Metody Ekonometryczne
Niesferyczność macierzy wariancji kowariancji składnika losowego

Jakub Mućk
Szkoła Główna Handlowa w Warszawie

Estymator MNK

- Dla jednorównaniowego modelu liniowego:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{1}$$

estymator MNK (OLS) przyjmuję postać:

$$\hat{\beta}^{OLS} = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbf{y}. \tag{2}$$

- Estymator wariancji-kowariancji można zapisać jako:

$$Var(\hat{\beta}^{OLS}) = \mathbb{S}_{\varepsilon}^2(\mathbf{X}^T\mathbf{X})^{-1}, \tag{3}$$

gdzie wariancja składnika losowego można oszacować jako $\mathbb{S}_{\varepsilon}^2$:

$$\mathbb{S}_{\epsilon}^2 = \frac{e^Te}{n-(k+1)} = \frac{SSE(\hat{\beta}^{OLS})}{df} \tag{4}$$

gdzie $SSE(\hat{\beta}^{OLS})$ to suma kwadratów reszt, a df to **liczba stopni swobody**.

Założenia:

1. $rz(\mathbf{X}) = k + 1 \leq n$
2. $\mathbb{E}(\mathbf{X}\varepsilon) = 0$
3. $\mathbb{E}(\varepsilon) = 0$
4. $Var(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon^T) = I\sigma^2$
5. $\varepsilon_i \sim N(0, \sigma^2)$

Twierdzenie Gaussa - Markowa

Estymator $\hat{\beta}$ uzyskany Klasyczną Metodą Najmniejszych Kwadratów jest estymatorem **BLUE** [*best linear unbiased estimator*], tj. zgodnym, nieobciążonym i najefektywniejszy w klasie liniowych estymatorów wektora β.

- nieobciążoność, czyli: $\mathbb{E}(\hat{\beta}^{OLS}) = \beta$
- najefektywniejszy, czyli posiadający najmniejszą wariancję w swojej klasie
- zgodny, czyli: $\text{plim}_{n\to\infty}\hat{\beta}_n^{OLS} = \beta$.

Niesferyczność macierzy wariancji-kowariancji składnika losowego

- Przypomnijmy estymator macierzy wariancji-kowariancji oszacowań $(\hat{\beta}^{OLS})$:

$$Var(\hat{\beta}^{OLS}) = \mathbb{E}\left[\left(\hat{\beta}^{OLS} - \beta\right)\left(\hat{\beta}^{OLS} - \beta\right)^T\right] \tag{5}$$

- Korzystając z zapisu macierzowego:

$$\begin{aligned} Var(\hat{\beta}^{OLS}) &= \mathbb{E}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\left(\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\right)^T\right] \\ &= \mathbb{E}\left[\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\varepsilon\varepsilon^T\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\right] \\ &= \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbb{E}\left[\varepsilon\varepsilon^T\right]\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1} \end{aligned}$$

- Następnie korzystając z założenie o **sferyczności macierzy wariancji-kowariancji składnika losowego**, tj. $\mathbb{D}^2(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon^T) = \sigma^2 I$, można uprościć wzór na estymator wariancji kowariancji oszacowań do:

$$Var(\hat{\beta}^{OLS}) = \sigma^2\left(\mathbf{X}^T\mathbf{X}\right)^{-1}. \tag{6}$$

Konsekwencje niesferyczności macierzy wariancji-kowariancji składnika losowego

- **Brak sferyczności macierzy wariancji-kowariancji składnika losowego** prowad do **obciążenia macierzy wariancji-kowariancji parametrów strukturalnych** $Var(\hat{\beta}^{OLS})$.
- Naturalną konsekwencją jest **brak wiarygodności błędów standardowych**.
- **Obciążene są również wyniki testów statystycznych** bazujących na macierzy wariancji-kowariancji wektora parametrów strukturalnych. W szczególności test t-studenta czy test liniowych restrykcji (test Walda).
- **Brak sferyczności macierzy wariancji-kowariancji składnika losowego** może świadczyć o poważniejszych problemach jak np.
  - **problemie pominiętych zmiennych (*omitted variable bias*)** .

- Autokorelacja składnika losowego
- Heteroskedastyczność

- Autokorelacja składnika losowego
- Heteroskedastyczność
- Efekt ARCH (*Autoregressive Conditional Heteroscedasticity*).
- Zależność przestrzenna
- Zależność pomiędzy jednostkami panelu (*cross-sectional dependence*)

- Próba wyjaśnienia źródeł niesferyczności.
- Zmiana specyfikacji modelu.
- Odporne estymatory wariancji-kowariancji (*robust covariance estimators*).
  - Konsekwencją niesferyczności składnika losowego jest obciążoność macierzy wariancji-kowariancji. Dlatego rozwiązaniem estymatora wariancji-kowariancji uwzględniającego (*odpornego*) tę własność skłanika losowego.
- Inne metody estymacji parametrów strukturalnych:
  - Uogólniona MNK (*GLS - Generalized Least Squares*)
  - Ważona MNK w przypadku heteroskedastyczności.

Autokorelacja składnika losowego

- **Autokorelacja składnika losowego** jest problemem najczęściej występującym w przypadku **szeregów czasowych** i polega na zależności (skorelowaniu) bieżących wartości składnika losowego od wartości przeszłych.
- **Autokorelację składnika losowego** można wiązać z **inercją/persystencją** zmiennych/procesów ekonomicznych. Jest to własność polegająca na rozłożonej w czasie absorbcji czynników zewnętrznych.
- Indeks t będzie oznaczać czas obserwacji.
- Zgodnie z założenia MNK:

$$Var(\varepsilon_t) = \begin{bmatrix} \sigma^2 & 0 & \dots & \dots & 0 \\ 0 & \ddots & \ddots & \ddots & \vdots \\ \vdots & \ddots & \sigma^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \ddots & 0 \\ 0 & \dots & \dots & 0 & \sigma^2 \end{bmatrix}$$

co jest równoznaczne:

