Matematyka stosowana

Ekonometria

Piotr Władysław Jaworski
P.Jaworski@mimuw.edu.pl

Uniwersytet Warszawski, 2011

Streszczenie. Celem wykładu jest przedstawienie teoretycznych podstaw współczesnej ekonometrii

Wersja internetowa wykładu:
http://mst.mimuw.edu.pl/lecture.php?lecture=ekn

(może zawierać dodatkowe materiały)

Niniejsze materiały są dostępne na licencji Creative Commons 3.0 Polska:
Uznanie autorstwa — Użycie niekomercyjne — Bez utworów zależnych.

Copyright © P.Jaworski, Uniwersytet Warszawski, Wydział Matematyki, Informatyki i Mechaniki, 2011. Niniejszy plik PDF został utworzony 5 czerwca 2011.

Projekt współfinansowany przez Unię Europejską w ramach Europejskiego Funduszu Społecznego.

Skład w systemie LaTeX, z wykorzystaniem m.in. pakietów beamer oraz listings. Szablony podręcznika i prezentacji: Piotr Krzyżanowski; koncept: Robert Dąbrowski.

Spis treści

1. Wstęp - Co to jest ekonometria? . . . 5
- 1.1. Informacje wstępne . . . 5
- 1.2. Etapy modelowania . . . 6
- 1.3. Przykłady . . . 6
- 1.4. Klasyfikacja modeli ekonometrycznych . . . 8

2. Metoda najmniejszych kwadratów (MNK) . . . 9
- 2.1. Wprowadzenie . . . 9
- 2.2. Odrobina algebry liniowej . . . 12

3. MNK w terminach statystyki opisowej . . . 14
- 3.1. Notacja statystyki opisowej . . . 14
- 3.2. MNK z wyrazem wolnym . . . 15
- 3.3. Przypadek $m = 2$ i $X_2 = e$ . . . 18

4. Klasyczny model regresji . . . 19
- 4.1. Notacja macierzowa dla zmiennych losowych . . . 19
- 4.2. Warunkowa wartość oczekiwana . . . 20
- 4.3. Założenia klasycznego modelu regresji . . . 21
- 4.4. Estymacja parametrów modelu metodą MNK . . . 22

5. Klasyczny model regresji z gaussowskim składnikiem losowym . . . 27
- 5.1. Testowanie pojedynczego parametru strukturalnego β_k . . . 27
- 5.2. Testowanie hipotezy liniowości . . . 29

6. Modele produkcji - funkcja Cobba-Douglasa . . . 35
- 6.1. Funkcja Cobba-Douglasa . . . 35
  - 6.1.1. Wprowadzenie . . . 35
  - 6.1.2. Efekt skali . . . 36
  - 6.1.3. Koszty produkcji . . . 36
- 6.2. Przykład Nerlove'a . . . 37
  - 6.2.1. Charakterystyka danych . . . 37
  - 6.2.2. Konstrukcja modelu . . . 38
  - 6.2.3. Estymacja parametrów modelu 6.2 . . . 38
  - 6.2.4. Estymacja parametrów modelu 6.3 . . . 38
  - 6.2.5. Test jednorodności modelu . . . 39
  - 6.2.6. Test braku efektów skali dla modelu ograniczonego 6.3 . . . 39

7. Modele nieliniowe . . . 40
- 7.1. Zadanie aproksymacyjne . . . 40
- 7.2. Założenia modelu i estymacja parametrów . . . 41
- 7.3. Przykłady . . . 42
  - 7.3.1. Funkcja Törnquista I typu . . . 43
  - 7.3.2. Funkcja Törnquista II typu . . . 44
  - 7.3.3. Funkcja Törnquista III typu . . . 45
  - 7.3.4. Funkcja logistyczna . . . 46

8. Metody asymptotyczne . . . 47
- 8.1. Zbieżność zmiennych losowych . . . 47
- 8.2. Estymatory jako ciągi zmiennych losowych . . . 50

8.3. Stacjonarność i ergodyczność procesów stochastycznych . . . 51
8.3.1. Definicje i podstawowe własności . . . 51
8.3.2. Przykłady . . . 52
8.4. Martyngały i przyrosty martyngałowe . . . 54

