

Testy diagnostyczne (II)

Natalia Nehrebecka
Stanisław Cichocki

Wykład 10

Plan wykładu

1. Testy diagnostyczne
2. Testowanie prawidłowości formy funkcyjnej modelu
3. Testowanie normalności składników losowych
- 4. Testowanie stabilności parametrów**
5. Testowanie heteroskedastyczności

Testowanie stabilności parametrów – Test Chowa

- Służy do weryfikacji czy parametry modelu będą takie same dla kilku różnych próbek
- Załóżmy, że modele dla próbek: $y_s = X_s \beta_s + \varepsilon_s$,
gdzie $s=1, \dots, m$ oznacza numer próbki

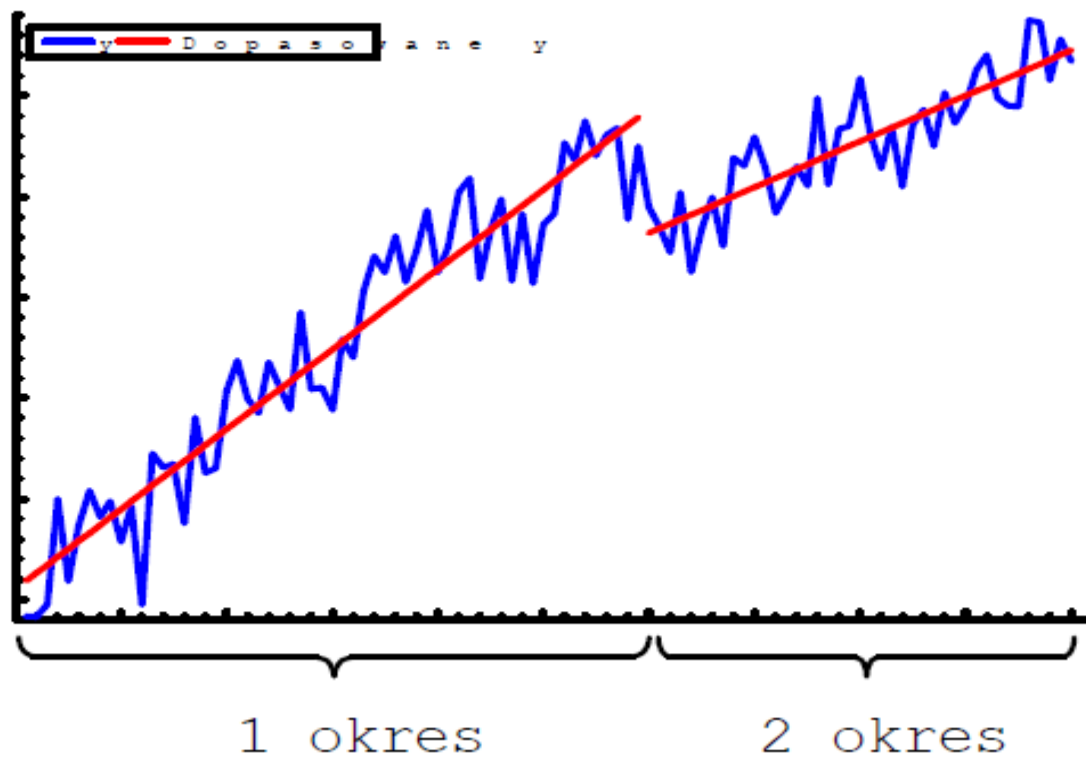
$H_0 : \beta_1 = \beta_2 = \dots = \beta_m$ - parametry są takie same w próbkach

