Testy diagnostyczne

Stanisław Cichocki

Natalia Nehrebecka

Wykład 12

Plan wykładu

- 1. Testy diagnostyczne
  - Testowanie stabilności parametrów modelu: test Chowa
  - Testowanie heteroskedastyczności

Testowanie stabilności parametrów

- **Test Chowa:**

Służy do weryfikacji czy parametry modelu będą takie same dla kilku różnych podpróbek

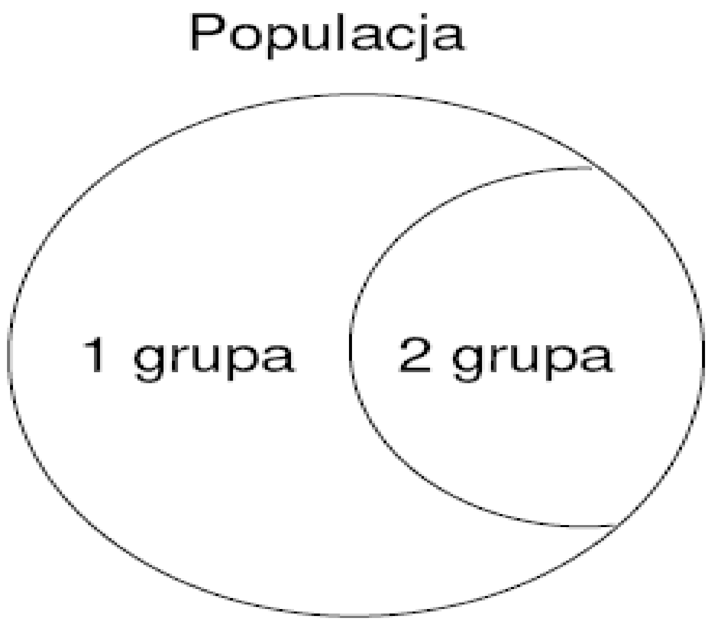
$$H_0 : \beta_1 = \beta_2 = \ldots = \beta_m$$

- parametry są takie same w podpróbkach

$$H_1 : \beta_r \neq \beta_s$$

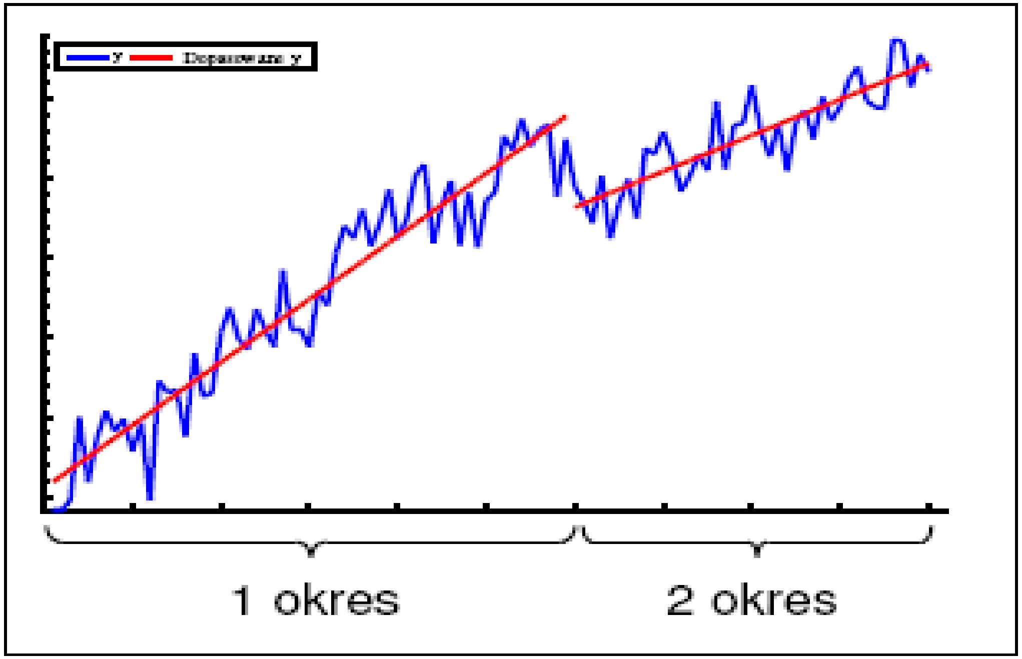
- parametry różnią się w podpróbkach

Testowanie stabilności parametrów - dla próby przekrojowej



Rysunek 4: Podgrupy o podobnej wielkości

Testowanie stabilności parametrów – dla szeregu czasowego



Rysunek 3: Załamanie w parametrach trendu liniowego

Jakie założenie KMRL nie jest spełnione przy odrzuceniu H0?

