# Testy diagnostyczne

**Stanisław Cichocki**

**Natalia Nehrebecka**

Wykład 11

# Plan wykładu

- 1. Testy diagnostyczne
  - Testowanie normalności składników losowych: test Jarque-Berra
  - Testowanie stabilności parametrów modelu: test Chowa
- 2. Heteroskedastyczność
  - Testowanie heteroskedastyczności

# Plan wykładu

- 1. Testy diagnostyczne
  - Testowanie normalności składników losowych: test Jarque-Berra
  - Testowanie stabilności parametrów modelu: test Chowa
- 2. Heteroskedastyczność
  - Testowanie heteroskedastyczności

# Testy diagnostyczne

- Dla każdego testu:

1. Nazwa testu
2. Hipotezy
3. Jakie założenie KMRL nie jest spełnione w przypadku odrzucenia H0?
4. Jakie są konsekwencje niespełnienia założenia KMRL?
5. W jaki sposób można rozwiązać problemy zasygnalizowane przez wynik testu?

# Testowanie normalności składników losowych

- Test Jarque – Bera (Test JB):

$$H_0 : \varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$ - składnik los. ma rozkład normalny

$$H_1 : \varepsilon \nsim N(\mathbf{0}, \sigma^2 \mathbf{I})$$ - składnik los. nie ma rozkładu normalnego

# Testowanie normalności składników losowych

- Sposób przeprowadzenia testu:

1. estymujemy model: $y_i = x_i\beta + \varepsilon_i$ i uzyskujemy reszty

2. liczymy współczynnik skośności i kurtozę dla rozkładu reszt:

$$\hat{\Theta}_1 = \frac{\sum_{i=1}^{N} e_i^{\,3} / N}{\hat{\sigma}^3}$$

$$\hat{\Theta}_2 = \frac{\sum_{i=1}^{N} e_i^{\,4} / N}{\hat{\sigma}^4}$$

gdzie $\hat{\sigma} = \sqrt{\frac{\sum_{i=1}^{N} e_i^2}{N}}$

# Testowanie normalności składników losowych

- Sposób przeprowadzenia testu:

3. Porównujemy wielkość skośności i kurtozy uzyskanych z rozkładu reszt z oczekiwanymi wielkościami tych statystyk dla rozkładu normalnego :

$$\Theta_1 = 0$$

$$\Theta_2 = 3$$

statystka testowa: $$LM = N\left[\frac{\hat{\Theta}_1}{6} + \frac{(\hat{\Theta}_2 - 3)^2}{24}\right] \xrightarrow{D} \chi^2$$

