Testy diagnostyczne (cz. III)

Wykład 13

Natalia Nehrebecka
Stanisław Cichocki

Plan zajęć

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

Plan zajęć

1. Testy diagnostyczne

2. Testowanie prawidłowości formy funkcyjnej modelu

3. Testowanie normalności składników losowych

4. Testowanie stabilności parametrów

5. Testowanie heteroskedastyczności

6. Testowanie autokorelacji

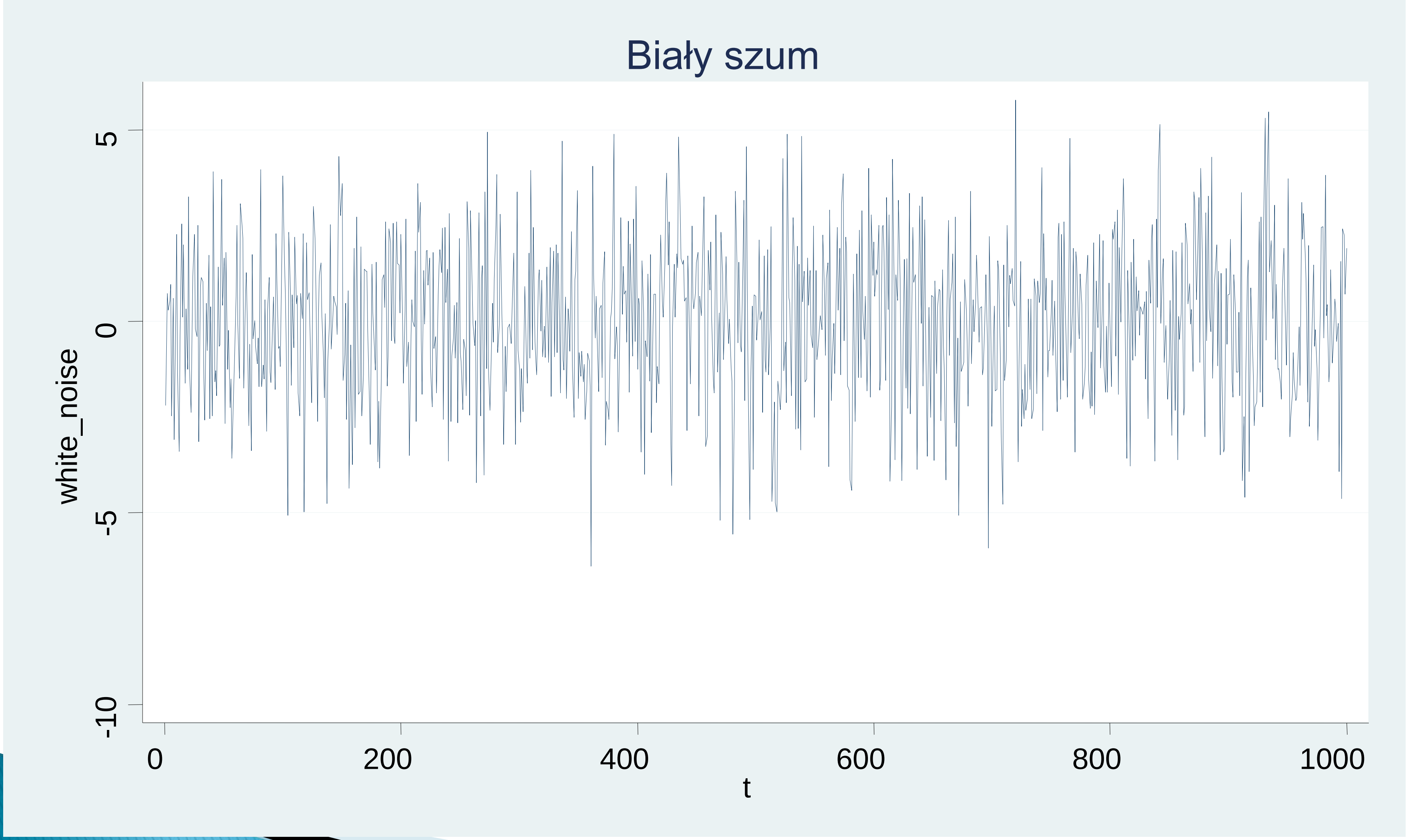
Testowanie autokorelacji

Przypomnienie: Co to znaczy, że w modelu występuje autokorelacja?

- Brak autokorelacji

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

Sferyczne błędy losowe



Testowanie autokorelacji

- Test Durbina-Watsona (Test DW):

- Test Durbina-Watsona jest powszechnie stosowanym testem wykrywania autokorelacji pierwszego rzędu, a więc *autokorelacji miedzy sąsiednimi zaburzeniami losowymi*.

