

# Testowanie hipotez

## Testy diagnostyczne

**Stanisław Cichocki**

**Natalia Nehrebecka**

Wykład 11

# Plan wykładu

- ▶ 1. Testowanie hipotez łącznych
- ▶ 2. Testy diagnostyczne
  - Testowanie prawidłowości formy funkcyjnej: test RESET
  - Testowanie normalności składników losowych: test Jarque-Berra

# Plan wykładu

- ▶ 1. Testowanie hipotez łącznych
- ▶ 2. Testy diagnostyczne
  - Testowanie prawidłowości formy funkcyjnej: test RESET
  - Testowanie normalności składników losowych: test Jarque-Berra

# Testowanie hipotez złożonych

- ▶ Hipoteza łączna:

$$H_0: H\beta = h$$

- ▶ Jest to układ  $g$  równań liniowych
- ▶ Macierz  $H$  ma pełen rząd wierszowy równy  $g$  (liczba ograniczeń)
- ▶ Poszczególne równania powinny być liniowo niezależne
- ▶ Układ równań nie powinien być sprzeczny

# Testowanie hipotez złożonych

- ▶ Przykład – model:

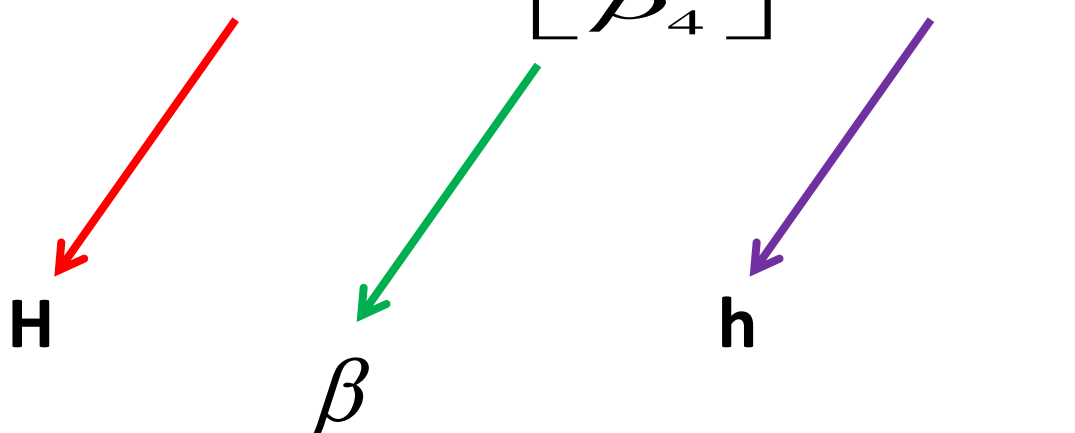
$$y_i = \beta_1 + \beta_2 X_{2i} + \beta_3 X_{3i} + \beta_4 X_{4i} + \varepsilon_i$$

Testujemy hipotezę:

$$H_0: \begin{cases} \beta_1 = 1 \\ \beta_2 = 2 \\ \beta_3 = \beta_4 \end{cases}$$

# Testowanie hipotez złożonych

$$\begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & -1 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_2 \\ \beta_3 \\ \beta_4 \end{bmatrix} = \begin{bmatrix} 1 \\ 2 \\ 0 \end{bmatrix}$$



**H**                       $\beta$                       **h**

# Testowanie hipotez złożonych

- ▶ Testowanie hipotez prostych nie jest równoważne testowaniu hipotezy łącznej, że wszystkie rozważane hipotezy proste są łącznie prawdziwe

# Testowanie hipotez złożonych

$$y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_K x_{Ki} + \varepsilon_i$$

