Wybór formy funkcyjnej modelu

Wykład 5

Natalia Nehrebecka Stanisław Cichocki

7 listopada 2015

Plan zajęć

Etapy doboru formy funkcyjnej
w praktyce

1. Dobór formy, tak aby możliwa była odpowiedź na pytanie badawcze
2. Szacunek modelu i ocena jakości na podstawie testów diagnostycznych

W praktyce wybiera się tę formę funkcyjną, która jest najsłabiej odrzucana przez testy diagnostyczne.

- Niech

$$y_i = f(\mathbf{x}_i, \theta, \epsilon_i) \tag{1}$$

będzie funkcją **nieliniową**.

- Pod pewnymi warunkami istnieją takie funkcje, g, h i β, że model (1) można sprowadzić do modelu **liniowego**:

$$g(y_i) = h(\mathbf{x}_i)'\beta(\theta) + \epsilon_i \tag{2}$$

- Oznaczając $g(y_i) = y_i^\star$, $h(\mathbf{x}_i) = \mathbf{x}_i^\star$ oraz $\beta(\theta) = \beta^\star$ mamy

$$y_i^\star = \mathbf{x}_i^\star \beta^\star + \epsilon_i \tag{3}$$

- Niech
$$y_i = f(\mathbf{x}_i, \theta, \epsilon_i) \tag{1}$$
będzie funkcją **nieliniową**.
- Pod pewnymi warunkami istnieją takie funkcje, g, h i β, że model (1) można sprowadzić do modelu **liniowego**:
$$g(y_i) = h(\mathbf{x}_i)'\beta(\theta) + \epsilon_i \tag{2}$$
- Oznaczając $g(y_i) = y_i^\star$, $h(\mathbf{x}_i) = \mathbf{x}_i^\star$ oraz $\beta(\theta) = \beta^\star$ mamy
$$y_i^\star = \mathbf{x}_i^\star \beta^\star + \epsilon_i \tag{3}$$

- Niech
$$y_i = f(\mathbf{x}_i, \theta, \epsilon_i) \tag{1}$$
będzie funkcją **nieliniową**.
- Pod pewnymi warunkami istnieją takie funkcje, g, h i β, że model (1) można sprowadzić do modelu **liniowego**:
$$g(y_i) = h(\mathbf{x}_i)'\beta(\theta) + \epsilon_i \tag{2}$$
- Oznaczając $g(y_i) = y_i^\star$, $h(\mathbf{x}_i) = \mathbf{x}_i^\star$ oraz $\beta(\theta) = \beta^\star$ mamy
$$y_i^\star = \mathbf{x}_i^\star\beta^\star + \epsilon_i \tag{3}$$

- Dobry model dobrze opisuje zjawisko i posiada parametry o intuicyjnie oczywistej interpretacji.
- Interpretacja wyników jest ważna, bo umożliwia porównanie wyników z teorią, zdrowym rozsądkiem, wynikami z innych źródeł itp.

Efekt cząstkowy
ang. *partial effect*

Efekt cząstkowy zmiana oczekiwanego y_i w reakcji na zmianę x_{ki}; wynosi $\frac{\Delta E(y)}{\Delta x_k}$ przy założeniu *ceteris paribus.*

W modelu liniowym

$$E(y_i) = \beta_1 x_{1i} + \beta_2 x_{2i} + \dots \beta_K x_{Ki}$$

o ile zmienne $x_{1i}, x_{2i}, \dots x_{Ki}$ oraz parametry $\beta_1, \beta_2, \dots, \beta_K$ są nielosowe.

Zatem **efekt cząstkowy** dla $\Delta x_k = 1$ równy jest β_k.

Wniosek

Parametr β_k w modelu **liniowym** opisuje zmianę oczekiwanej wartości y_i na skutek jednostkowej zmiany x_{ki}, przy założeniu, że wszystkie pozostałe zmienne nie ulegają zmianie.