# Testowanie normalności składników losowych

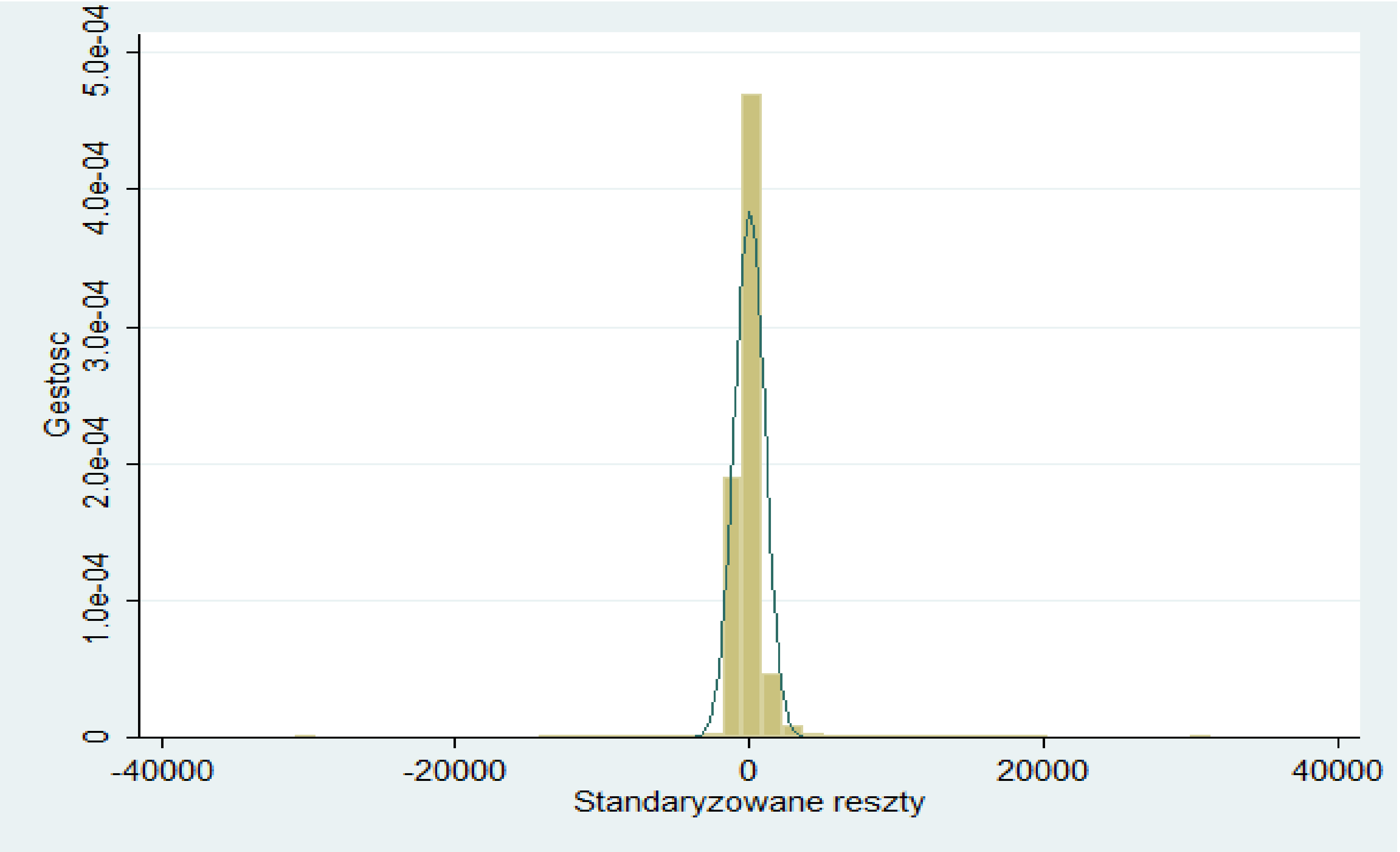
- Przykład

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
