Heteroskedastyczność i autokorelacja (cz. II)

Natalia Nehrebecka
Stanisław Cichocki

Wykład 12

Plan zajęć

1. Konsekwencje heteroskedastyczności i autokorelacji

2. Uogólniona MNK

3. Stosowalna Uogólniona MNK

4. Odporne macierze wariancji i kowariancji b

Plan zajęć

1. Konsekwencje heteroskedastyczności i autokorelacji

2. Uogólniona MNK

3. Stosowalna Uogólniona MNK

4. Odporne macierze wariancji i kowariancji b

Heteroskedastyczność i autokorelacja

- Jeżeli założenie o homoskedastyczności lub założenie o braku autokorelacji nie jest spełnione, to mówimy o niesferyczności błędów losowych

- W przypadku występowania heteroskedastyczności lub autokorelacji:

$$Var(\varepsilon) = \Omega = \sigma^2 V$$

SUMNK

- Gdy elementy macierzy V są nieznane $\longrightarrow$ używamy estymatora UMNK zastępując V oszacowaniem $\hat{V}$

$\longrightarrow$ metoda ta to Stosowalna Uogólniona Metoda Najmniejszych Kwadratów (*Feasible Generalized Least Squares*)

Stosowalna Uogólniona MNK

- **Istota:** budujemy dodatkowy model dla wariancji i kowariancji tak aby znaleźć funkcję $\Omega(\theta)$

- Zakładamy, że

$$Var(\varepsilon) = \Omega(\theta)$$

  - gdzie wymiar $\boldsymbol{\theta}$ nie zależy od liczby obserwacji.
  - Dla znanego $\boldsymbol{\theta}$ moglibyśmy zastosować UMNK, jednak zwykle $\boldsymbol{\theta}$ jest nieznane.
  - Wiemy, że estymator MNK jest zgodny nawet wtedy, gdy występuje autokorelacja lub heteroskedastyczność.

Przykład

- Typowy przykład zastosowania SUMNK – usuwanie heteroskedastyczności z modelu

- Heteroskedastyczność ma postać zależności funkcyjnej:

$$\sigma_i{}^2 = \alpha_0 + z_i\alpha$$

- a więc

$$Var(\varepsilon) = \Omega(\alpha_0,\, \alpha)$$

- Można ją wykryć za pomocą testu Breuscha-Pagana.

SUMNK

- Etapy estymacji za pomocą SUMNK:

1. Szacujemy model za pomocą MNK, uzyskując zgodny estymator b_{MNK}

2. Na podstawie wektora reszt e szacujemy model pomocniczy, z którego otrzymujemy zgodny estymator θ

3. Powtórnie szacujemy parametry modelu stosując estymator UMNK, nieznaną macierz Ω zastępujemy jej oszacowaniem $\Omega(\hat{\theta})$

SUMNK

4. Uzyskany estymator SUMNK:

$$b_{SUMNK} = \left[X'\Omega^{-1}\left(\widehat{\theta}\right)X \right]^{-1} X'\Omega^{-1}\left(\widehat{\theta}\right)y$$

Przykład 1

- Typowy przykład zastosowania SUMNK – usuwanie heteroskedastyczności z modelu
- Heteroskedastyczność ma postać zależności funkcyjnej:

$$\sigma_i{}^2 = \alpha_0 + z_i\alpha$$

- a więc

$$Var(\varepsilon) = \Omega(\alpha_0, \alpha)$$

- Można ją wykryć za pomocą testu Breuscha-Pagana.

Przykład 1

- **Najczęściej jednak** α_0, α **są nieznane**
- W tym przypadku możemy posłużyć się metodą dwustopniową
- Estymator MNK jest nieobciążony i zgodny
- $\boldsymbol{e^2}$ - kwadraty reszt z MNK stanowią pewne oszacowanie $\boldsymbol{\sigma^2}$
- Regresja $\boldsymbol{e^2}$ na stałej i $\boldsymbol{z_i}$ da zgodny estymator $\boldsymbol{\alpha_0, \alpha}$
- Dla tak policzonych estymatorów $\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}, \widehat{\boldsymbol{\alpha}}$ można policzyć wartości teoretyczne $\hat{\sigma}_i{}^2 = \widehat{\boldsymbol{\alpha}}_{\mathbf{0}} + \boldsymbol{z_i}\widehat{\boldsymbol{\alpha}}$ dla poszczególnych obserwacji i uzyskać macierz $\Omega(\widehat{\boldsymbol{\alpha}}_{\mathbf{0}}, \widehat{\boldsymbol{\alpha}})$

Przykład 1

Stosowalna UMNK polega na wyestymowaniu MNK równania:

$$y_i{}^* = x_i{}^* \beta + \varepsilon_i{}^*$$

- gdzie:

- $y_i{}^* = \frac{y_i}{\hat{\sigma}_i}$
- $x_i{}^* = \frac{x_i}{\hat{\sigma}_i}$

- Uzyskany w ten sposób estymator będzie asymptotycznie zgodnym i efektywnym

Przykład 2

- **Wydatki na żywność**

Source	SS	df	MS
Model	89.6721174	6	14.9453529
Residual	275.09173	3339	.08238746
Total	364.763848	3345	.109047488

Number of obs	=	3346	
F(6, 3339)	=	181.40	
Prob > F	=	0.0000	
R-squared	=	0.2458	
Adj R-squared	=	0.2445	
Root MSE	=	.28703	

lq	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
linc	.3540467	.0113581	31.17	0.000	.3317772	.3763162
_Iklm_2	-.0334229	.0197196	-1.69	0.090	-.0720866	.0052408
_Iklm_3	-.0584767	.0213409	-2.74	0.006	-.1003194	-.0166341
_Iklm_4	-.0325534	.0176063	-1.85	0.065	-.0670736	.0019668
_Iklm_5	-.0423542	.019028	-2.23	0.026	-.0796619	-.0050465
_Iklm_6	-.0203535	.0189354	-1.07	0.283	-.0574796	.0167726
_cons	3.749705	.090792	41.30	0.000	3.571692	3.927719

Przykład 2

- Regresja ***e*²** na stałej i *klm*

resid2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
_Iklm_2	.0068093	.0091102	0.75	0.455	-.0110528	.0246715
_Iklm_3	-.016938	.0098739	-1.72	0.086	-.0362975	.0024215
_Iklm_4	-.0129282	.0080402	-1.61	0.108	-.0286926	.0028361
_Iklm_5	-.0088624	.0086485	-1.02	0.306	-.0258194	.0080945
_Iklm_6	-.0133894	.0084823	-1.58	0.115	-.0300205	.0032416
_cons	.0906276	.0066867	13.55	0.000	.0775171	.1037381

- Test na łączną nieistotność zmiennej *klm:*

```
F(  5,  3340) =    2.19
     Prob > F =    0.0526
```

Przykład 2

- Regresja ***e²*** na stałej i *linc*

resid2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
linc	.017435	.005053	3.45	0.001	.0075278	.0273423
_cons	-.0504988	.0385321	-1.31	0.190	-.1260476	.02505

Przykład 2

- Przyjęto, że zależność między wariancją błędu losowego i ma postać:

$$\hat{\sigma}_i^{\,2} = \hat{\alpha}_0 + \hat{\alpha}_1 linc$$

resid2	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
linc	.017435	.005053	3.45	0.001	.0075278	.0273423
_cons	-.0504988	.0385321	-1.31	0.190	-.1260476	.02505

Przykład 2

- Z regresji tej najpierw wartości dopasowane

$$\hat{\sigma}_i^{\,2} = \hat{\alpha}_0 + \hat{\alpha}_1 linc$$

- Podzielono **zmienną zależną** i **wszystkie zmienne niezależne** (łącznie ze stałą i zmiennymi zerojedynkowymi) w pierwotnej regresji przez $\hat{\sigma}_i$

$$\hat{\sigma}_i = \sqrt{\hat{\alpha}_0 + \hat{\alpha}_1 linc}$$

Przykład 2

```
      Source |       SS       df       MS              Number of obs =    3346
-------------+------------------------------           F(  7,  3339) =       .
       Model |  1682677.41     7  240382.487           Prob > F      =  0.0000
    Residual |  3345.60972  3339  1.00197955           R-squared     =  0.9980
-------------+------------------------------           Adj R-squared =  0.9980
       Total |  1686023.02  3346  503.892116           Root MSE      =   1.001

------------------------------------------------------------------------------
        t_lq |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      t_linc |   .3603508   .0114135    31.57   0.000     .3379727     .382728
    t_Iklm_2 |  -.0330761   .0199833    -1.66   0.098    -.0722569     .006104
    t_Iklm_3 |  -.0565737   .0216039    -2.62   0.009    -.0989319    -.014215

    t_Iklm_4 |  -.0290265   .0178167    -1.63   0.103    -.0639593     .00590
    t_Iklm_5 |  -.0374901   .0191635    -1.96   0.051    -.0750634     .00009
    t_Iklm_6 |   -.017306   .0190144    -0.91   0.363    -.054587      .0199
      t_cons |   3.699076   .0906672    40.80   0.000     3.521307     3.876
------------------------------------------------------------------------------
```

Oszacowanie błędu standardowego *s* powinno być równe 1

Przykład 2

- Standardowo uzyskiwana **wielkość statystyki *F*** jest **niepoprawna** (przekształcona stała jest traktowana jako jedna ze zmiennych). Poprawny test ***F*** powinien testować łączną istotność wszystkich zmiennych poza przekształconą stałą.

```
F(  6,  3339) =  183.74
     Prob > F =    0.0000
```