Elastycznością cząstkową nazywamy procentową zmianę oczekiwanego y w reakcji na 1% zmianę x_k:

$$e_{y,x_k} = \frac{\Delta E(y)}{E(y)} \Big/ \frac{\Delta x_k}{x_k} = \frac{\Delta E(y)}{\Delta x_k} \frac{x_k}{E(y)} \approx \frac{\partial E(y)}{\partial x_k} \frac{x_k}{E(y)}$$

ale $\frac{\partial ln(\varphi(x))}{\partial \varphi(x)} = \frac{1}{\varphi(x)}$, stąd $\frac{\partial ln E(y)}{\partial E(y)} = \frac{1}{E(y)}$ oraz $\frac{x_k}{\partial x_k} = \frac{1}{\partial ln(x_k)}$, więc:

$$e_{y,x_k} = \frac{\partial ln(E(y))}{\partial ln x_k}.$$

W modelu logliniowym

$$lnE(Y_i) = \beta_1 lnX_{1i} + \beta_2 lnX_{2i} + \cdots + \beta_K lnX_{Ki}$$

wtedy

$$\frac{\partial lnE(Y_i)}{\partial lnX_{ki}} = \beta_k$$

o ile zmienne $x_{1i}, x_{2i}, \ldots x_{Ki}$ oraz parametry $\beta_1, \beta_2, \ldots, \beta_K$ są nielosowe.

Wniosek

Parametr β_k w modelu **logliniowym** jest elastycznością y względem x_k.

Semielastycznością cząstkową nazywamy procentową zmianę oczekiwanego y w reakcji na jednostkową zmianę x_k:

$$\ddot{e}_{y,x_k} = \frac{\Delta E(y)}{E(y)} \Big/ \Delta x_k \approx \frac{\partial E(y)}{\partial x_k} \frac{1}{E(y)} = \frac{\partial \ln E(y)}{\partial x_k}$$

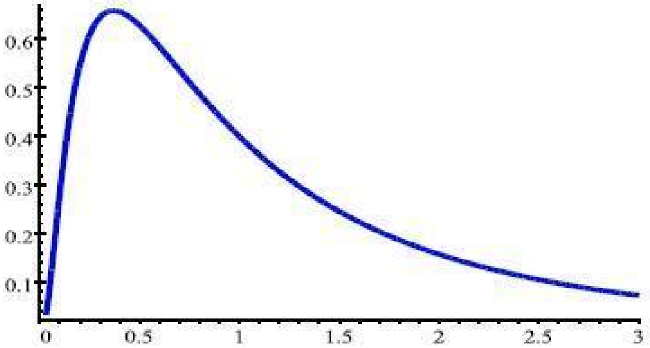
- Stosując model potęgowy logarytmujemy zmienną zależną jak i zmienne niezależne.
- Zmienna zależna i zmienne niezależne powinny przyjmować wartości dodatnie, ponieważ w innym przypadku nie da się ich zlogarytmować.

$$ln(Y_i) = \beta_1 lnX_{1i} + \beta_2 lnX_{2i} + \ldots \beta_K lnX_{Ki} + \varepsilon_i$$

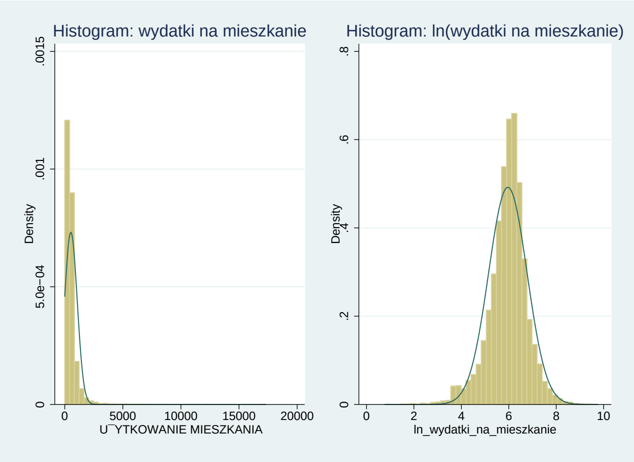
Wybór pomiędzy modelem liniowym a potęgowym

- W badaniach empirycznych często stwierdzamy, że zmienne w modelu mają rozkład zbliżony do normalnego bądź do lognormalnego.
- Jeśli ϵ ma rozkład lognormalny, to $ln(\epsilon)$ ma rozkład normalny.