$$\forall_{t \neq s} cov(\varepsilon_t, \varepsilon_s) = 0$$

- **Autokorelacja składnika losowego pierwszego rzędu AR(1)**:
$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \tag{7}$$
gdzie $\eta_t \sim \mathcal{N}(0, \sigma^2_\eta)$, $\mathbb{E}(\eta) = 0$ oraz $Var(\eta)^2 = I\sigma^2_\eta$. Zakładamy również, że $|\rho| < 1$
- Wariancja składnika losowego:
$$Var(\varepsilon_t) = Var(\rho\varepsilon_{t-1} + \eta_t). \tag{8}$$
Korzystając z definicji (60) można podmienić ε_{t-1}:
$$Var(\varepsilon_t) = Var(\rho^2\varepsilon_{t-2} + \rho\eta_{t-1} + \eta_t). \tag{9}$$
- Iterując powyższą czynność uzsykujemy:
$$Var(\varepsilon_t) = Var(\eta_t + \rho\eta_{t-1} + \rho^2\eta_{t-2} + \ldots). \tag{10}$$
- η_t są niezależne w czasie a więc
$$Var(\varepsilon_t) = \sigma^2_\eta + \rho\sigma^2_\eta + \rho^2\sigma^2_\eta + \ldots, \tag{11}$$
- a więc wariancja składnika losowego jest równa
$$Var(\varepsilon_t) = \frac{\sigma^2_\eta}{1-\rho}. \tag{12}$$
- W przypadku kowariancji warto *iść krok po kroku*:
$$cov(\varepsilon_t, \varepsilon_{t-1}) = cov(\varepsilon_{t-1} + \eta_t, \varepsilon_{t-1}) = \rho cov(\varepsilon_{t-1}, \varepsilon_{t-1}) = \rho Var(\varepsilon_{t-1}) = \rho\frac{\sigma^2_\eta}{1-\rho}. \tag{13}$$

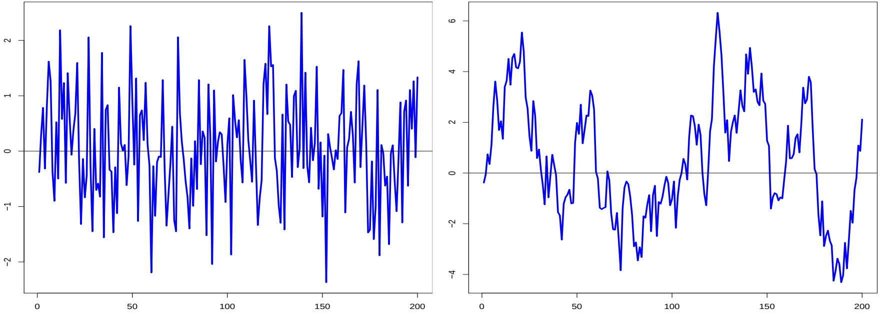
- Łatwo pokazać ogólną zależność:

$$cov(\varepsilon_t, \varepsilon_{t-k}) = \rho^k \frac{\sigma_\eta^2}{1-\rho}. \tag{14}$$

- Zatem, w przypadku autokorelacji składnika losowego macierz wariancji-kowariancji składnika losowego nie jest diagonalna:

$$Var(\varepsilon_t) = \frac{\sigma_\eta^2}{1-\rho} \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix}.$$

Przykład idiosynkratycznego zaburzenia losowego (L) oraz AR(1) (P)



Przyczyny autokorelacji składnika losowego:

- Wysoka **inercja (persystencja)** zjawisk gospodarczych.
- Psychologia podejmowanych zjawisk.
- Problem pominięcia ważnej zmiennej.
- Niepoprawna postać funkcyjna; wadliwa struktura dynamiczna, brak uwzględnienie czynników cyklicznych/sezonowych.
- Przekształcenia statystyczne.

Analiza reszt (e_t):

- wykresy reszt w czasie,
- scatterploty, tj. wykres reszt e_t względem opóźnionych reszt, np .e_{t-1}.

Testowanie autokorelacji składnika losowego

- test Durbina-Watsona,
- testy *portmanteau*,
- test Breuscha-Godfreya.

- Test Durbina-Watsona umożliwia sprawdzenie jedynie autokorelacji pierwszego rzędu.

SGH

- Statystyka testu DW opiera się na oszacowaniu współczynnika korelacji pomiędzy e_t a e_{t-1}:

$$d = \frac{\sum_{t=2}^{n} (e_t - e_{t-1})^2}{\sum_{t=1}^{n} e_{t-1}^2} \tag{15}$$

Łatwo zauważyć, że $d \approx 2(1 - \hat{\rho})$. Hipotezą zerową jest brak autokorelacji, tj.:

$$\mathcal{H}_0 : \qquad \rho = 0 \tag{16}$$

Natomiast **hipoteza alternatywna testu DW zależy od wartości statystyki testowej:**, tj.

$$\mathcal{H}_1 : \qquad \rho > 0 \qquad \text{gdy} \quad d \in (0, 2) \tag{17}$$

$$\mathcal{H}_1 : \qquad \rho < 0 \qquad \text{gdy} \quad d \in (2, 4) \tag{18}$$

- Wartości krytyczne d_U i d_L są stablicowane.

$\mathcal{H}_1 : \quad \rho > 0$		$\mathcal{H}_1 : \quad \rho < 0$	
Statystyka d	Decyzja	Statystyka d	Decyzja
$(0, d_L)$	są podstawy do odrzucenia $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$ o dodatniej autokorelacji	$(4 - d_L, 4)$	są podstawy do odrzucenia $\mathcal{H}_0$ na rzecz $\mathcal{H}_1$ o ujemnej autokorelacji
(d_L, d_U)	brak decyzji	$(4 - d_L, 4 - d_U)$	brak decyzji
$(d_U, 2)$	nie ma podstaw do odrzucenia $\mathcal{H}_0$	$(4 - d_U, 2)$	nie ma podstaw do odrzucenia $\mathcal{H}_0$

- Test DW posiada **obszary niekonkluzywności**.
- Test Durbina-Watsona umożliwia weryfikację autokorelacji jedynie pierwszego rzędu.
- W specyfikacji modelu ekonometrycznego nie może zostać uwzględniona część autoregresyjna zmiennej objaśniane (późnione wartości zmiennej objaśnianej), ponieważ wtedy statystyka DW jest obciążona.
- Test Durbina-Watsona można stostować w przypadku modeli z wyrazem wolnym.