- ***R*²** uzyskane w takiej regresji **nie jest interpretowalne** zmienna zależna została stworzona sztucznie
- Można policzyć R^2 dla oryginalnej regresji i uzyskanych z UMNK wartości dopasowanych

$$R^2_{UMNK} = \frac{\sum \left(\widehat{y}_{UMNK,i} - \overline{\widehat{y}}_{UMNK} \right)^2}{\sum \left(y_i - \overline{y} \right)^2} = \rho^2_{\widehat{y},y}$$

Policzona wartość R^2_{UMNK} =.28674733

Przykład 2

- Wynik standardowego testu Breuscha-Pagana jest teraz następujący:

```
Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         chi2      =    0.00905
         Prob > chi2  =   0.9242
```

- Heteroskedastyczność udało się usunąć!

Przykład 2

lq	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
linc	.3540467	.0113581	31.17	0.000	.3317772	.3763162
_Iklm_2	-.0334229	.0197196	-1.69	0.090	-.0720866	.0052408
_Iklm_3	-.0584767	.0213409	-2.74	0.006	-.1003194	-.0166341
_Iklm_4	-.0325534	.0176063	-1.85	0.065	-.0670736	.0019668
_Iklm_5	-.0423542	.019028	-2.23	0.026	-.0796619	-.0050465
_Iklm_6	-.0203535	.0189354	-1.07	0.283	-.0574796	.0167726
_cons	3.749705	.090792	41.30	0.000	3.571692	3.927719

t_lq	Coef.	Std. Err.	t	P>|t|	[95% Conf.	Interval]
t_linc	.3603508	.0114135	31.57	0.000	.3379727	.3827289
t_Iklm_2	-.0330761	.0199833	-1.66	0.098	-.0722569	.0061048
t_Iklm_3	-.0565737	.0216039	-2.62	0.009	-.0989319	-.0142156
t_Iklm_4	-.0290265	.0178167	-1.63	0.103	-.0639593	.0059063
t_Iklm_5	-.0374901	.0191635	-1.96	0.051	-.0750634	.0000832
t_Iklm_6	-.017306	.0190144	-0.91	0.363	-.054587	.0199751
t_cons	3.699076	.0906672	40.80	0.000	3.521307	3.876845

Przykład 3

- SUMNK w modelach z autokorelacją
  - Najpopularniejszym zastosowaniem tej metody estymacji jest estymator Praisa-Winstena
  - SAMODZIELNIE!

Plan zajęć

1. Konsekwencje heteroskedastyczności i autokorelacji

2. Uogólniona MNK

3. Stosowalna Uogólniona MNK

4. Odporne macierze wariancji i kowariancji b

Odporne macierze wariancji i kowariancji b

- Istnieją **estymatory macierzy wariancji i kowariancji**, które są **zgodne** w przypadku występowania heteroskedastyczności lub autokorelacji

→ odporne (robust) estymatory wariancji

- Stosujemy różne estymatory w zależności od tego czy w modelu występuje heteroskedastyczność czy autokorelacja

  - Najpopularniejszym odpornym na **heteroskedastyczność** estymatorem macierzy wariancji i kowariancji *b* jest estymator White'a

  - Estymator Newey'a-Westa macierzy wariancji i kowariancji *b* stosujemy gdy w modelu występuje **heteroskedastyczność i autokorelacja**

  - Estymator warstwowy (autokorelacja błędu losowego może być skutkiem częściowo nielosowego doboru próby)

Konsekwencje heteroskedastyczności i autokorelacji

- Macierz wariancji i kowariancji b:

$$Var(b) = E\left((X'X)^{-1}X'\varepsilon\varepsilon'X(X'X)^{-1}\right) =$$

$$(X'X)^{-1}X'\Omega X(X'X)^{-1} =$$

$$\sigma^2(X'X)^{-1}X'VX(X'X)^{-1}$$

Estymator odporny White

- Jeśli znalibyśmy macierz wariancji-kowariancji ***Ω***, wtedy estymatorem macierzy wariancji-kowariancji wektora parametrów **β** byłoby

$$Var(b) = \frac{1}{N}\left(\frac{1}{N}X'X\right)^{-1}\left(\frac{1}{N}X'\Omega X\right)\left(\frac{1}{N}X'X\right)^{-1}$$

- Jednak macierz ***Ω*** nie jest znana.
- Zachodzi wiec konieczność oszacowania **N(N+1)/2** nieznanych parametrów macierzy na podstawie **N** obserwacji.
- White (1980) pokazał, ze rozwiązaniem jest odmienne spojrzenie na problem.
  - To co jest istotne to uzyskanie ***zgodnego estymatora dla macierzy X'ΩX***, która ma wymiar $K \times K$
  - Ponadto liczba zmiennych w modelu jest zazwyczaj stała i nie zależy od rozmiaru próby.