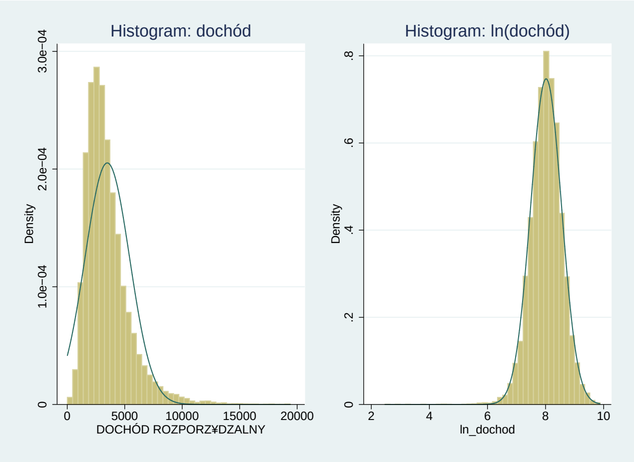
Rozkład lognormalny



Przykład: Wydatki mieszkaniowe gospodarstw domowych



Dochody gospodarstw domowych



Wyniki regresji: poziomy

```
reg w04 dochg  if typ==1

      Source |       SS       df       MS              Number of obs =   18328
-------------+------------------------------           F(  1, 18326) = 1196.98
       Model |   334324900     1   334324900           Prob > F      =  0.0000
    Residual |  5.1186e+09 18326  279307.612           R-squared     =  0.0613
-------------+------------------------------           Adj R-squared =  0.0613
       Total |  5.4529e+09 18327  297534.578           Root MSE      =   528.5

------------------------------------------------------------------------------
         w04 |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dochg |    .069476   .0020081    34.60   0.000     .0655398    .0734121
       _cons |   285.6029   8.012152    35.65   0.000     269.8983    301.3074
------------------------------------------------------------------------------
```

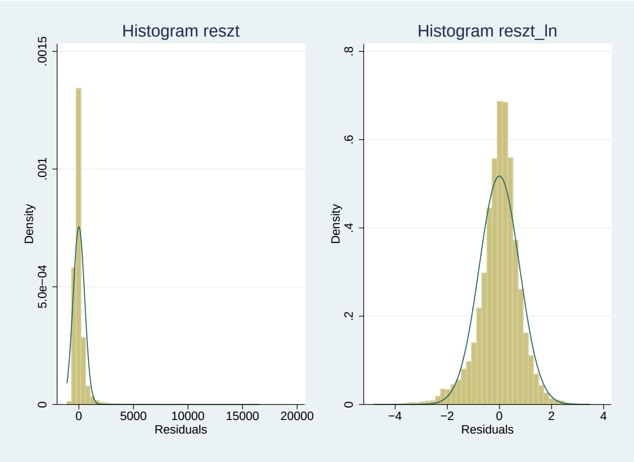
Wyniki regresji: logarytmy

```
reg ln_wydatki_na_mieszkanie ln_dochod  if typ==1

      Source |       SS       df       MS              Number of obs =   18328
-------------+------------------------------           F(  1, 18326) = 1955.92
       Model |   1161.9238     1   1161.9238           Prob > F      =  0.0000
    Residual |  10886.6431 18326  .59405452            R-squared     =  0.0964
-------------+------------------------------           Adj R-squared =  0.0964
       Total |  12048.5669 18327  .657421669           Root MSE      =  .77075

------------------------------------------------------------------------------
ln_wydatkiów |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
   ln_dochod |   .4716229    .010664    44.23   0.000     .4507205    .4925253
       _cons |   2.178381    .085701    25.42   0.000     2.010399    2.346363
------------------------------------------------------------------------------
```

Reszty z regresji



1. Kiedy mówimy, że model można sprowadzić do modelu liniowego względem przekształconych zmiennych?
2. Wyjaśnić, co to jest efekt cząstkowy?
3. Podaj definicję elastyczności cząstkowej.
4. Podaj definicję semielastyczności cząstkowej.