- Statystyka **uogólnionego testu Durbina-Watsona** nie wykazuje obciążenia w przypadku modeli autoregresyjnych.
- Hipoteza zerowa odnosi się do braku autokorelacji.
- Statystyka testowa:

$$d^{\mathcal{G}} = \left(1 - \frac{d}{2}\right)\sqrt{\frac{T}{1 - TVar(\hat{\beta}_y)}} \tag{19}$$

  gdzie $Var(\hat{\beta}_y)$ to wariancja szacunku parametru autoregresji, a d to statystyka podstawowego testu Durnina-Watsona.
- Statystyka $d^{\mathcal{G}}$ posiada standardowy rozkład normalny, tj. $d^{\mathcal{G}} \sim \mathcal{N}(0, 1)$.

- Generalna konstrukcja testów *portmanteau* polega na weryfikacji statystycznej zależności w czasie (tj. autokorelacji) do rzędu P włącznie.
- Hipotezy zerowa postuluje **brak autokorelacji do rzędu P włącznie.**
- Niech $\hat{\rho}_k$ będzię korelacją pomiedzy resztami e_t a resztami opóźnionymi o k okresów, tj. e_{t-k}.
- **Statystyka Boxa-Pierca:**

$$Q_{BP} = T \sum_{i=1}^{P} \hat{\rho}_T^2 \tag{20}$$

ma rozkład χ^2 z P stopniami swobody a T oznacza wielkość próby.
- **Statystyka Ljunga-Boxa:**

$$Q_{LB} = T(T+2) \sum_{i=1}^{P} \frac{\hat{\rho}_k^2}{T-i} \tag{21}$$

ma rozkład χ^2 z P stopniami swobody.
- Statytyka Q_{LB} ma *lepsze* własności od Q_{PB} bez względu na wielkość próby.

- Test mnożnika Lagrange'a (LM) zaproponowany przez Breuscha i Godfreya pozwala na testowanie autokorelacji zarówno pierwszego jak i wyższych rzędów.
- W pierwszym kroku szacowane są parametry modelu:

$$y_t = \beta_0 + \beta_1 x_{1,t} + \ldots + \beta_k x_{k,t} + \varepsilon_t \tag{22}$$

- W drugim kroku szacowane są parametry modelu, w którym wyjaśniany jest składnik resztowy z modelu (22). Dodatkowo, uwzględniane są opóźnienia do rzędu P włącznie:

$$e_t = \beta_0 + \underbrace{\beta_1 x_{1,t} + \beta_2 x_{2,t} + \ldots + \beta_k x_{k,t}}_{\text{zmienne objaśniające z modelu (22)}} + \underbrace{\beta_{k+1} e_{t-1} + \ldots + \beta_{k+P,t} e_{t-P}}_{\text{opóźnione reszty z modelu (22)}} + \eta_t \tag{23}$$

- Hipoteza zerowa testu LM jest równoznaczna braku autokorelacji do rzędu P włącznie:

$$\mathcal{H}_0 : \beta_{k+1} = ... = \beta_{k+P} \quad = \quad 0 \tag{24}$$

$$\mathcal{H}_1 : \exists_{l \in (1,..,P)} \beta_{k+l} \quad \neq \quad 0 \tag{25}$$

- Statystyka testowa:

$$LM = nR^2 \tag{26}$$

posiada rozkład χ^2 z P stopniami swobody (rząd weryfikowanej autokorelacji składnika losowego). Są podstawy do odrzucenia $\mathcal{H}_0$, jeżeli LM jest większa od wartości krytycznej χ^2.

- Estymator wariancji-kowariancji wektora β:

$$Var(\hat{\beta}^{OLS}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\mathbb{E}\left[\varepsilon\varepsilon^T\right]\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}. \tag{27}$$

- Niech Ω będzie niesferyczną macierzą wariancji kowariancji składnika losowego, tj.

$$Var(\hat{\beta}^{OLS}) = \left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T\Omega\mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}. \tag{28}$$

- Newey i West (1987) proponują następujący estymator wariancji-kowariancji:

$$\mathbf{X}^T\hat{\Omega}\mathbf{X} = \mathbf{X}^T\hat{\Omega}_0\mathbf{X} + \sum_{j=1}^{L}\sum_{t=j+1}^{T} w_j e_t e_{t-j}\left[x_t x_{t-j}^T + x_{t-j}x_t^T\right], \tag{29}$$

gdzie L to maksymalna liczba opóźniej, x_t to wektor obserwacji zmiennych objaśniających w momencie t, w_j to wagi dla j-tego opóźnienia, a $\hat{\Omega}_0$ jest wyznaczana następująco:

$$\mathbf{X}^T\hat{\Omega}_0\mathbf{X} = \frac{T}{T-k}\sum_{t=1}^{T} e_t^2 x_t^T x_t. \tag{30}$$

- Newey i West (1987) proponują następujące wagi

$$w_j = 1 - j/L. \tag{31}$$

- **Wybór maksymalnej liczby opóźniej** L jest kluczowy. Im mniejsze L tym mniejsza wariancja ale większe obciążenie.
  - metoda Andrewsa (1991) czy Neweya i Westa (1994),
  - metoda prób i błędów,
- **Wybór wag** w_j.
  - Możliwe wykorzystanie estymatora jądra gęstości spektralnej (*kernel spectral density*), np. Barletta czy Parzena.
- Częstą praktyką mającą na celu ograniczenie obciążenia (wynikającego z persystencji obserwacji empirycznych) jest tzw. *prewhitening* przy pomocy modelu VAR (wektorowej autoregeresji).

- **[Hill, Griffiths i Lim]:** krzywa Phillipsa dla Australii.
- **Dane:** szeregi czasowe od 1987Q1 do 2009Q3.
  - inf_t - inflacja w okresie t.
  - Δu_t - zmiana stopy bezrobocia w okresie t.
- Model:

$$inf_t = inf_t^E - \gamma \Delta u_t + \varepsilon_t \tag{32}$$

- Zakładając, że oczekiwania inflacyjne są stałe w czasie:

$$inf_t = \beta_0 + \beta_1 \Delta u_t + \varepsilon_t \tag{33}$$

gdzie $\beta_1 = -\gamma$.