- Związek pomiędzy zmienną zależną a zmiennymi niezależnymi opisany jest równaniem:

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \cdots + \beta_K x_{Ki} + \varepsilon_i \qquad i = 1,2,3 \cdots n$$

Jakie są skutki niespełnienia założenia KMRL

Odrzucenie hipotezy zerowej o tym, że parametry są stabilne

- **podważa interpretacje ekonomiczną modelu (interpretacja oszacowanych parametrów)**
- **niemożliwe udowodnienie własności estymatora MNK (nieobciążoność czy efektywność estymatora MNK)**

W jaki sposób można rozwiązać problemy zasygnalizowane przez wynik testu?

- Problem niestabilności parametrów można rozwiązać poprzez:
  - wprowadzenie do modelu interakcji pomiędzy zmiennymi 0-1 związanymi z podziałem na grupy a odpowiednimi zmiennymi objaśniającymi (w przypadku gdy jedynie część parametrów jest różna dla analizowanych podprób)
  - estymacje osobnych regresji na wyodrębnionych podpróbach

Testowanie heteroskedastyczności

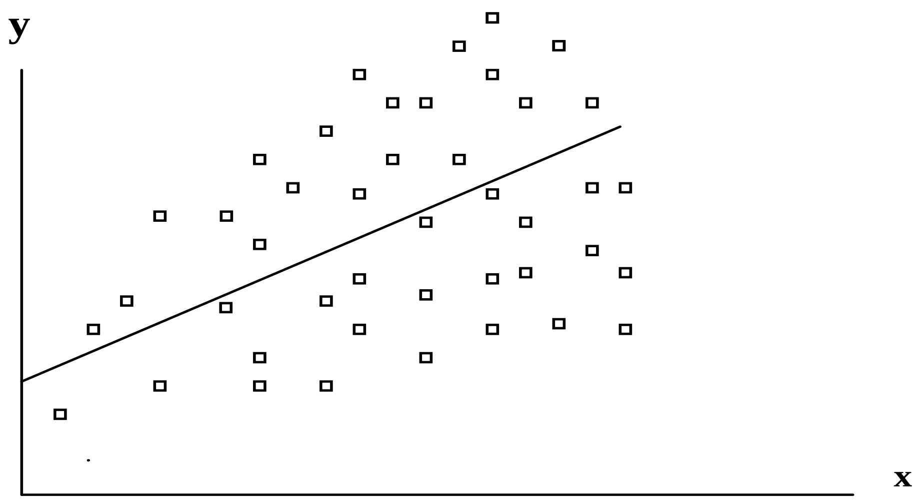
Przypomnienie: Co to znaczy, że w modelu występuje homoskedastyczność/heteroskedastyczność?

- heteroskedastyczność

$$Var(\varepsilon)=\begin{bmatrix} Var(\varepsilon_1) & Cov(\varepsilon_1,\varepsilon_2) & \cdots & Cov(\varepsilon_1,\varepsilon_n) \\ Cov(\varepsilon_2,\varepsilon_1) & Var(\varepsilon_2) & \cdots & Cov(\varepsilon_2,\varepsilon_n) \\ \vdots & \vdots & & \vdots \\ Cov(\varepsilon_n,\varepsilon_1) & Cov(\varepsilon_n,\varepsilon_1) & \cdots & Var(\varepsilon_n) \end{bmatrix}=\begin{bmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{bmatrix}$$