9. Teoria dużej próbki . . . 56

9.1. Założenia modelu . . . 56
9.2. Asymptotyka estymatorów MNK . . . 57

10. Teoria dużej próbki cd . . . 61

10.1. Testy asymptotyczne . . . 61
10.1.1. Testowanie pojedynczego parametru strukturalnego b_k . . . 61
10.1.2. Testowanie hipotezy liniowości . . . 62
10.1.3. Testowanie nieliniowych zależności między parametrami modelu . . . 63
10.1.4. Testowanie warunkowej homoskedastyczności – test White'a . . . 64

11. Testowanie autokorelacji składnika losowego i składnika resztowego . . . 66

11.1. Autokorelacja składnika losowego . . . 66
11.2. Autokorelacja składnika resztowego . . . 69

12. Hipoteza efektywnego rynku - ekonometria racjonalnych oczekiwań . . . 71

12.1. Przykład E.Fama - konstrukcja modelu . . . 71
12.2. Hipoteza efektywnego rynku . . . 72
12.3. Analiza danych empirycznych . . . 73
12.3.1. Test na autokorelację realnych stóp zwrotu . . . 73
12.3.2. Test predykcji stopy inflacji π w oparciu o nominalną stopę zwrotu R . . . 74
12.3.3. Dyskusja wyników . . . 75

13. Regresja względem czasu. . . . 76

13.1. Model tendencji rozwojowej z liniowym trendem . . . 76
13.1.1. Założenia modelu . . . 76
13.1.2. Estymacja parametrów modelu . . . 76
13.1.3. Testowanie parametrów strukturalnych . . . 78

14. Liniowe szeregi czasowe . . . 80

14.1. Szeregi czasowe stacjonarne rzędu 2 . . . 80
14.2. Sploty vel filtry . . . 82
14.3. Funkcje tworzące . . . 86
14.4. Operator przesunięcia . . . 88
14.5. Przykłady . . . 88
14.6. Procesy o przyrostach stacjonarnych . . . 89
14.7. Ułamkowy ruch Browna . . . 90

15. Nieliniowe szeregi czasowe . . . 92

15.1. Wstęp. . . . 92
15.2. Ogólne własności modelu GARCH(1,1). . . . 92
15.3. Ograniczenia na parametry modelu GARCH(1,1). . . . 94
15.4. Stacjonarność modeli GARCH. . . . 94