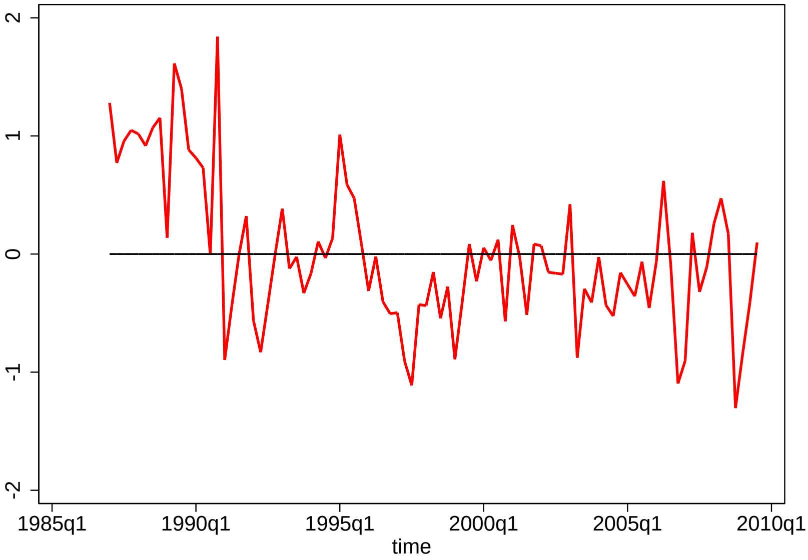
- Rozważany model:

$$inf_t = \beta_0 + \beta_1 \Delta u_t + \varepsilon_t \tag{34}$$

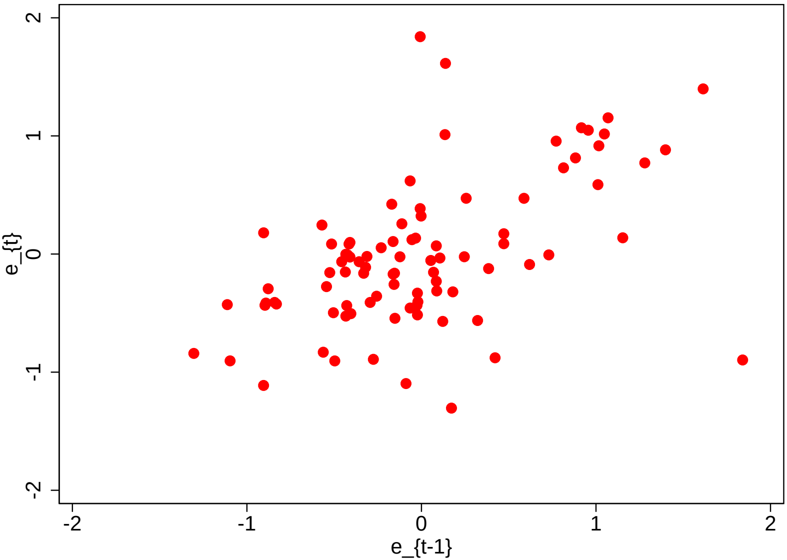
	Oszacowanie	Błąd stand.	t-Studenta	wartość p
β_0	0.777621	0.0658249	11.8135	0.0000
β_1	−0.527864	0.229405	−2.3010	0.0238

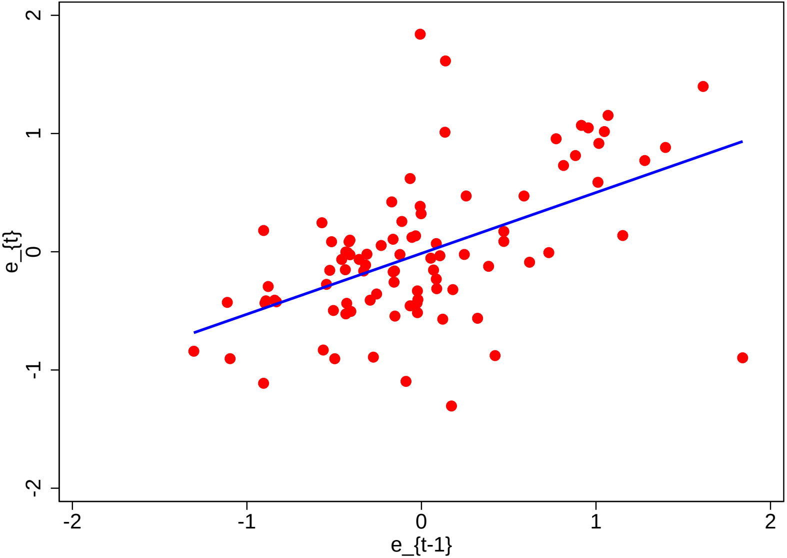
- Czy oszacowanie parametru β_1 jest statystycznie istotne?

Przykład – wykres reszt modelu względem czasu



Przykład – zależność między resztami a ich pierwszym opóźnieniem SGH




e_{t}
e_{t-1}
-2
-1
0
1
2
-2
-1
0
1
2

- Test Durbina-Watsona:
  - Statystyka testowa: 0.887
  - Wartości krytyczne ($N = 90$ i $k = 1$) przy 5% poziomie istotności:
    d_L : 1.6345
    d_U : 1.6794

- Test Durbina-Watsona:
  - Statystyka testowa: 0.887
  - Wartości krytyczne ($N = 90$ i $k = 1$) przy 5% poziomie istotności:
    d_L : 1.6345
    d_U : 1.6794

- Statystyka testu LM:

p	Wartość statystyki	p-value
1	27.592	0.000
4	36.672	0.000

Rozważany model:

$$inf_t = \beta_0 + \beta_1 \Delta u_t + \varepsilon_t \quad (35)$$

Podstawowe oszacowania oszacowania

	Oszacowanie	Błąd stand.	t-Studenta	wartość p
β_0	0.777621	0.0658249	11.8135	0.0000
β_1	−0.527864	0.229405	−2.3010	0.0238