- Autokorelacje pierwszego rzędu opisuje równanie:

$$\varepsilon_t = \rho \varepsilon_{t-1} + u_t$$

- gdzie: ρ jest współczynnikiem autokorelacji zaburzeń

$$Var(u) = \sigma_u^2 I$$

Testowanie autokorelacji

- Test Durbina-Watsona (Test DW):

$$H_0 : Cov(\varepsilon_t, \varepsilon_{t-1}) = 0$$ **- brak autokorelacji**

$$H_1 : Cov(\varepsilon_t, \varepsilon_{t-1}) \neq 0$$ **- autokorelacja**

gdzie: $t = 1, \ldots, T$

Testowanie autokorelacji

Sposób przeprowadzenia testu:

- **Krok 1:** Estymujemy model

$$y_t = x_t \beta + \varepsilon_t$$

- **Krok 2:** Na podstawie reszt z modelu liczymy następującą statystykę:

$$\boldsymbol{DW} = \frac{\sum_{t=2}^{T} (e_t - e_{t-1})^2}{\sum_{t=1}^{T} e_t^2} = \frac{\boxed{\sum_{t=2}^{T} e_t^{\,2} + \sum_{t=2}^{T} e_{t-1}^{\,2}} - 2\sum_{t=2}^{T} e_t e_{t-1}}{\sum_{t=1}^{T} e_t^2}$$

$$\frac{\boxed{2\sum_{t=1}^{T} e_t^{\,2}} - 2\sum_{t=1}^{T} e_t e_{t-1} - \boxed{e_1^{\,2} - e_T^{\,2}} + 2 e_1 e_0}{\sum_{t=1}^{T} e_t^2} \approx 2(1-\hat{\rho}) - \frac{e_1^{\,2} + e_T^{\,2} - 2 e_1 e_0}{\sum_{t=1}^{T} e_t^2} \xrightarrow{p} 2(1-\rho)$$

Testowanie autokorelacji

- **Test Durbina-Watsona (Test DW):**

- **Do badania autokorelacji I rzędu (między $\varepsilon_t, \varepsilon_{t-1}$)**

- **Rozkład statystki testowej wyprowadzony dla małych prób**

- **Nie można go stosować w modelach gdzie jedną ze zmiennych objaśniających jest opóźniona zmienna zależna**

Przykład

- Dane roczne za okres 1960 – 1995 dotyczące rynku paliwowego w Stanach Zjednoczonych.

- Zmienne, które wykorzystamy w regresji:
  - *G – konsumpcja benzyny wyrażona jako całkowite wydatki podzielone przez indeks cen;*
  - *Pg – indeks cen benzyny;*
  - *Y – PKB.*

Przykład

```
. tsset year /*zdefiniowanie zmiennej mierzącej przebieg czasu*/
        time variable:  year, 1960 to 1995
                delta:  1 unit

. regress g pg y
      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  2,    33) =  987.12
       Model |  88110.8184     2  44055.4092           Prob > F      =  0.0000
    Residual |  1472.79845    33  44.6302562           R-squared     =  0.9836
-------------+------------------------------           Adj R-squared =  0.9826
       Total |  89583.6168    35  2559.53191           Root MSE      =  6.6806
------------------------------------------------------------------------------
           g |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pg |  -15.12235   1.880337    -8.04   0.000    -18.94792   -11.29677
           y |   .0369204   .0013176    28.02   0.000     .0342397     .039601
       _cons |   -79.7535   8.672551    -9.20   0.000    -97.39794   -62.10906
------------------------------------------------------------------------------

. dwstat /*statystyka Durbina - Watsona*/
Durbin-Watson d-statistic(  3,    36) =  .4742979
```