- ▶ **Krok 1.** Stawiamy przykładową hipotezę zerową  $H_0$ :

$$H_0 : \begin{cases} \beta_2 = 0 \\ \beta_3 = 0 \end{cases}$$

- ▶ Brak podstaw do odrzucenia tej hipotezy oznacza, że zmienne  $x_{2i}, x_{3i}$  są łącznie nieistotne
- ▶ **Model bez ograniczeń**  $y_i = \beta_1 + \beta_2 x_{2i} + \beta_3 x_{3i} + \dots + \beta_K x_{Ki} + \varepsilon_i$
- ▶ **Model z ograniczeniami**  $y_i = \beta_1 + \beta_4 x_{4i} + \dots + \beta_K x_{Ki} + \varepsilon_i$



# Testowanie hipotez złożonych

- ▶ **Krok 2.** Przy założeniu, że postawiona hipoteza zerowa jest prawdziwa, wyznaczamy statystykę testową z rozkładu F:

$$F = \frac{(\mathbf{e}'_R \mathbf{e}_R - \mathbf{e}'\mathbf{e}) / g}{\mathbf{e}'\mathbf{e} / (N - K)} = \frac{(R^2 - R_R^2) / g}{(1 - R^2) / (N - K)}$$

- ▶ Gdzie:
- ▶  $R^2, \mathbf{e}'\mathbf{e}$  oznaczają współczynnik determinacji i sumę kwadratów reszt dla modelu bez ograniczeń
- ▶  $R^2_R, \mathbf{e}'_R \mathbf{e}_R$  to te same wielkości, ale dla modelu z ograniczeniami,
- ▶  $g$  oznacza liczbę ograniczeń,
- ▶  $K$  – ilość szacowanych parametrów w modelu bez ograniczeń,
- ▶  $N$  – liczba obserwacji

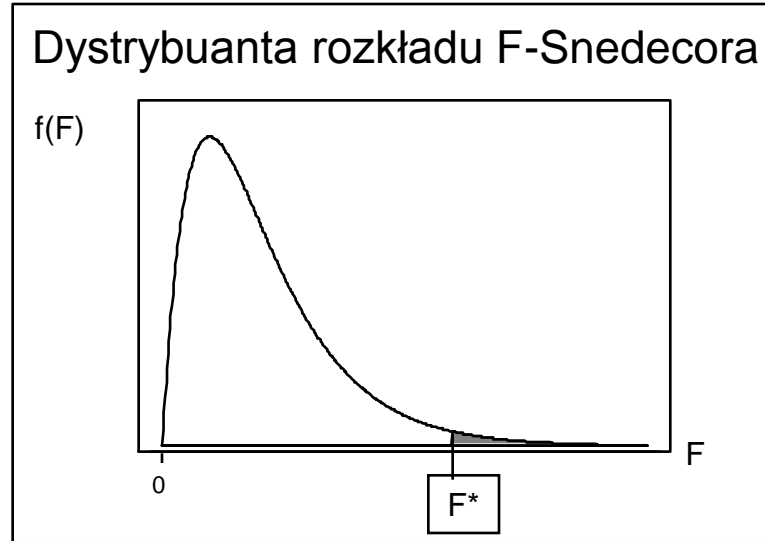
# Testowanie hipotez złożonych

- ▶ **Krok 3.** Odczytujemy z tablic rozkładu F wartość krytyczną ( $\alpha$  - poziom istotności)

$$F^* = F(g, n - K)$$

# Testowanie hipotez złożonych

- ▶ **Krok 4.** Podjęcie decyzji



- ▶  $F \geq F^*$  - odrzucamy hipotezę
- ▶  $F < F^*$  - nie ma podstaw do odrzucenia hipotezy zerowej

# Testowanie hipotez złożonych

## ► Przykład

xi: reg wydg dochg i.klm (model bez ograniczeń)