Heteroscedastyczność

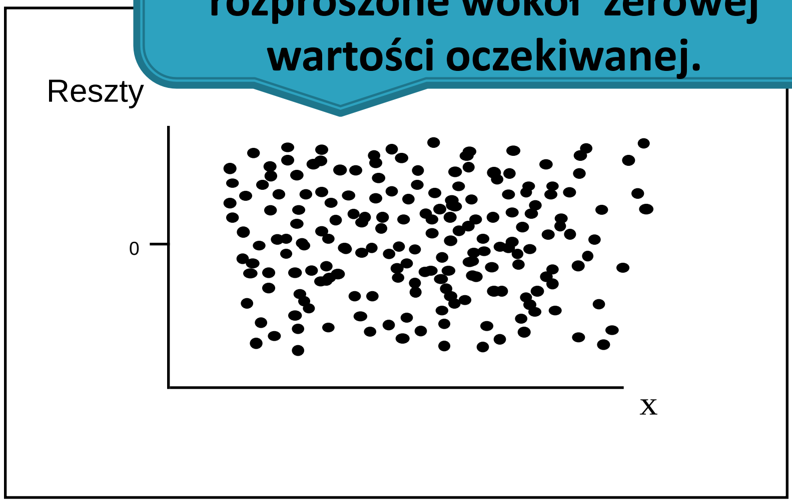
- Stałość wariancji zaburzeń nazywamy homoskedastycznością zaburzeń. **Oznacza to, że zaburzenia losowe są jednakowo rozproszone wokół zerowej wartości oczekiwanej.** Jeśli wariancje nie byłyby jednakowe, to sytuację taką nazywamy heteroskedastycznością.

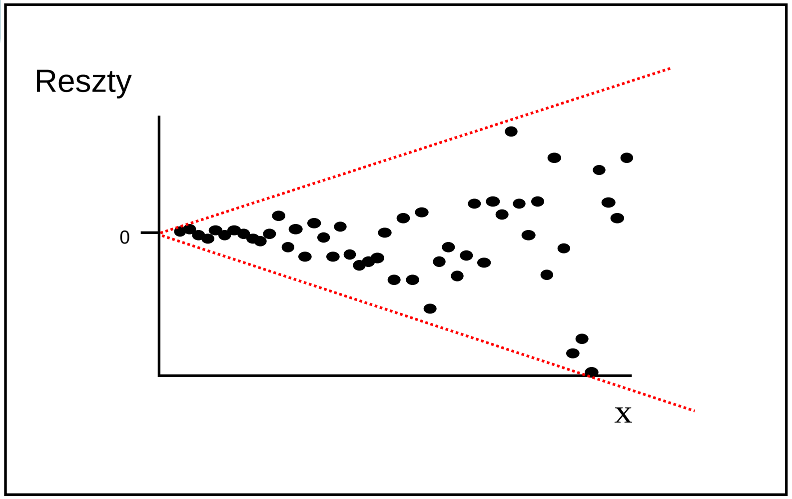


Rys.1. H**eteroskedastyczność**

Heteroskedastyczność

Oznacza to, że zaburzenia losowe są jednakowo rozproszone wokół zerowej wartości oczekiwanej.