Przykład

Wartości krytyczne testu Durbina-Watsona, $\alpha = 0.05$, K zawiera stałą

	$K=2$		$K=3$		$K=4$		$K=5$		$K=6$		$K=7$		$K=8$	
T	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U	d_L	d_U
6	0.61	1.40												
7	0.70	1.36	0.47	1.90										
8	0.76	1.33	0.56	1.78	0.37	2.29								
9	0.82	1.32	0.63	1.70	0.45	2.13	0.30	2.59						
10	0.88	1.32	0.70	1.64	0.53	2.02	0.38	2.41	0.24	2.82				
11	0.93	1.32	0.76	1.60	0.59	1.93	0.44	2.28	0.32	2.64	0.20	3.00		
12	0.97	1.33	0.81	1.58	0.66	1.86	0.51	2.18	0.38	2.51	0.27	2.83	0.17	3.15
13	1.01	1.34	0.86	1.56	0.71	1.82	0.57	2.09	0.44	2.39	0.33	2.69	0.23	2.99
14	1.04	1.35	0.91	1.55	0.77	1.78	0.63	2.03	0.51	2.30	0.39	2.57	0.29	2.85
15	1.08	1.36	0.95	1.54	0.81	1.75	0.69	1.98	0.56	2.22	0.45	2.47	0.34	2.73
16	1.11	1.37	0.98	1.54	0.86	1.73	0.73	1.94	0.61	2.16	0.50	2.39	0.40	2.62
17	1.13	1.38	1.02	1.54	0.90	1.71	0.78	1.90	0.66	2.10	0.55	2.32	0.45	2.54
18	1.16	1.39	1.05	1.54	0.93	1.70	0.82	1.87	0.71	2.06	0.60	2.26	0.50	2.46
19	1.18	1.40	1.07	1.54	0.97	1.69	0.86	1.85	0.75	2.02	0.65	2.21	0.55	2.40
20	1.20	1.41	1.10	1.54	1.00	1.68	0.89	1.83	0.79	1.99	0.69	2.16	0.59	2.34
21	1.22	1.42	1.12	1.54	1.03	1.67	0.93	1.81	0.83	1.96	0.73	2.12	0.64	2.29
22	1.24	1.43	1.15	1.54	1.05	1.66	0.96	1.80	0.86	1.94	0.77	2.09	0.68	2.25
23	1.26	1.44	1.17	1.54	1.08	1.66	0.99	1.79	0.89	1.92	0.80	2.06	0.71	2.21
24	1.27	1.45	1.19	1.55	1.10	1.66	1.01	1.78	0.92	1.90	0.84	2.04	0.75	2.17
25	1.29	1.45	1.21	1.55	1.12	1.65	1.04	1.77	0.95	1.89	0.87	2.01	0.78	2.14
26	1.30	1.46	1.22	1.55	1.14	1.65	1.06	1.76	0.98	1.87	0.90	1.99	0.82	2.12
27	1.32	1.47	1.24	1.56	1.16	1.65	1.08	1.75	1.00	1.86	0.92	1.97	0.85	2.09
28	1.33	1.48	1.26	1.56	1.18	1.65	1.10	1.75	1.03	1.85	0.95	1.96	0.87	2.07
29	1.34	1.48	1.27	1.56	1.20	1.65	1.12	1.74	1.05	1.84	0.97	1.94	0.90	2.05
30	1.35	1.49	1.28	1.57	1.21	1.65	1.14	1.74	1.07	1.83	1.00	1.93	0.93	2.03
31	1.36	1.50	1.30	1.57	1.23	1.65	1.16	1.74	1.09	1.83	1.02	1.92	0.95	2.02
32	1.37	1.50	1.31	1.57	1.24	1.65	1.18	1.73	1.11	1.82	1.04	1.91	0.97	2.00
33	1.38	1.51	1.32	1.58	1.26	1.65	1.19	1.73	1.13	1.81	1.06	1.90	0.99	1.99
34	1.39	1.51	1.33	1.58	1.27	1.65	1.21	1.73	1.14	1.81	1.08	1.89	1.01	1.98
35	1.40	1.52	1.34	1.58	1.28	1.65	1.22	1.73	1.16	1.80	1.10	1.88	1.03	1.97
36	1.41	1.52	1.35	1.59	1.30	1.65	1.24	1.72	1.18	1.80	1.11	1.88	1.05	1.96

Testowanie autokorelacji

- Test Breuscha-Godfreya (Test BG):

- Do badania autokorelacji wyższego rzędu

- Można go stosować w modelach, gdzie występują opóźnione zmienne zależne

Testowanie autokorelacji

- Test Breuscha-Godfreya (Test BG):

$$H_0 : Cov(\varepsilon_{t,} \varepsilon_{t-i}) = 0$$ **gdzie** $i = 1, \dots, s$

$$H_1 : \varepsilon_t = \gamma_1 \varepsilon_{t-1} + \dots + \gamma_s \varepsilon_{t-s} + u_t$$ **gdzie** $Var(u) = \sigma_u^2 I$

- Hipoteza zerowa: brak autokorelacji

- Hipoteza alternatywna: autokorelacja

Testowanie autokorelacji – Test Breuscha-Godfreya (Test BG)

- **Sposób przeprowadzania:**

Krok 1: Estymujemy model $y_t = x_t\beta + \varepsilon_t$ i uzyskujemy reszty e_t