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

wydg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg	.5818533	.0040164	144.87	0.000	.573981	.5897256
_Ik1m_2	-40.65607	23.26644	-1.75	0.081	-86.2592	4.947067
_Ik1m_3	-70.57179	25.89099	-2.73	0.006	-121.3191	-19.82444
_Ik1m_4	-109.2499	20.60656	-5.30	0.000	-149.6395	-68.86021
_Ik1m_5	-153.3497	22.98153	-6.67	0.000	-198.3944	-108.305
_Ik1m_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 hipotez złożonych

## ▶ Przykład

xi: reg wydg dochg (model z ograniczeniami)

Source	SS	df	MS			
Model	2.3571e+10	1	2.3571e+10	Number of obs	=	31705
Residual	3.4401e+10	31703	1085097.53	F( 1, 31703)	=	21722.15
Total	5.7971e+10	31704	1828523.21	Prob > F	=	0.0000
				R-squared	=	0.4066
				Adj R-squared	=	0.4066
				Root MSE	=	1041.7

wydg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg	.5874177	.0039856	147.38	0.000	.5796057	.5952296
_cons	714.5099	10.00872	71.39	0.000	694.8924	734.1274

# Testowanie hipotez złożonych

$$F = \frac{(\mathbf{e}'_R \mathbf{e}_R - \mathbf{e}'\mathbf{e}) / g}{\mathbf{e}'\mathbf{e} / (N - K)} = \frac{(3,4401e+10 - 3,4278e+10) / 5}{(3,4278e+10) / (31705 - 7)} \approx 22,65$$

# Testowanie hipotez złożonych

```
test _Iklm_2 _Iklm_3 _Iklm_4 _Iklm_5 _Iklm_6
```

```
( 1)  _Iklm_2 = 0
```

```
( 2)  _Iklm_3 = 0
```

```
( 3)  _Iklm_4 = 0
```

```
( 4)  _Iklm_5 = 0
```

```
( 5)  _Iklm_6 = 0
```

```
F( 5, 31698) = 22.65
```

```
Prob > F = 0.0000
```

# Plan wykładu

- ▶ 1. Testowanie hipotez łącznych
- ▶ 2. Testy diagnostyczne
  - Testowanie prawidłowości formy funkcyjnej: test RESET
  - Test normalności składników losowych: test Jarque-Berra



# Testy diagnostyczne

- ▶ Służą do weryfikacji założeń KMRL
- ▶ Sprawdzenie założeń KMRL jest ważne  $\longrightarrow$  na nich opierają się własności estymatorów MNK
- ▶ Jeśli któreś z założeń nie jest spełnione  $\longrightarrow$  należy zastanowić się nad przeformułowaniem modelu lub zastosować bardziej zaawansowane narzędzia ekonometryczne
- ▶ Testy są stosowane po wyestymowaniu modelu

# Testy diagnostyczne

- ▶ W praktyce do testowania jednego założenia KMRL używa się często kilku testów
- ▶ Czasami różne testy zastosowane do testowania tej samej hipotezy zerowej dają sprzeczne wnioski

# Testowanie prawidłowości formy funkcyjnej

- Test RESET (*Regression Specification Error Test*):

$$H_0 : y_i = x_i\beta + \varepsilon_i \quad - \text{liniowa postać modelu}$$

$$H_1 : y_i = f(x_i\beta) + \varepsilon_i \quad - \text{nieliniowa postać modelu}$$

gdzie  $f(\bullet)$  jest nieliniowa

# Testowanie prawidłowości formy funkcyjnej

► Sposób przeprowadzenia testu:

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

$$\hat{y} = x_i\hat{b}$$

2. przeprowadzamy regresję pomocniczą:

$$y_i = x_i\beta + \alpha_1 y_i + \dots + \alpha_p y_i + u_i$$

i za pomocą testu F testujemy  $H_0$ :

$$\alpha_1 = \alpha_2 = \dots = \alpha_p = 0$$

# Testowanie prawidłowości formy funkcyjnej

- ▶ W dużych próbach rozkład statystyki będzie dążył do rozkładu F-Snedecora o  $N-p$  i  $p$  stopniach swobody

