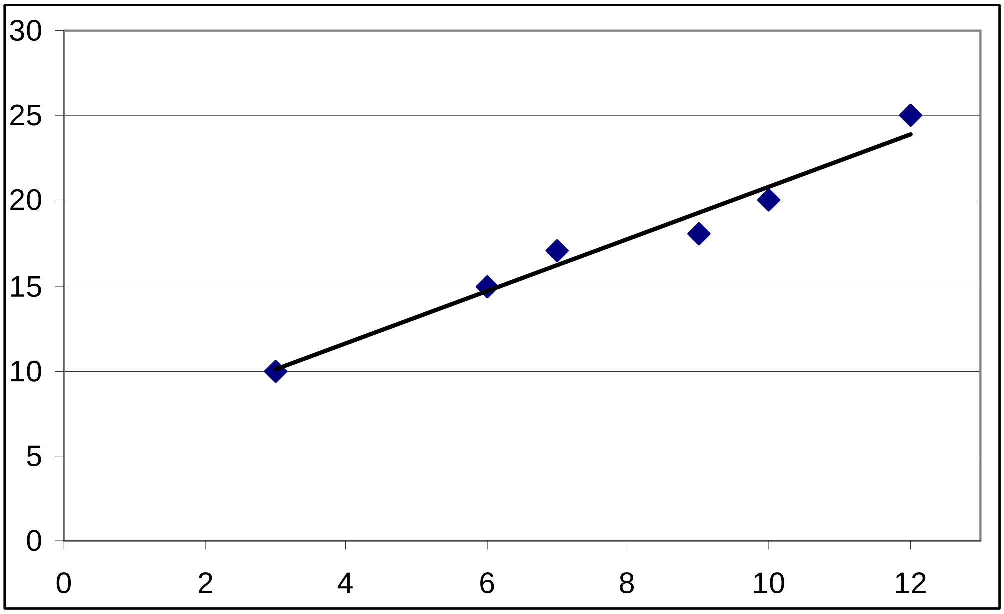
$$\sum (y_t - \hat{y_t})^2 \rightarrow min$$

Przykład, dzięki któremu można zrozumieć „intuicję“ metody MNK - dla uproszczenia model z 1 zmienną.

Wyobraźmy sobie model opisujący popyt na piwo (w sztukach miesięcznie) w zależności od rocznego dochodu (w tys. zł). Zebrano dane statystyczne - udało się przeankietować jedynie 6 studentów. Oto one:

nr obserwacji	1	2	3	4	5	6
y (popyt na piwo)	10	15	17	18	20	25
x (dochody roczne)	3	6	7	9	10	12

Umieszczając te punkty na wykresie możemy znaleźć linię „najbliżej“ (w sensie minimalnej sumy kwadratów reszt) - dla podanego przykładu jest to ta czarna prosta na wykresie.



Jak interpretujemy parametry? Wzrost wartości zmiennej objaśniającej o jednostkę powoduje zmianę wartości zmiennej objaśnianej o tyle, ile wynosi oszacowana wartość parametru przy tej zmiennej, przy założeniu że inne zmienne objaśniające nie uleganą zmianie (*ceteris paribus*).

Dla podanego przykładu z piwem model z oszacowanymi parametrami przedstawia się następująco (*wartości parametrów zostały zaokrąglone*):

$$y_t = 5 + 2 \cdot x_t$$

Interpretacja oszacowań parametrów przebiega następująco: wzrost x_t o jednostkę - czyli rocznych dochodów o 1 tys. zł, powoduje wzrost wartości y_t o 2 jednostki - czyli popytu na piwo o 2 szt. miesięcznie (nie potrzebujemy dodawać, że przy założeniu, iż inne zmienne objaśniające nie ulegną zmianie, gdyż tu jest jedynie 1 zmienna objaśniająca).

W celu oceny stopnia dopasowania modelu do danych („zgodności z rzeczywistością“) używamy najczęściej **współczynnika determinacji** (R^2). Jest on określony wzorem:

$$R^2 = 1 - \frac{\sum(y_t - \hat{y}_t)^2}{\sum(y_t - \overline{y})^2}$$

gdzie: $\overline{y}$ - średnia arytmetyczna zmiennej objaśnianej, pozostałe oznaczenia jak wyżej.

Interpretacja współczynnika determinacji - w jakim stopniu (tj. w ilu %) model wyjaśnia zmienność (wahania) zmiennej objaśnianej. Np. 0,93 oznacza, że model wyjaśnił 93% zmienności zmiennej objaśnianej. Generalnie im wyższa wartość tego współczynnika tym model lepszy (chociaż wysoki wsp. determinacji nie oznacza jeszcze, że model jest dobry; znaczenie ma np. zgodność modelu z teorią ekonomii itp.). Współczynnik R^2 przyjmuje wartości od 0 do 1.

Badanie istotności parametrów ma na celu sprawdzenie, czy wpływ danej zmiennej (lub zmiennych) jest statystycznie istotny. Innymi słowy, czy wpływ ten jest silny (istotny statystycznie) czy też pomijanie mały (nieistotny statystycznie).

W celu zbadania istotności parametru obliczamy **statystyki t-Studenta**:

$$t_{OBL} = \frac{a_i}{S_{a_i}}$$

gdzie:
a_i- oszacowana wartość parametru α_i,
S_{a_i}- średni błąd szacunku parametru α_i,
t_{OBL} - obliczona wartość statystyki t-Studenta.

W teście tym stawiamy hipotezy:

$$\mathbf{H_0} : \alpha_i = 0$$

$$\mathbf{H_1} : \alpha_i \neq 0$$

,

Brak podstaw do odrzucenia hipotezy zerowej $\mathbf{H_0}$ oznacza zatem, iż dana zmienna objaśniająca zmienna nie jest istotna statystycznie (nie wpływa silnie na zmienną objaśnianą). Z kolei odrzucenie hipotezy zerowej na rzecz hipotezy alternatywnej $\mathbf{H_1}$ oznacza, że zmienna istotnie statystycznie wpływa na zmienną objaśnianą.

W celu rozstrzygnięcia czy zmienna jest istotna czy nie porównujemy t_{OBL} z wartością krytyczną statystyki t-Studenta na przyjętym poziomie istotności t_{KRYT} (o N-K stopniach swobodby - gdzie N to liczba obserwacji, a K - liczba szacowanych parametrów[1]).

Wartość statystyki testowej a wartość krytyczna	Wniosek
$\lvert t_{OBL}\rvert \leq t_{KRYT}$	Nie można odrzucić $\mathbf{H_0}$ - parametr nieistotny
$\lvert t_{OBL}\rvert > t_{KRYT}$	Odrzucić $\mathbf{H_0}$ na korzyść $\mathbf{H_1}$ - parametr istotny

Trzy informacje praktyczne. Po pierwsze - wartość krytyczna na standardowym poziomie istotności 5% wynosi około 2, więc jeśli nie mamy tablic z wartościami krytycznymi, wówczas sprawdzać czy $|t_{OBL}| > 2$. Po drugie - jeżeli w modelu występuje zmienna nieistotna, wówczas powinniśmy usunąć ją z modelu albo poprawić model tak, aby była ona istotna. Po trzecie - (w zdecydowanej większości przypadków) nie testujemy istotności wyrazu wolnego (stałej).

[1] Wróćmy znów do przykładu z popytem na piwo. Mieliśmy 6 obserwacji, stąd $N=6$. Oszacowaniu podlegały dwa parametry - współczynnik przy zmiennej x oraz wyraz wolny, stąd $K=2$.