Krok 2: Przeprowadzamy regresję pomocniczą reszt z tej regresji na resztach opóźnionych:

$$e_t = x_t\mu + \gamma_1 e_{t-1} + \cdots + \gamma_s e_{t-s} + u_t$$

Krok 3: Przetestować hipotezę o braku autokorelacji możemy testując:

$$H_0: \gamma_1 = \cdots = \gamma_s = 0$$

Można to zrobić za pomocą statystyki F lub statystyki

$$TR^2 \xrightarrow{D} {\chi_s}^2$$

Przykład

1) Szacujemy regresję i wyznaczamy z niej reszty:

```
. tsset year /*zdefiniowanie zmiennej mierzącej przebieg czasu*/
        time variable:  year, 1960 to 1995
                delta:  1 unit

. regress g pg y
      Source |       SS       df       MS              Number of obs =      36
-------------+------------------------------           F(  2,    33) =  987.12
       Model |  88110.8184     2  44055.4092           Prob > F      =  0.0000
    Residual |  1472.79845    33  44.6302562           R-squared     =  0.9836
-------------+------------------------------           Adj R-squared =  0.9826
       Total |  89583.6168    35  2559.53191           Root MSE      =  6.6806
------------------------------------------------------------------------------
           g |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pg |  -15.12235   1.880337    -8.04   0.000    -18.94792   -11.29677
           y |   .0369204   .0013176    28.02   0.000     .0342397     .039601
       _cons |   -79.7535   8.672551    -9.20   0.000    -97.39794   -62.10906
------------------------------------------------------------------------------

. predict e, residual /*Tworzymy reszty*/
```

Przykład

2) Przeprowadzamy regresję pomocniczą:

$$e_t = \alpha_0 + \alpha_1 pg_t + \alpha_2 y_t + \boxed{\gamma_1 e_{t-1} + \gamma_2 e_{t-2} + \gamma_3 e_{t-3} + \gamma_4 e_{t-4}} + \xi_t, \quad t = 1, 2, ..T$$

```
. /*Regresja pomocnicza*/

. regress e pg y l.e l2.e l3.e l4.e /*l.e,l2.e,l3.e,l4.e - odpowiednio reszty opóźnione o 1, 2, 3 i 4
okresy*/

      Source |       SS       df       MS              Number of obs =      32
-------------+------------------------------           F(  6,    25) =    6.94
       Model |  919.397609     6  153.232935           Prob > F      =  0.0002
    Residual |  551.671362    25  22.0668545           R-squared     =  0.6250
-------------+------------------------------           Adj R-squared =  0.5350
       Total |  1471.06897    31  47.4538378           Root MSE      =  4.6975

------------------------------------------------------------------------------
           e |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
          pg |   -1.07159   1.711641    -0.63   0.537    -4.596781    2.453601
           y |   .0005598   .0012961     0.43   0.670    -.0021097    .0032292
             |
           e |
         L1. |   .9448478   .2016664     4.69   0.000     .5295081    1.360187
         L2. |  -.3996707   .2731132    -1.46   0.156    -.9621578    .1628165
         L3. |   .3243404   .2747752     1.18   0.249    -.2415697    .8902505
         L4. |  -.0580117   .2226776    -0.26   0.797    -.5166249    .4006014
             |
       _cons |  -2[illegible]2475   8.886971    -0.32   0.753    -21.12553    15.48058
------------------------------------------------------------------------------
```

Przykład

$$LM = T \cdot R^2 \xrightarrow{D} \chi^2_4$$

```
. /*Statystyka testowa*/
. display e(r2)*(e(N)) /*e(r2) - R^2; e(N) - liczba obseracji*/
19.999554
```

```
. /*P-value*/
. display chi2tail(4, e(r2)*(e(N)))
.0004995
```

Przykład

```
. bgodfrey, lags(4) nomiss0 /* test Breuscha - Godfrey'a na autokorelacje rzędu 4;
lags( ) - definiuje maksymalny rząd opóźnienia reszt w regresji pomocniczej; nomiss -
powoduje, że regresja pomocnicza jest przeprowadzana tylko na obserwacjach, które nie
zawierają braków danych. Opóźnianie zmiennej e powoduje pojawianie się braków danych.
Domyślnie w teście Breuscha - Godfreya braki danych zastępowane są przez 0 i regresja
pomocnicza przeprowadzana jest na całej próbie. */

Breusch-Godfrey LM test for autocorrelation
---------------------------------------------------------------------------
    lags(p)   |          chi2               df                 Prob > chi2
-------------+-------------------------------------------------------------
       4      |         20.000               4                   0.0005
---------------------------------------------------------------------------
                        H0: no serial correlation
```

Jakie założenie KMRL nie jest spełnione przy odrzuceniu H0?

- Brak autokorelacji błędu losowego – kowariancja dwóch różnych błędów losowych jest zerowa:

$$\operatorname{cov}(\varepsilon_i, \varepsilon_j) = 0 \quad dla \quad i \neq j$$

Pytania teoretyczne

1. Za pomocą jakich testów testuje się autokorelację? Jakiemu założeniu KMRL odpowiada H_0 w tych testach? Jakie są hipotezy alternatywne w tym testach?

Dziękuję za uwagę