# Testowanie prawidłowości formy funkcyjnej

## ► Przykład

xi: reg wydg dochg i.klm

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

wydg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg	.5818533	.0040164	144.87	0.000	.573981	.5897256
_Ik1m_2	-40.65607	23.26644	-1.75	0.081	-86.2592	4.947067
_Ik1m_3	-70.57179	25.89099	-2.73	0.006	-121.3191	-19.82444
_Ik1m_4	-109.2499	20.60656	-5.30	0.000	-149.6395	-68.86021
_Ik1m_5	-153.3497	22.98153	-6.67	0.000	-198.3944	-108.305
_Ik1m_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 prawidłowości formy funkcyjnej

Ramsey RESET test using powers of the fitted values of `wydg`

Ho: model has no omitted variables

F(3, 31695) = 907.11

Prob > F = 0.0000

## Jakie założenie KMRL nie jest spełnione przy odrzuceniu $H_0$ ?

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

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



# Jakie są skutki niespełnienia założenia KMRL

Odrzucenie hipotezy  
zerowej o  
poprawności przyjętej  
formy funkcyjnej

- 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?

Przebudowanie modelu aby uwzględnił nieliniowość relacji między zmiennymi (możliwe, że zmienne modelu powinny być poddane jakiejś transformacji – logarytmowanie, potęgowanie, itp.)

# Testowanie normalności składników losowych

- Test Jarque – Berra (Test JB):

$$H_0 : \varepsilon \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

- składnik los. ma rozkład normalny

$$H_1 : \varepsilon \not\sim N(\mathbf{0}, \sigma^2 \mathbf{I})$$

- 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:

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

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

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

# 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 :

$$\Theta_1 = 0$$

$$\Theta_2 = 3$$

statystyka 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

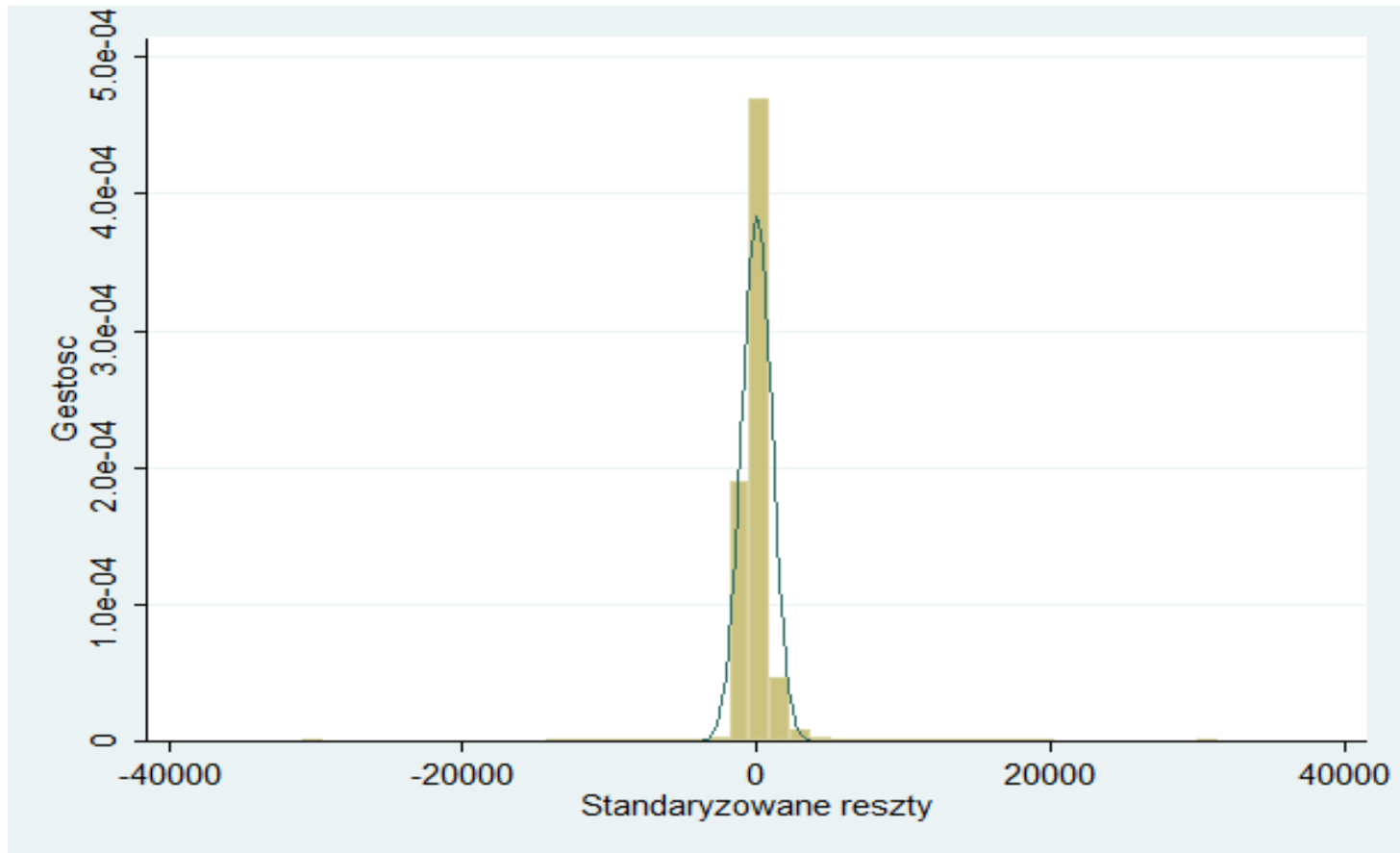
xi: reg wydg dochg i.klm

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

wydg	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
dochg	.5818533	.0040164	144.87	0.000	.573981	.5897256
_Ik1m_2	-40.65607	23.26644	-1.75	0.081	-86.2592	4.947067
_Ik1m_3	-70.57179	25.89099	-2.73	0.006	-121.3191	-19.82444
_Ik1m_4	-109.2499	20.60656	-5.30	0.000	-149.6395	-68.86021
_Ik1m_5	-153.3497	22.98153	-6.67	0.000	-198.3944	-108.305
_Ik1m_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

Variable	Obs	Pr(Skewness)	Pr(Kurtosis)	----- joint ----- chi2(2)	Prob>chi2
e	3.2e+04	0.0000	0.0000	34060.88	0.0000



## Jakie założenie KMRL nie jest spełnione przy odrzuceniu $H_0$ ?

- ▶ 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  $\longrightarrow$  można znaleźć estymator *nieliniowy i nieobciążony* o wariancji mniejszej niż estymator  $b$

# Pytania teoretyczne

1. Jak należy testować hipotezę postaci:  $H_0: H\beta = h$ , używając do tego sumy kwadratów reszt z modelu bez ograniczeń i z ograniczeniami?
2. Wyjaśnić, jakie korzyści i niebezpieczeństwa łączą się z narzucaniem ograniczeń na model.
3. Do czego służą testy diagnostyczne?
4. Za pomocą jakiego testu testujemy prawidłowość formy funkcyjnej? Jakiemu założeniu KMRL odpowiada  $H_0$  w tym teście? Jaka jest hipoteza alternatywna w tym teście?
5. (\*) Opisz dwa sposoby przeprowadzania testu RESET.

# Pytania teoretyczne

6. Za pomocą jakiego testu weryfikowana jest normalność składnika losowego? Jakiemu założeniu KMRL odpowiada  $H_0$  w tym teście? Jaka jest hipoteza alternatywna w tym teście? Jakie są konsekwencje dla własności MNK, jeśli  $H_0$  jest fałszywe?

**Dziękuję za uwagę**