Literatura . . . 96

Bibliografia . . . 96

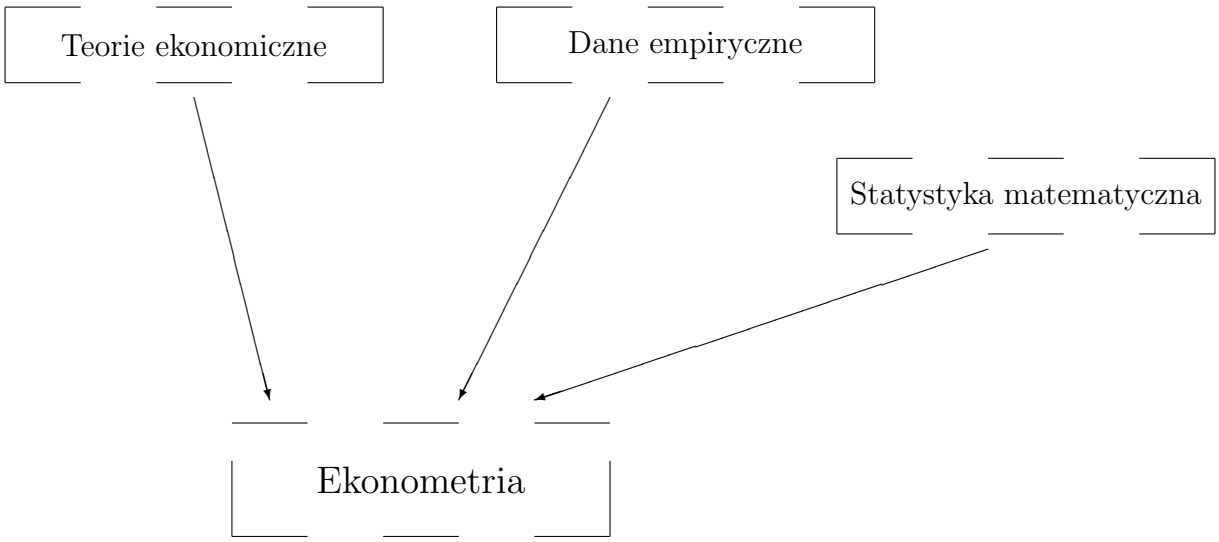
Literatura . . . 97

1. Wstęp - Co to jest ekonometria?

Podstawowe metody i cele. Przykłady modeli ekonometrycznych. Ogólna klasyfikacja modeli ekonometrycznych. (1 wykład)

1.1. Informacje wstępne

W skrócie można powiedzieć, że ekonometria to zestawienie danych empirycznych z teoriami ekonomicznymi przy zastosowaniu statystyki matematycznej.



Rysunek 1.1. Ekonometria - schemat powiązań.

Podstawowe cele ekonometrii to:
1. Analiza danych empirycznych i prognozowanie na ich podstawie;
2. Weryfikacja i kalibrowanie teorii ekonomicznych.

Kluczowym obiektem w ekonometrii jest tzw. *model ekonometryczny*. Zapisujemy go w postaci

$$Y_t = F(t, X_t, \varepsilon_t),$$

gdzie $t \in \mathbb{N}$ zwykle oznacza czas – kolejny moment lub kolejny przedział czasowy (dzień, miesiąc, rok ...), ale może też oznaczać numer porządkowy obserwacji (np. firmy, której dotyczą dane czy województwa).
$X_t \in \mathbb{R}^k$ to wektor zmiennych objaśniających,
$Y_t \in \mathbb{R}^m$ to wektor zmiennych objaśnianych,
F nazywa się postacią analityczną modelu, jest to funkcja o wartościach w $\mathbb{R}^m$;
a ε_t nazywa się składnikiem losowym.

1.2. Etapy modelowania

Przedstawimy teraz uproszczony schemat konstrukcji modelu ekonometrycznego. Możemy wyróżnić trzy operacje:

Zbieramy dane historyczne (empiryczne) x_t, y_t.

$\Downarrow$ Aproksymacja

Konstruujemy model $y_t = f(t, x_t) + \xi_t$, gdzie ξ_t - błąd przybliżenia.

$\Downarrow$ Estymacja

Konstruujemy model stochastyczny $Y_t = f(t, X_t) + \varepsilon_t$, gdzie X_t i Y_t to zmienne losowe, których realizacją są nasze obserwacje x_t i y_t, a ε_t to składnik losowy (też zmienne losowe).

$\Downarrow$ Ekstrapolacja

Zakładamy, że w przyszłości X_t i Y_t będą związane tą samą zależnością jak dotychczas.

Proszę zwrócić uwagę, że dwie pierwsze operacje aproksymację i estymację możemy wykonać dowolnie dokładnie. Natomiast o ekstrapolacji zawsze "matematyk teoretyk" będzie mógł powiedzieć, że to "wróżenie z fusów".

1.3. Przykłady

1. Model konsumpcji
Przez Y_t oznaczamy całkowity popyt konsumpcyjny w miesiącu t, a przez X_t dochody gospodarstw domowych w tym okresie. Przyjmujemy, że

$$Y_t = \alpha_0 + \alpha_1 X_t + \varepsilon_t,$$

gdzie α_0 wydatki stałe, α_1 część dochodów przeznaczona na konsumpcję, a ε_t składnik losowy. Zauważmy, że składnik losowy "zawiera w sobie" wszystkie czynniki nie uwzględnione w sposób jawny w modelu.

Uwagi:
W modelu zakładamy, że α_0 i α_1 są stałe, a w rzeczywistości są one tylko wolno-zmienne. Istotną wadą powyższego modelu jest nieuwzględnienie oszczędności.

2. Model oszczędności
Przez Y_t oznaczamy stan oszczędności na koniec miesiąca t, a przez X_t dochody gospodarstw domowych w tym miesiącu. Przyjmujemy, że

$$Y_t = Y_{t-1} - \beta_0 + \beta_1 X_t - \beta_2 Y_{t-1} + \varepsilon_t,$$

gdzie β_0 wydatki stałe, β_1 część dochodów przeznaczona na oszczędności, β_2 część oszczędności przeznaczona na konsumpcję, a ε_t składnik losowy.