```

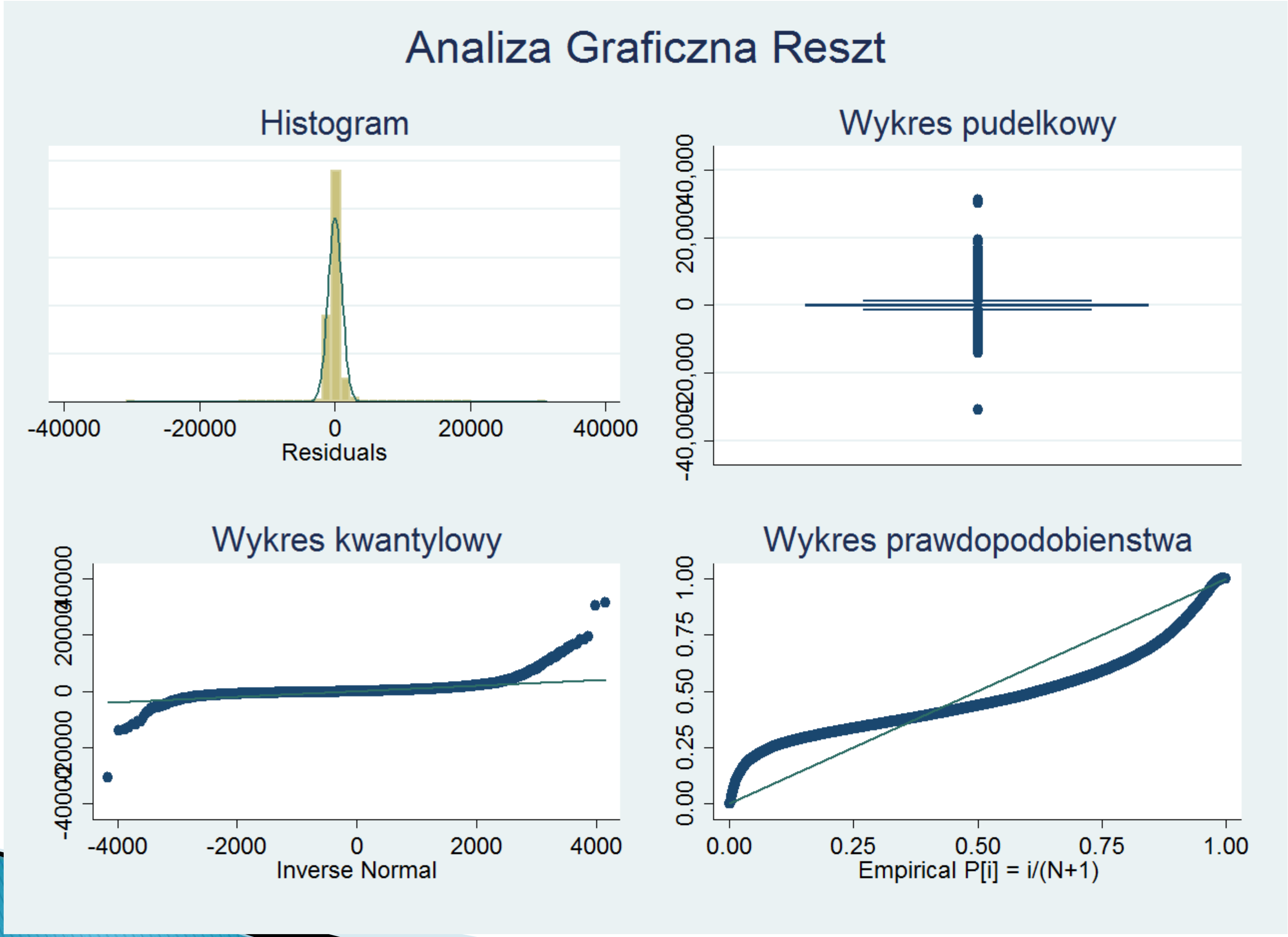
# Testowanie normalności składników losowych



# Testowanie normalności składników losowych