Odporne błędy standardowe

	Oszacowanie	Błąd stand.	t-Studenta	wartość p
β_0	0.777621	0.101848	7.6351	0.0000
β_1	−0.527864	0.309225	−1.7071	0.0913

Heteroskedastyczność składnika losowego

- Heteroskedastyczność składnika losowego jest drugą formą niespełnienia założenia o sferyczności macierzy wariancji-kowariancji składnika losowego.
- Zjawisko heteroskedastyczności składnika losowego charakteryzuje przede wszystkim modele oparte o **dane przekrojowe**.
- Ogólny zapis hetereoskedastyczności składnika losowego:

$$Var(\varepsilon) = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_n^2 \end{bmatrix},$$

gdzie

$$\sigma_1^2 \neq \sigma_2^2 \neq \dots \sigma_k^2.$$

- Heteroskedastyczność składnika losowego można wiązać ze zmiennością roli nieobserowalnej heterogeniczności (różnorodności):

$$\sigma_1^2 \neq \sigma_2^2 \neq \ldots \sigma_k^2.$$

- Zasadniczo σ_i^2 może w tym przypadku zależeć od zmiennych objaśniającyh, tj. X_i:

$$\sigma_i^2 = f\left(x_{1i}, x_{2i}, \ldots, x_{ki}\right), \tag{36}$$

  gdzie $f(\cdot)$ jest pewną funkcją.
  - Kluczowe jest zdiagnozowanie tej zależności oraz próba wytłumaczenia.

Analiza reszt (e_t):

- wykresy kwadratów reszt względem zmiennych objaśniających,

Testowanie heteroskedatyczności składnika losowego

- test Godfleda i Quandta,
- Test Breuscha i Pagana,
- test White'a.

- Test Goldfelda Quandta polega na porównaniu wariancji w dwóch grupach.
- Kluczowa jest tutaj identyfikacja grup.
  - zmienne binarne,
  - porządkowanie (sortowanie) próby względem pewnej (ciągłej) zmiennej objaśniającej.
- W pierwszym kroku szacowane są parametry strukturalne modeli dla obu grup **osobno**, a następnie wyznaczane są reszty.
- W drugim kroku porównywana jest wariancja składnika losowego przy pomocy statystyki $\mathcal{F}$:

$$\mathcal{F} = \frac{SSE_1/(N_1 - K)}{SSE_2/(N_2 - K)}, \tag{37}$$

  gdzie SSE_i i N_i to suma kwadratów reszt i liczebność i-tej podpróby. **Wariancja reszt pierwszej podpróby jest większa.**
- Hipoteza zerowa postuluje brak różnic w wariancji pomiędzy grupami.
- Statystyka $\mathcal{F}$ ma rozkład F-Snedecora z $N_1 - K$ oraz $N_2 - K$ stopniami swobody.

- W teście Breuscha i Pagana zakłada się specyficzną zależność:

$$\sigma_i^2 = \sigma^2 f\left(\alpha_0 + \alpha X\right). \tag{38}$$

- Pierwszy krok: szacujemy parametry strukturalne i wyznaczamy reszty.
- Drugi krok: obliczamy kwadrat reszt w relacji do ich wariancji w całej próbie, tj. $\hat{e}_i^2/\hat{\sigma}^2$.
- Trzeci krok: regresja pomocnicza, w której zmienną objaśnianą są $\hat{e}^2/\hat{\sigma}^2$:

$$\hat{e}_i^2/\hat{\sigma}^2 = \alpha_0 + \alpha_1 x_{1i} + \ldots + \alpha_k x_{ki} + \eta_i, \tag{39}$$

gdzie $\eta_i \sim \mathcal{N}(0, \sigma_\eta^2)$. Hipoteza zerowa postuluje brak heteroskedastyczności (w rozważanej formie):

$$\mathcal{H}: \alpha_1 = \ldots = \alpha_k \quad \text{tj.} \quad \sigma_i^2 = \sigma^2. \tag{40}$$

- Statystyka testowa:

$$LM = nR^2 \tag{41}$$

posiada rozkład χ^2 z k stopniami swobody.

- Test White'a jest najogólniejszym testem pozwalającym zbadać heteroskedastyczność składnika losowego.
- W pierwszym kroku szacowane są parametry modelu:

$$y_i = \beta_0 + \beta_1 x_{1,i} + \ldots + \beta_k x_{k,i} + \varepsilon_i \tag{42}$$

- W drugim kroku, zmienną objaśnianą są kwadraty reszt z oszacowanego modelu (42). Ponadto uwzględniane są kwadraty oraz interacje zmiennych objaśniających z modelu (42), tj.:

$$\begin{aligned} \hat{e}_i^2 &= \alpha_0 + \alpha_1 x_{1,i} + \ldots + \alpha_k x_{k,i} + \beta_{k+1} x_{1,i}^2 + \ldots + \alpha_{k+k} x_{k,i}^2 + \\ &\quad + \alpha_{k+k+1} x_{1,i} x_{2,i} + \ldots + \alpha_{k+k+S} x_{k-1,i} x_{k,i} + \eta_i \end{aligned}$$

- Hipotezą zerową jest homoskedastyczność składnika losowego:

$$\mathcal{H}_0 : \quad \sigma_i^2 = \sigma^2 \qquad \mathcal{H}_1 : \quad \sigma_i^2 \neq \sigma^2 \tag{43}$$

- Statystyka testowa:

$$LM = nR^2 \tag{44}$$

posiada rozkład χ^2 z M stopniami swobody (liczba wszystkich wszystkich zmiennych objaśniających w regresji testowej, tj. $M = 2k+s$). Są podstawy do odrzucenia $\mathcal{H}_0$, jeżeli LM jest większa od wartości krytycznej χ^2.

- Test White'a jest ogólny ponieważ szczegółowa postać heteroskedastyczności jest nieznana. Tym samym, wynik tego testu może świadczyć np. o braku poprawnej specyfikacji (np. brak uwzględnienia nieliniowości).