$H_1 : \beta_r \neq \beta_s$ - parametry różnią się w próbkach

dla pewnych r, s

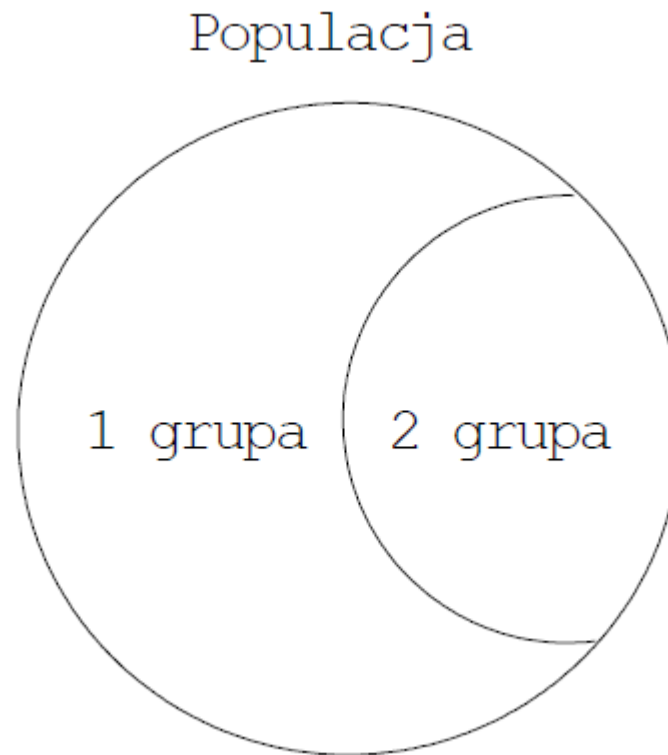
Dla szeregu czasowego

1. W momencie t^* parametry modelu ulegają zmianie (zmiany strukturalne w gospodarce)



Dla próby przekrojowej

2. Podgrupa analizowanej populacji opisywana innym modelem niż reszta populacji



Procedura przeprowadzenia testu Chowa

Sprawdzimy, czy parametry regresji są takie same w próbkach wyodrębnionych za pomocą zmiennej np. płeć

KROK 1: przeprowadzamy regresję na całej próbie (nie wprowadzamy do modelu zmiennej płęć!)

- obliczamy: $S = \text{RSS}$ - Suma kwadratów z regresji na całej próbie

KROK 2: przeprowadzamy regresję na próbce kobiet

- obliczamy: $S_1 = \text{RSS}_1$ - Suma kwadratów z regresji na próbce zawierającej kobiety

KROK 3: przeprowadzamy regresję na próbce mężczyzn

- obliczamy: $S_2 = \text{RSS}_2$ - Suma kwadratów z regresji na próbce zawierającej mężczyzn

Testowanie stabilności parametrów – test Chowa

Statystyka opisowa:

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

Gdzie:

S – suma kwadratów reszt z regresji na całej próbie,

S_s – suma kwadratów reszt z regresji na j -tej podpróbie,

m – liczba wyodrębnionych próbek,

K – liczba szacowanych parametrów (*taka sama we wszystkich regresjach*),

N – liczba obserwacji.

Przykład – test Chow

```
. xi: regress siops age age3 ds dw pasiops masiops
```

Source	SS	df	MS
Model	51114.1379	6	8519.02298
Residual	96395.3946	1225	78.690118
Total	147509.532	1231	119.829027

```
Number of obs = 1232
F( 6, 1225) = 108.26
Prob > F = 0.0000
R-squared = 0.3465
Adj R-squared = 0.3433
Root MSE = 8.8707
```

siops	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.2190567	.05151	4.25	0.000	.1179991	.3201142
age3	-.0000216	6.55e-06	-3.30	0.001	-.0000345	-8.76e-06
ds	4.84467	.6812481	7.11	0.000	3.508127	6.181212
dw	13.31385	.7250675	18.36	0.000	11.89133	14.73636
pasiops	.084663	.0303691	2.79	0.005	.0250818	.1442441
masiops	.1042921	.0317458	3.29	0.001	.0420101	.1665742
_cons	18.39664	2.241034	8.21	0.000	13.99995	22.79333

Przykład – test Chowa

```
. xi: regress siops age age3 ds dw pasiops masiops if sex==2
```

Source	SS	df	MS
Model	31352.4708	6	5225.41181
Residual	55567.3783	656	84.7063694
Total	86919.8492	662	131.298866

Number of obs = 663
 F(6, 656) = 61.69
 Prob > F = 0.0000
 R-squared = 0.3607
 Adj R-squared = 0.3549
 Root MSE = 9.2036

siops	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.1967849	.0736225	2.67	0.008	.0522207	.3413492
age3	-.0000169	9.01e-06	-1.88	0.061	-.0000346	7.79e-07
ds	6.909362	.9927858	6.96	0.000	4.959941	8.858783
dw	12.12589	.9545302	12.70	0.000	10.25158	14.00019
pasiops	.0999609	.0424373	2.36	0.019	.0166316	.1832903
masiops	.0573792	.045161	1.27	0.204	-.0312982	.1460567
_cons	18.25259	3.203996	5.70	0.000	11.96127	24.54391