```
 Skewness/Kurtosis tests for Normality
                                                        ------- joint ------
    Variable |     Obs   Pr(Skewness)   Pr(Kurtosis)      chi2(2)     Prob>chi2
-------------+-------------------------------------------------------------------
           e |   3.2e+04   0.0000          0.0000     34060.88         0.0000
```

# Testowanie normalności składników losowych



# Jakie założenie KMRL nie jest spełnione przy odrzuceniu H0?

- Niespełnione dodatkowe założenie o tym, że składnik losowy ma rozkład normalny

# Jakie są skutki niespełnienia założenia KMRL

- **Próba duża:** rozkłady statystyk są bliskie standardowym rozkładom

- **Mała próba:** jest problemem, gdyż:
  - To założenie jest niezbędne do wyprowadzenie rozkładów statystyk testowych oraz prawidłowego wnioskowania statystycznego.
  - Estymator b uzyskany metoda MNK jest najlepszym estymatorem wśród *liniowych i nieobciążonych* estymatorów → można znaleźć estymator *nieliniowy i nieobciążony o* wariancji mniejszej niż estymator b

# Testowanie stabilności parametrów

**Test Chowa:**

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

Załóżmy, że modele dla podpróbek: $y_s = X_s \beta_s + \varepsilon_s$

gdzie *s=1,...,m* oznacza numer podpróbki.

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

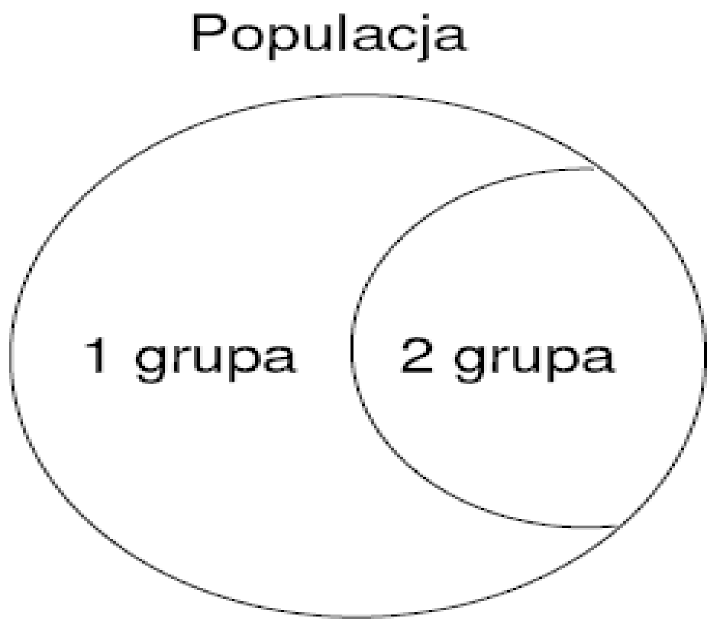
- parametry są takie same w podpróbkach

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

dla pewnych r i s

- parametry różnią się w podpróbkach

# Testowanie stabilności parametrów - dla próby przekrojowej



# Testowanie stabilności parametrów

Sposób przeprowadzania testu:

1. przeprowadzamy regresję na całej próbie (bez zmiennej wyodrębniającej podpróbki);
2. przeprowadzamy regresję na każdej z propróbek;
3. obliczamy statystkę:

$$F = \frac{(S - \sum_{s=1}^{m} S_j)/[K(m-1)]}{\sum_{s=1}^{m} S_j/(N-mK)} \sim F(K(m-1), N-mK)$$

# Testowanie stabilności parametrów

Gdzie:

$S$ – suma kwadratów reszt z regresji na całej próbie;

$S_j$ – suma kwadratów reszt z regresji na j-tej podpróbie;

$m$ - liczba podprób;

$K$ - liczba szacowanych parametrów;

$N$ - liczba obserwacji

# 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

# Plan wykładu

- 1. Testy diagnostyczne
  - Testowanie normalności składników losowych: test Jarque-Berra
  - Testowanie stabilności parametrów modelu: test Chowa
- 2. Heteroskedastyczność