- Macierz projekcji (*projection/hat matrix*) $\mathbf{P}$ określa zależność pomiędzy obserwacjami empirycznymi a wartościami teoretycznym:

$$\hat{\mathbf{y}} = \mathbf{P}\mathbf{y}. \tag{45}$$

- Macierz projekcji w przypadku estymatora MNK:

$$\mathbf{P} = \mathbf{X}\left(\mathbf{X}^T\mathbf{X}\right)^{-1}\mathbf{X}^T. \tag{46}$$

- Wybrane własność macierzy projekcji:
  - Symetryczność, tj. $\mathbf{P} = \mathbf{P}^T$.
  - Idempotentność, tj. $\mathbf{P}\mathbf{P} = \mathbf{P}$.

- Najczęstsze estymatory macierzy wariancji-kowariancji (Ω) uwzględniające heteroskedaczność HC (*heteroskedasticity-consistent*):

$$\begin{aligned}
&\text{homoskedatyczność:} & \omega_i &= \hat{\sigma}^2_{\varepsilon}(\text{const}),\\
&HC_0 \text{ (White)} & \omega_i &= \hat{e}_i^2,\\
&HC_1 & \omega_i &= \tfrac{N}{N-k}\hat{e}_i^2,\\
&HC_2 & \omega_i &= \tfrac{\hat{e}_i^2}{1-p_i},\\
&HC_3 & \omega_i &= \tfrac{\hat{e}_i^2}{(1-p_i)^2},\\
&HC_4 & \omega_i &= \tfrac{\hat{e}_i^2}{(1-p_i)^{\delta_i}}.
\end{aligned}$$

gdzie p_i to diagonalny element macierzy projekcji P, a $\delta_i = min\,(4, p_i/\bar{p})$.

Estymator GLS (UMNK)

- **Kluczowy problem:**
$$Var(\varepsilon) = \sigma^2\Omega \neq \sigma^2 I. \tag{47}$$
- Konsekwencje:
  - Obciążenie estymatora wariancji-kowariancji $Var(\beta^{OLS})$.
  - Niewiarygodność błędów standardowych, statystyk testu t-studenta, statystyki testu liniowych resktykcji.
- Rozwiązania:
  - Alternatywna postać funkcyjna
    - **Szeregi czasowe:** modele autoregresyjne.
    - **Dane przekrojowe:** transformacja logarytmiczna, uwzględnienie pominiętych zmiennych, uwzględnienie nieliniowości, itp.
  - Odporny estymator wariancji-kowariancji.
  - **Estymator GLS (UMNK)**.

- Model regresji liniowej:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{48}$$

gdzie

- $\mathbf{y}$ – wektor zmiennej objaśnianej,
- $\mathbf{X}$ – macierz zmiennych objaśniających,
- β – wektor parametrów strukturalnych,
- ε – składnik losowy.

- Model regresji liniowej:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{48}$$

gdzie
  - $\mathbf{y}$ – wektor zmiennej objaśnianej,
  - $\mathbf{X}$ – macierz zmiennych objaśniających,
  - β – wektor parametrów strukturalnych,
  - ε – składnik losowy.

- **Estymator GLS (UMNK) – założenia:**
  1. $rz(\mathbf{X}) = k + 1 \leq n$
  2. $\mathbb{E}(\mathbf{X}\varepsilon) = 0$
  3. $\mathbb{E}(\varepsilon) = 0$
  4. $Var(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon^T) = \sigma^2\Omega$
  5. $\varepsilon_i \sim N(0, \sigma^2)$

- Model regresji liniowej:

$$\mathbf{y} = \mathbf{X}\beta + \varepsilon, \tag{48}$$

gdzie

  - $\mathbf{y}$ – wektor zmiennej objaśnianej,
  - $\mathbf{X}$ – macierz zmiennych objaśniających,
  - β – wektor parametrów strukturalnych,
  - ε – składnik losowy.

- **Estymator GLS (UMNK) – założenia:**

  1. $rz(\mathbf{X}) = k + 1 \leq n$
  2. $\mathbb{E}(\mathbf{X}\varepsilon) = 0$
  3. $\mathbb{E}(\varepsilon) = 0$
  4. $Var(\varepsilon) = \mathbb{E}(\varepsilon\varepsilon^T) = \sigma^2\Omega$
  5. $\varepsilon_i \sim N(0, \sigma^2)$

- **Estymator Uogólnionej Metody Najmniejszych Kwadratów UMNK (GLS – generalized least squares):**

$$\hat{\beta}^{GLS} = \left(\mathbf{X}^T\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{y} \tag{49}$$

- **Estymator GLS (UMNK):**

$$\hat{\beta}^{GLS} = \left(\mathbf{X}^T \Omega^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \Omega^{-1} \mathbf{y} \tag{50}$$

- Geneza:

$$\hat{\beta}^{GLS} = arg \min_{\beta} \left(\mathbf{y} - \mathbf{X}\beta\right)^T \Omega^{-1} \left(\mathbf{y} - \mathbf{X}\beta\right). \tag{51}$$

- Estymator $\hat{\beta}^{GLS}$ jest zgodny, nieobciążony i najefektywniejszy przy spełnieniu założeń z poprzedniego slajdu.
  - Jeżeli $\Omega = I$ to wtedy $\hat{\beta}^{GLS} = \hat{\beta}^{OLS}$.
- Kluczowa jest znajomość Ω, a więc macierzy wariancji-kowariancji składnika losowego.

- Kluczowe założeniem o macierzy wariancji-kowariancji składnika losowego jest fakt, że Ω jest symetryczną pozytywnie zdefiniowaną macierzą. Wtedy,

$$\Omega = C\Lambda C^T, \tag{52}$$

gdzie C to macierz wektorów własnych a Λ to macierz diagonalna, której elementami są wartości własne macierzy Ω.