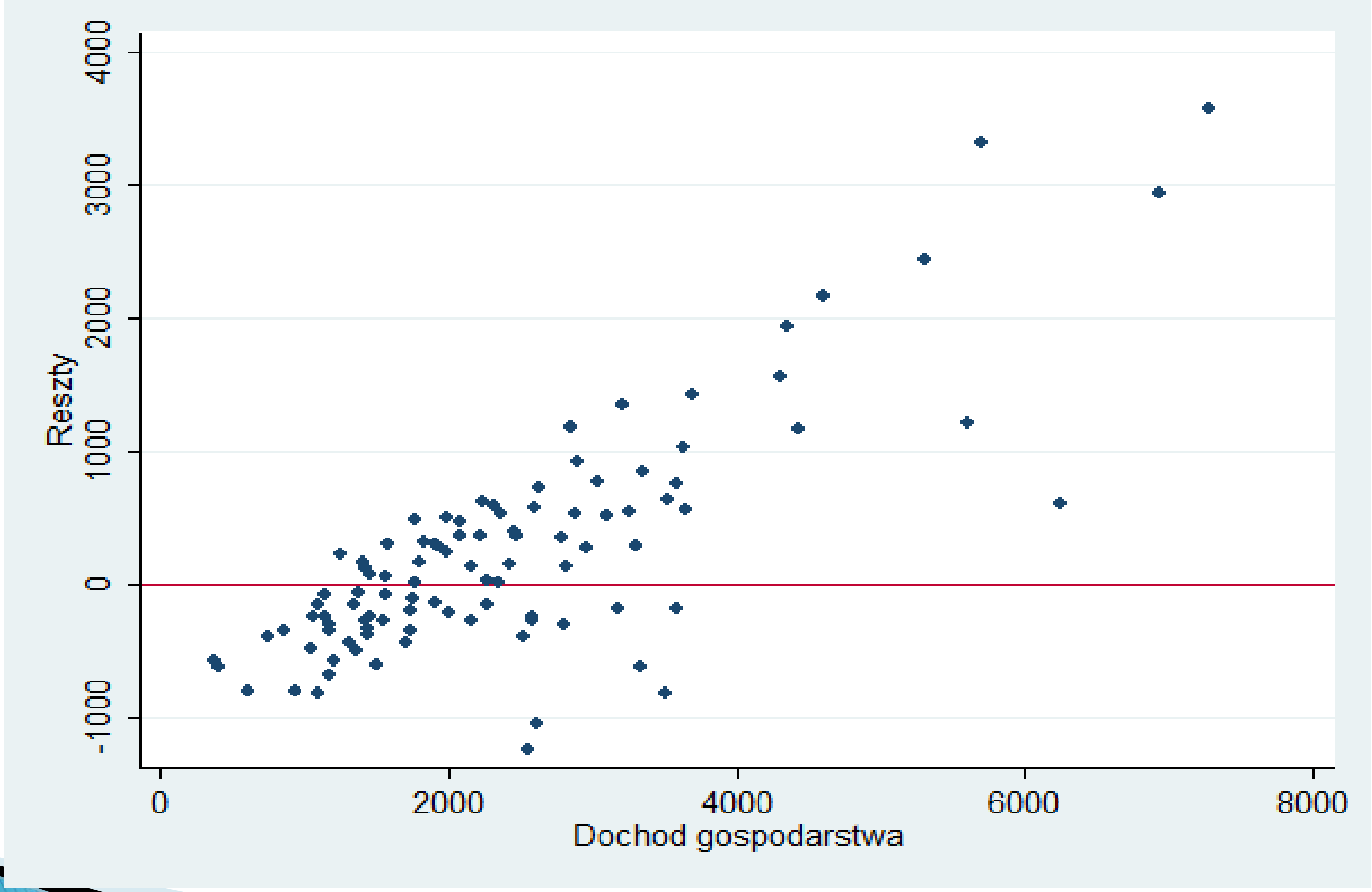


Homoskedastyczność: reszty zachowują się losowo.

Heteroskedastyczność: Wariancja reszt zmienia się wraz ze zmianą zmiennej niezależnej X.

Założenia klasycznego modelu regresji liniowej

- Regresja wydatków na dochodzie



Testowanie heteroskedastyczności

- Test Goldfelda-Quandta (Test GQ):

$$H_0 : Var(\varepsilon_i) = \sigma^2$$ dla $i = 1, \ldots, N$

$$H_1 : Var(\varepsilon_i) > Var(\varepsilon_j)$$ dla $z_i > z_j$ gdzie z jest pewną zmienną, od której zależy wariancja błędu losowego

- Hipoteza zerowa: homoskedastyczność

- Hipoteza alternatywna: heteroskedastyczność

Testowanie heteroskedastyczności

- Test Goldfelda-Quandta (Test GQ):
  - z jego konstrukcji wynika, iż można go stosować do wykrywania zależności między wariancją błędu losowego a wielkością jednej zmiennej
  - jako jedyny z testów na heteroskedastyczność ma rozkład wyprowadzony dla małych prób