  - Testowanie heteroskedastyczności

# Heteroskedastyczności

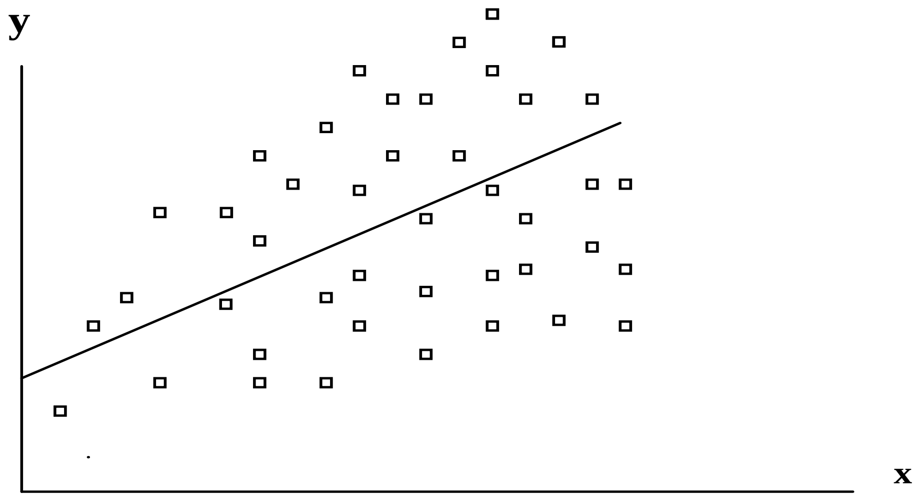
Przypomnienie: Co to znaczy, że w modelu występuje homoskedastyczność/heteroskedastyczność?

**- Heteroskedastyczność**

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

# Heteroskedastyczność

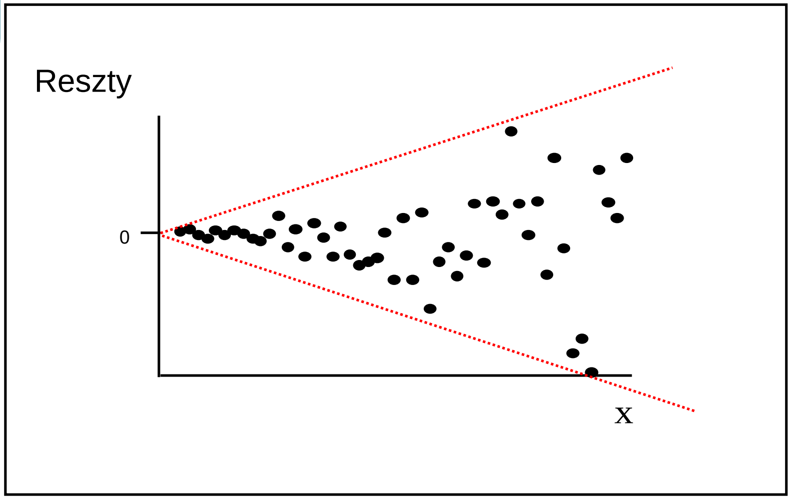
- Stałość wariancji zaburzeń nazywamy homoskedastycznością zaburzeń. **Oznacza to, że zaburzenia losowe są jednakowo rozproszone wokół zerowej wartości oczekiwanej.** Jeśli wariancje nie byłyby jednakowe, to sytuację taką nazywamy heteroskedastycznością.




Rys.1. Heteroskedastyczność

# Heteroskedastyczność

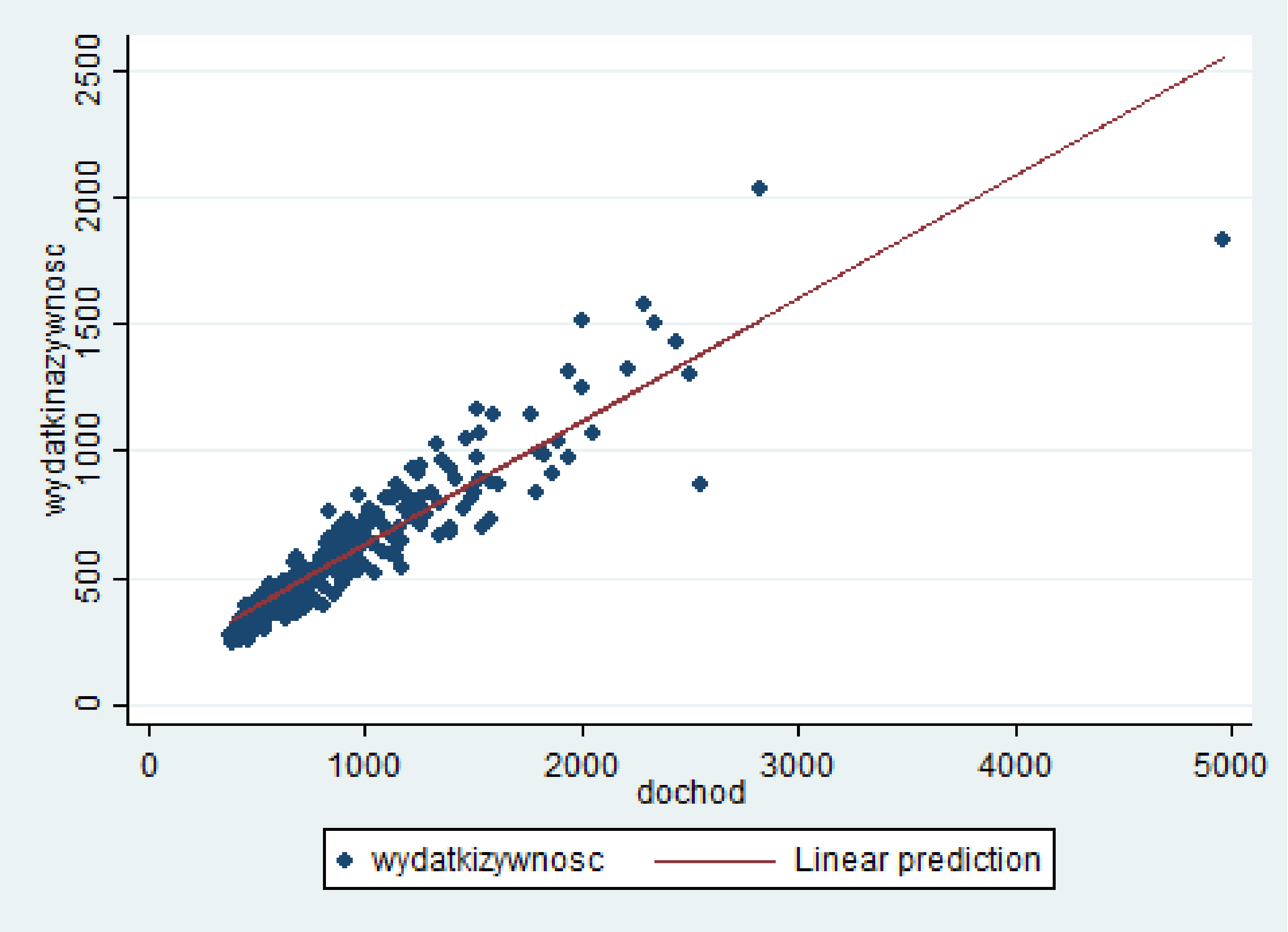
**Oznacza to, że zaburzenia losowe są jednakowo rozproszone wokół zerowej wartości oczekiwanej.**

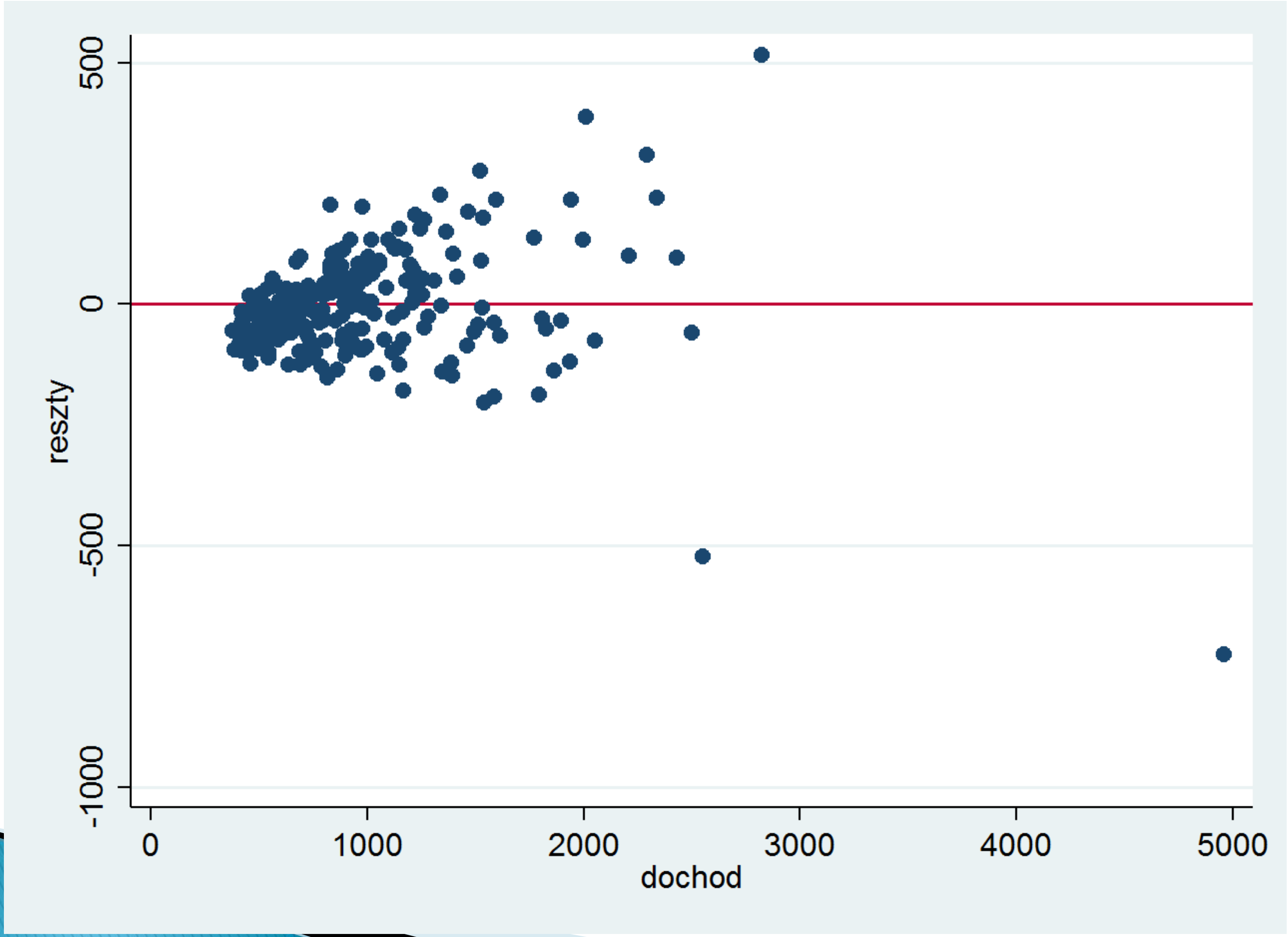

**Homoskedastyczność**: reszty zachowują się losowo.

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

# Heteroskedastyczność



# Heteroskedastyczność



# Testowanie heteroskedastyczności

- Test Goldfelda-Quandta – **samodzielnie**:
  - 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 $i = 1, \ldots, N$

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

**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:

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

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:

$$LM = NR^2 \xrightarrow{D} \chi^2_p$$

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 Breuscha-Pagana i test White'a są bardziej uniwersalne niż test Goldfelda-Quandta

- Jednak statystyki testowe dla testu Breuscha-Pagana i testu White'a 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:

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

# 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
```

# Pytania teoretyczne

1. Za pomocą jakiego testu weryfikowana jest normalność składnika losowego? Jakiemu założeniu KMRL odpowiada H0 w tym teście? Jaka jest hipoteza alternatywna w tym teście? Jakie są konsekwencje dla własności MNK, jeśli H0 jest fałszywe?
2. Za pomocą jakiego testu testowana jest stabilność parametrów? Jakiemu założeniu KMRL odpowiada H0 w tym teście? Jaka jest hipoteza alternatywna w tym teście?

# Pytania teoretyczne

3. Za pomocą jakich testów można testować heteroskedastyczność? Jakiemu założeniu KMRL odpowiada H0 w tych testach? Jakie są hipotezy alternatywne w tych testach?

# Dziękuję za uwagę