- Zdefiniujmy macierz P:

$$P^T = C\Lambda^{-1/2}, \tag{53}$$

a więc w szczególności:

$$\Omega^{-1} = P^T P. \tag{54}$$

- Wprowadźmy transformację elementów polegającą na lewostronnym pomnożeniu przez macierz P, tj. $\mathbf{y}_* = P\mathbf{y}$, wtedy zapis:

$$P\mathbf{y} = P\mathbf{X}\beta + P\varepsilon, \tag{55}$$

jest tożsamy zapisowi:

$$\mathbf{y}_* = \mathbf{X}_*\beta + \varepsilon_*. \tag{56}$$

- Natomiast wariancja składnika losowego jest sferyczna (pamiętając, że $\Omega^{-1} = P^T P$):

$$Var(\varepsilon_*) = P\Omega P^T = \sigma^2 I. \tag{57}$$

- Znając powyższą transformację, zastosujemy teraz estymator MNK dla przetransformowanych obserwacji, ponieważ $Var(\varepsilon_*)$ jest sferyczna:

$$\begin{aligned} \hat{\beta}_*^{OLS} &= \left(\mathbf{X}_*^T\mathbf{X}_*\right)^{-1}\mathbf{X}_*^T\mathbf{y}_* \\ &= \left(\mathbf{X}^T P^T P\mathbf{X}\right)^{-1}\mathbf{X}^T P^T P\mathbf{y} \\ &= \left(\mathbf{X}^T\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{y} = \hat{\beta}^{GLS}. \end{aligned}$$

- **Kluczowym praktycznym wyzwaniem jest znajomość macierzy wariancji-kowariancji składnika losowego $\sigma^2\Omega$.**
- W empirycznych aplikacjach wykorzystywany jest estymator FGLS (*feasible generalized squares*). Ogólna idea:
  1. Szacujmy parametry modelu MNK. Dzięki wektorowi $\hat{\beta}^{OLS}$ uzyskujemy reszty.
  2. Na podstawie (i) **uzyskanych reszt** oraz (ii) **założenia o strukturze niesferyczności składnika losowego (np. heteroskedastyczność lub autokorelacja)** wyznaczamy szacunek macierzy $\hat{\Omega}$.
  3. Mając macierz $\hat{\Omega}$ możemy wyznaczyć estymator (F)GLS, tj. :

$$\hat{\beta}^{FGLS} = \left(\mathbf{X}^T \hat{\Omega}^{-1} \mathbf{X}\right)^{-1} \mathbf{X}^T \hat{\Omega}^{-1} \mathbf{y}, \tag{58}$$

  ...lub zastosować odpowiednią tranformację danych.
- Rozszerzeniem estymatora FGLS jest iterowany estymator FGLS.
  - Transformacja wynikająca z estymatora GLS jest stosowana wielokrotnie do momentu spełnienia pewnego kryterium (np. brak autokorelacji).

Metoda Cochrane'a -Orcutta

- Metoda Cochrane'a -Orcutta polega na wykorzystaniu estymatora FGLS w przypadku autokorelacji składnika losowego pierwszego rzędu.
- Model regresji liniowej dla szeregów czasowych:

$$y_t = X_t\beta + \varepsilon_t \tag{59}$$

- **Autokorelacja składnika losowego pierwszego rzędu AR(1)**:

$$\varepsilon_t = \rho\varepsilon_{t-1} + \eta_t \tag{60}$$

gdzie $\eta_t \sim \mathcal{N}(0, \sigma^2_\eta)$, $\mathbb{E}(\eta) = 0$ oraz $Var(\eta)^2 = I\sigma^2_\eta$. Zakładamy również, że $|\rho| < 1$.

- Przy takich założeniach, macierz wariancji-kowariancji składnika losowego nie jest diagonalna:

$$Var(\varepsilon_t) = \frac{\sigma^2_\eta}{1-\rho}\begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{n-1} \\ \rho & 1 & \rho & \dots & \rho^{n-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{n-3} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \rho^{n-1} & \rho^{n-2} & \rho^{n-3} & \dots & 1 \end{bmatrix} = \sigma^2_\eta\Omega.$$

- Macierz odwrotna do macierzy Ω:

$$\Omega^{-1} = \begin{bmatrix} 1 & -\rho & 0 & \dots & 0 \\ -\rho & 1+\rho^2 & -\rho & \dots & 0 \\ 0 & -\rho & 1+\rho^2 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1+\rho^2 \end{bmatrix}$$

- Korzystając z dekompozycji macierzy $\Omega^{-1} = P^T P$, otrzymujemy następującą macierz P:

$$P = \begin{bmatrix} \sqrt{1-\rho^2} & 0 & 0 & \dots & 0 \\ -\rho & 1 & 0 & \dots & 0 \\ 0 & -\rho & 1 & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & 1 \end{bmatrix}.$$

- Powyższy zapis sugeruje następującą transformację (dla $t > 1$):

$$y_t - \hat{\rho} y_{t-1} = X_t \beta - \hat{\rho} X_{t-1} \beta + \eta_t. \tag{61}$$

- Transformacja ta zmniejsza liczebność obserwacji wykorzystywanych w próbie. Prozszerzeniem metody Cochrane'a-Orcutta jest metoda Praisa-Winstena, z godnie z którą proponowana jest następująca tranformacja dla $t = 1$ (pierwszej obserwacji):

$$\sqrt{1-\hat{\rho}^2} y_1 = \sqrt{1-\hat{\rho}^2} X_1 \beta + \eta_t. \tag{62}$$

- Oszacowanie współczynnika autokorelacji pierwszego rzędu ($\hat{\rho}$) można wyznaczyć na podstawie reszt z modelu, którego parametry szacowano MNK.
- Powyższa transformacja jest nazywana quasi różnicowaniem *quasi-differencing*. Przypominając operator opóźnień $L y_t = y_{t-1}$, dla $t > 1$ mamy:

$$y_t \left(1 - \hat{\rho} L\right) = \left(1 - \hat{\rho} L\right) X_t \beta + \eta_t, \tag{63}$$

a operator różnicowania jest definiowany jako $\Delta y_t = (1 - L) y_t$.