Przykład – test Chow

```
. xi: regress siops age age3 ds dw pasiops masiops if sex==1
```

Source	SS	df	MS			
Model	21630.7481	6	3605.12468	Number of obs =	569	
Residual	38951.28	562	69.3083274	F(6, 562) =	52.02	
Total	60582.0281	568	106.6585	Prob > F =	0.0000	
				R-squared =	0.3570	
				Adj R-squared =	0.3502	
				Root MSE =	8.3252	

siops	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
age	.2618198	.0716213	3.66	0.000	.1211416	.402498
age3	-.0000284	9.46e-06	-3.00	0.003	-.000047	-9.81e-06
ds	2.201164	.928647	2.37	0.018	.3771207	4.025207
dw	15.73118	1.160005	13.56	0.000	13.45271	18.00966
pasiops	.0491113	.0431628	1.14	0.256	-.0356689	.1338915
masiops	.1614052	.0436714	3.70	0.000	.0756261	.2471843
_cons	18.85553	3.117649	6.05	0.000	12.73186	24.9792

Przykład – test Chowa

$$F = \frac{[S - (S_1 + S_2)] / (K(m - 1))}{(S_1 + S_2) / (N - mK)}$$

$$F = \frac{(96395,3946 - 55567,3783 - 38951,28) / (7 * 1)}{(55567,3783 + 38951,28) / (1232 - 7 * 2)} = 3,4548958$$

$$F^* = F(7, 1218) = 2,0170835$$

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 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

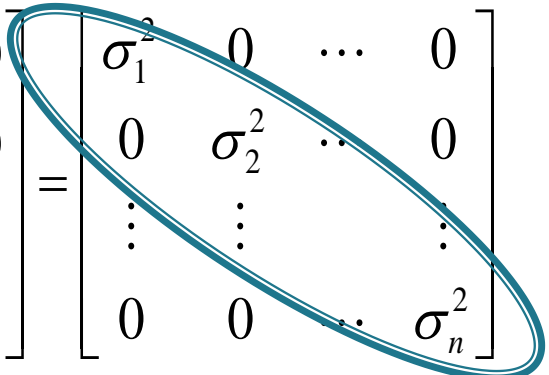
Plan wykładu

1. Testy diagnostyczne
2. Testowanie prawidłowości formy funkcyjnej modelu
3. Testowanie normalności składników losowych
4. Testowanie stabilności parametrów
5. Testowanie heteroskedastyczności

Testowanie heteroskedastyczności

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

- heteroskedastyczność

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


Testowanie heteroskedastyczności – Test Goldfelda-Quandt

- ▶ Test Goldfelda-Quandt – **samodzielnie!**
 - z 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 Breusch-Pagana (BP)

- ▶ Test BP stosowany jeśli istnieje podejrzenie, że wariancja błędów losowych w modelu zależy od pewnego wektora zmiennych z_i

$$H_0 : \text{Var}(\varepsilon_i) = \sigma^2 \quad i = 1, \dots, N$$

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

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

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

Testowanie heteroskedastyczności

- ▶ Szczególną postacią testu Breuscha-Pagana jest test White'a



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 Breusch-Pagana i White'a* są bardziej uniwersalne niż *test Goldfelda-Quandt* jednak rozkłady statystyk testowych dla tych testów są znane tylko **dla dużych prób**

Jakie założenie KMRL nie jest spełnione przy odrzuceniu H_0 ?

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

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

Pytania teoretyczne

1. Za pomocą jakich testów testujemy stabilność parametrów? Jakiemu założeniu KMRL odpowiada H_0 w tym testach? Jakie są hipotezy alternatywne w tym testach?
2. (*) W jaki sposób przeprowadzamy test Chowa?
3. Za pomocą jakich testów można testować heteroskedastyczność? Jakiemu założeniu KMRL odpowiada H_0 w tych testach? Jakie są hipotezy alternatywne w tym testach?
4. (*) Jak przeprowadza się test Goldfelda-Quandta?
5. (*) Podać postać regresji pomocniczej w przypadku testu Breuscha-Pagana.

Dziękuję za uwagę