- Oznaczmy przez x_i i-ty wiersz macierzy obserwacji *X*.

$$X'\Omega X = X'\sigma^2 VX = \sum_{i=1}^{N} {\sigma_i}^2 {x'}_i x_i$$

Estymator odporny White

- White zaproponował by nieznane wariancje zastąpić kwadratami reszt.
- W ten sposób uzyskany estymator jest zgodny.

$$\hat{S}_0 = \frac{1}{N}\sum_{i=1}^{N} e_i^2 x_i' x_i$$

- W rezultacie otrzymujemy estymator White'a, który jest zgodny w przypadku heteroscedastyczności.

$$\hat{\Sigma}_b = \frac{1}{N}\left(\frac{1}{N}X'X\right)^{-1}\hat{S}_0\left(\frac{1}{N}X'X\right)^{-1}$$

Przykład

- Przeanalizujemy model popytu na prace zgłaszanego przez belgijskie firmy. Próba zawiera dane z 570 firm z roku 1996. Dostępne są następujące zmienne:

- labor - zatrudnienie
- wage - suma pensji podzielona przez liczbę pracowników (w milionach jednostek)
- output - wartość dodana produkcji (w milionach jednostek)
- capital - wartość majątku trwałego (w milionach jednostek)

Przykład

```
. reg  lnlabor lnwage lnoutput lncapital

      Source |       SS       df       MS              Number of obs =     569
-------------+------------------------------           F(  3,   565) = 1011.02
       Model |  656.747032     3  218.915677           Prob > F      =  0.0000
    Residual |  122.338815   565  .216528876           R-squared     =  0.8430
-------------+------------------------------           Adj R-squared =  0.8421
       Total |  779.085847   568  1.37163001           Root MSE      =  .46533

------------------------------------------------------------------------------
     lnlabor |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lnwage |  -.9277642   .0714046   -12.99   0.000    -1.068015   -.7875132
    lnoutput |   .9900474   .0264103    37.49   0.000      .938173    1.041922
   lncapital |  -.0036975   .0187697    -0.20   0.844    -.0405644    .0331695
       _cons |  -.4480909   .0932397    -4.81   0.000    -.6312296   -.2649522
------------------------------------------------------------------------------
```

Przykład

```
. hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of lnlabor

         chi2(1)      =    19.49
         Prob > chi2  =   0.0000
```

Przykład

```
. reg  lnlabor lnwage lnoutput lncapital, robust

Regression with robust standard errors                 Number of obs =     569
                                                       F(  3,   565) =  544.73
                                                       Prob > F      =  0.0000
                                                       R-squared     =  0.8430
                                                       Root MSE      =  .46533

------------------------------------------------------------------------------
             |               Robust
     lnlabor |      Coef.   Std. Err.      t    P>|t|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
      lnwage |  -.9277642    .0866604   -10.71   0.000    -1.09798   -.7575483
    lnoutput |   .9900474    .0467902    21.16   0.000    .8981434    1.081951
   lncapital |  -.0036975     .037877    -0.10   0.922   -.0780944    .0706995
       _cons |  -.4480909    .1332882    -3.36   0.001   -.7098918     -.18629
------------------------------------------------------------------------------
```

Estymacja macierzy wariancji kowariancji metodą Newey'a-Westa

- Analogicznie do odpornego na heteroskedastyczność estymatora White`a *ekonometrycy Newey i West* zaproponowali odporny na heteroskedastyczność i na autokorelacje (o niesprecyzowanej strukturze) estymator macierzy wariancji-kowariancji dla *b*, oszacowanego za pomocą MNK.

$$\hat{\Sigma}_b = T\left(X'X\right)^{-1}\hat{S}_x\left(X'X\right)^{-1}$$

- Gdzie:

$$\hat{S}_x = \hat{S}_0 + \frac{1}{T}\sum_{l=1}^{L}\sum_{t=l+1}^{T}\omega_l e_t e_{t-l}(x_t x_{t-l}^{'} - x_{t-l} x_t^{'})$$

$$\omega_l = 1 - \frac{l}{L+1}$$

Pytania teoretyczne

1. Wyjaśnić różnicę między UMNK i SUMNK.
2. (*) Opisać proces szacowania modelu SUMNK na przykładzie usuwania heteroskedastyczności z modelu bądź estymacji za pomocą estymatora Praisa-Winsterna.
3. Jakie są zalety stosowania estymatora MNK w połączeniu z estymatorem odpornym macierzy wariancji i kowariancji w porównaniu do stosowania estymatora UMNK.

Dziękuję za uwagę