Ważona MNK i FGLS

- **Ważona Metoda Najmniejszych Kwadratów WLS (*weighted least squares*)** jest szczególnym przypadkiem estymatora GLS, który jest wykorzystywany w przypadku **heteroskedastyczności składnika losowego**.
- W przypadku heteroskedastyczności składnika losowego, macierz wariancji-kowariancji jest następująca:

$$Var(\varepsilon) = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \sigma_n^2 \end{bmatrix},$$

a samą wariancję składnika losowego możemy zapisać jako:

$$\sigma_i^2 = \omega_i \sigma^2 \tag{64}$$

wtedy macierz wariancji-kowariancji można zapisać:

$$Var(\varepsilon) = \sigma^2 \Omega = \sigma^2 \begin{bmatrix} \omega_1 & 0 & \dots & 0 \\ 0 & \omega_2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \omega_n \end{bmatrix}.$$

- Stosując estymator GLS

$$\hat{\beta}^{GLS} = \left(\mathbf{X}^T\Omega^{-1}\mathbf{X}\right)^{-1}\mathbf{X}^T\Omega^{-1}\mathbf{y} \tag{65}$$

możemy zuważyć, że nieznana macierz Ω^{-1} to macierz zawierająca elementy ω_i na przekątnej (*weights*):

$$\Omega^{-1} = \begin{bmatrix} \frac{1}{\omega_1} & 0 & \dots & 0 \\ 0 & \frac{1}{\omega_2} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{1}{\omega_n} \end{bmatrix}.$$

- Kluczowe jest zdefiniowanie macierz P, a więc macierzy transformacji GLS. Pamiętając, że $\Omega^{-1} = P^TP$ łatwo pokazać, że taką macierzą dla estymatora WLS można zapisać następująco:

$$P = \begin{bmatrix} \frac{1}{\sqrt{\omega_1}} & 0 & \dots & 0 \\ 0 & \frac{1}{\sqrt{\omega_2}} & \ddots & \vdots \\ \vdots & \ddots & \ddots & \vdots \\ 0 & \dots & 0 & \frac{1}{\sqrt{\omega_n}} \end{bmatrix}.$$

- Powyższe własności (macierz P) implikują następującą transformację danych:

$$y_{i*} = \frac{y_i}{\sqrt{\omega_i}}, \tag{66}$$

a następnie możliwość zastosowania estymatora OLS dla następującego modelu:

$$y_* = X_* \beta + \eta_i. \tag{67}$$

- **Intuicja:** w estymacji ważoną MNK niższe wagi (znaczenie) będą *otrzymywać* jednostki charakteryzujące się wyższą wariancją składnika losowego.

- Kluczowym założeniem estymatora WLS jest fakt, że znamy wagi wynikające z poniższej zależności
$$\sigma_i^2 = \omega_i \sigma^2. \tag{68}$$
- Jednak w aplikacjach empirycznych zazwyczaj nie znamy wag, tj. $\sqrt{\omega_i}$. Dlatego, kluczowe jest zrozumienie problemu heteroskedastyczności odpowienio wyznaczyć wagi.
- Popularne strategie wyznaczania wag:
  1. Założenie o zależności pomiędzy σ_i a x_{ij}, czyli pewną zmienną objaśniającą. Np.
  $$\sigma_i^2 = \sigma^2 x_{ij} \tag{69}$$
  wtedy wagami są naturalnie $1/\sqrt{x_{ij}}$.
  2. Uwzględnienie różnic pomiędzy grupami. Procedura jest następująca:
     - ★ Identyfikujemy K grup. Każda jednostka/obserwacja powinna być przyporządkowana dokładnie do jednej grupy.
     - ★ Dla każdej grupy szacujemy parametry strukturalne (MNK) a następnie wyznaczamy wariancję wariancję składnika losowego, tj. $\hat{\sigma}_i^2$ będzie taka sama dla wszystkich jednostek przynależących do pewnej grupy.
     - ★ Wagami w estymatorze WLS będą zatem $1/\hat{\sigma}_i$.

Estymator FGLS w przypadku heteroskedastyczności

SGH

- W przypadku heteroskedastyczności można rozważyć bardziej *elastyczną* zależność między wariancją składnika losowego σ_i^2 a zmiennymi objaśniającymi (x). Wtedy możliwe jest zastosowanie estymatora FGLS. Ogólna procedura jest następująca:
  1. Oszacuj parametry rozważanego modelu MNK.
  2. Na podstawie oszacowań z poprzedniego punktu wyznacza kwadraty reszt, $\hat{e}_i^2$.
  3. Rozważ model pomocniczy, w którym zmienną obajśnianą są $\hat{e}_i^2$ (lub logarytm naturalny $\ln \hat{e}_i^2$) a zmiennymi objaśniającymi x. Oszacuj parametry takiego modelu MNK.
  4. Wyznacz wartości teoretyczne na podstawie oszacowań z poprzedniego punktu, tj. $\hat{\hat{e}}_i^2$.
  5. Wykorzystaj $\sqrt{\hat{\hat{e}}_i^2}$ jako wagi w estymatorze WLS, a więc przeprowadź transformację wykorzystując (dzieląc przez) $\hat{\hat{e}}_i$.
- Zazwyczaj zależność między σ_i^2 a x_i ma charakter multiplikatywny. Przykładowe założenia o zależności między σ_i^2 a x_i
  - [Przykład #1] założenie $\sigma_i^2 = \sigma^2 x_{1i}^{\gamma_1} x_{2i}^{\gamma_2}$ będzie implikować następująca postać regresji pomocnicznej:
    $$\ln \hat{e}_i^2 = \gamma_0 + \gamma_1 \ln x_{1i} + \gamma_1 \ln x_{2i} + \eta_i. \tag{70}$$
  - [Przykład #2] założenie $\sigma_i^2 = \sigma^2 \exp(\gamma_1 + x_{1i}^{\gamma_2 x_{2i}})$ będzie implikować następująca postać regresji pomocnicznej:
    $$\ln \hat{e}_i^2 = \gamma_0 + \gamma_1 x_{1i} + \gamma_1 x_{2i} + \eta_i. \tag{71}$$
  - [Przykład #3] wykorzystać $\hat{y}$ i $\hat{y}^2$ zamiast x.

- Testowanie liniowych restrykcji ma sens jeżeli korzystamy z tych samych wag.
- Co jeżeli oszacowania WLS i OLS są statystycznie różne, tj. charakteryzują się przeciwnymi znakami?
- Czy R^2 ma interpretację?