Testowanie heteroskedastyczności

- Test Breuscha-Pagana (Test BP):

$$H_0: Var(\varepsilon_i) = \sigma^2$$

dla $i = 1, \dots, N$

$$H_1: Var(\varepsilon_i) = \sigma_i^2 = \sigma^2 f(\alpha_0 + z_i \alpha)$$

gdzie $f(\bullet)$ **- funkcja różniczkowalna**

z_i - wektor zmiennych, może zawierać zmienne występujące w wektorze zmiennych objaśniających

Testowanie heteroskedastyczności

- Test Breuscha-Pagana (Test BP):

  - Hipoteza zerowa: homoskedastyczność

  - Hipoteza alternatywna: heteroskedastyczność

  - Szczególnie przydatny, jeżeli wariancja błędu losowego zależy od kilku zmiennych

Testowanie heteroskedastyczności

- Test Breuscha-Pagana (Test BP) – sposób przeprowadzenia testu:

1. przeprowadzamy regresję y_i na x_i i uzyskujemy e_i

2. przeprowadzamy regresję pomocniczą:

$$\frac{e_i^2}{\tilde{\sigma}^2} = \hat{\alpha}_0 + z_i \hat{\alpha} + \hat{u}_i$$

i testujemy H0: $\alpha = 0$

Testowanie heteroskedastyczności

- Statystyka testowa:

$$LM = \frac{1}{2} ESS \xrightarrow{D} \chi_p^2$$

gdzie: ESS – wyjaśniona suma kwadratów w regresji pomocniczej

p- ilość zmiennych zawartych w z

Testowanie heteroskedastyczności

- Inna statystyka testowa:

$$LM = NR^2 \xrightarrow{D} \chi_p^2$$

gdzie: R^2 - współczynnik determinacji z regresji pomocniczej

Testowanie heteroskedastyczności

- Szczególną postacią testu BP jest test White'a $\longrightarrow z_i$ zawiera wszystkie kwadraty i iloczyny krzyżowe zmiennych objaśniających

- Stosujemy gdy interesuje nas samo wykrycie heteroskedastyczności a mniej wykrycie zmiennych, od których zależy wariancja błędu losowego

Testowanie heteroskedastyczności

- Test BP i White'a są bardziej uniwersalne niż test GQ jednak rozkłady statystyk testowych dla tych testów są znane tylko dla dużych prób

Jakie założenie KMRL nie jest spełnione przy odrzuceniu H0?

- Homoskedastyczność składnika losowego – wariancja błędu losowego jest stała dla wszystkich obserwacji:

$$\operatorname{var}(\varepsilon_i) = \sigma^2 \quad dla \quad i = 1, 2, \ldots, N$$

Testowanie heteroskedastyczności

- Przykład

xi: reg wydg dochg i.klm

```
      Source |       SS       df       MS              Number of obs =   31705
-------------+------------------------------           F(  6, 31698) = 3651.59
       Model |  2.3693e+10     6  3.9489e+09           Prob > F      =  0.0000
    Residual |  3.4278e+10 31698  1081405.34           R-squared     =  0.4087
-------------+------------------------------           Adj R-squared =  0.4086
       Total |  5.7971e+10 31704  1828523.21           Root MSE      =  1039.9

------------------------------------------------------------------------------
        wydg |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
       dochg |   .5818533   .0040164   144.87   0.000      .573981    .5897256
      _Iklm_2 |  -40.65607   23.26644    -1.75   0.081     -86.2592    4.947067
      _Iklm_3 |  -70.57179   25.89099    -2.73   0.006    -121.3191   -19.82444
      _Iklm_4 |  -109.2499   20.60656    -5.30   0.000    -149.6395   -68.86021
      _Iklm_5 |  -153.3497   22.98153    -6.67   0.000    -198.3944    -108.305
      _Iklm_6 |  -173.5506   18.96167    -9.15   0.000    -210.7162    -136.385
       _cons |   836.1774   18.74554    44.61   0.000     799.4354    872.9194
```

Testowanie heteroskedastyczności

- Przykład

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
        Ho: Constant variance
        Variables: fitted values of wydg
        chi2(1)      =129088.50
        Prob > chi2  =   0.0000


White's test for Ho: homoskedasticity
         against Ha: unrestricted heteroskedasticity

         chi2(12)     =   6142.84
         Prob > chi2  =    0.0000
```

Dziękuję za uwagę