Uwagi:
Zauważmy, że w powyższym modelu opóźniona zmienna objaśniana jest zmienną objaśniająca.

Model 1 i 2 można połączyć i otrzymać model dwurównaniowy.

3. Model konsumpcji z uwzględnieniem oszczędności
Przez $Y_{1,t}$ oznaczamy całkowity popyt konsumpcyjny w miesiącu t, przez $Y_{2,t}$ oznaczamy stan oszczędności, a przez X_t dochody gospodarstw domowych w tym okresie. Przyjmujemy, że

$$Y_{1,t} = \alpha_0 + \alpha_1 X_t + \alpha_2 Y_{2,t-1} + \varepsilon_{1,t},$$
$$Y_{2,t} = Y_{2,t-1} - \beta_0 + \beta_1 X_t - \beta_2 Y_{2,t-1} + \varepsilon_{2,t},$$

gdzie

$$\beta_0 = \alpha_0, \quad \beta_1 + \alpha_1 = 1, \quad \beta_2 = \alpha_2.$$

Uwagi:
Na powyższym przykładzie widzimy, jak z prostszych modeli można konstruować bardziej skomplikowane.
Pytanie: Czy w ten sposób uzyskujemy lepszy opis badanego zjawiska?
Okazuje się, że nie zawsze. Wyznaczanie wartości parametrów dla bardziej złożonego modelu, jest zwykle bardziej skomplikowane i mniej dokładne. W efekcie złożony model, który jest teoretycznie lepszy, w praktyce już takim być nie musi.

4. Model popytu dla dóbr konsumpcyjnych
Przez Y_t oznaczamy popyt dla wybranego dobra konsumpcyjnego, przez $X_{1,t}$ jego cenę, a przez $X_{2,t}$ dochody nabywcy. Przyjmujemy, że

$$Y_t = cX_{1,t}^{\alpha} X_{2,t}^{\beta} e^{\varepsilon_t}, \quad c > 0, \;\; \alpha < 0, \;\; \beta > 0.$$

Uwagi:
Jest to przykład modelu nieliniowego, który można zlinearyzować za pomocą logarytmowania.

$$\ln Y_t = \ln c + \alpha \ln X_{1,t} + \beta \ln X_{2,t} + \varepsilon_t.$$

5. Model stochastyczny kursu walutowego
Niech Y_t oznacza kurs 1 USD w EUR w dniu t. Przyjmujemy

$$Y_t = Y_{t-1} e^{\varepsilon_t}, \quad E(\varepsilon) = 0.$$

Po zlogarytmowaniu otrzymujemy model błądzenia przypadkowego

$$\ln Y_t = \ln Y_{t-1} + \varepsilon_t.$$

6. Model wydajności pracy
Niech Y_t oznacza wydajność pracy w PLN na 1 pracownika, a X_t techniczne uzbrojenie miejsca pracy też w PLN na 1 pracownika. Przyjmujemy

$$Y_t = \gamma X_t^{\alpha} e^{\delta t + \varepsilon_t}, \;\; \gamma, \alpha > 0.$$

Po zlogarytmowaniu otrzymujemy

$$\ln Y_t = \ln \gamma + \alpha \ln X_t + \delta t + \varepsilon_t.$$

Uwagi:
Współczynnik δ mierzy skalę postępu techniczno-organizacyjnego.

1.4. Klasyfikacja modeli ekonometrycznych

1. Klasyfikacja ze względu na dynamikę:
a. Modele statyczne (jednokresowe) charakteryzujące się brakiem zależności od czasu, tzn. F nie zależy od czasu i wśród zmiennych objaśniających nie ma opóźnionych zmiennych objaśnianych. Przykłady 1 i 4.
b. Modele dynamiczne – zależne od czasu lub od opóźnionych zmiennych objaśnianych. Przykłady 2, 3, 5 i 6.
W klasie modeli dynamicznych wyróżniamy modele autoregresyjne w których zależność od czasu wiąże się tylko z występowaniem zmiennych opóźnionych. Przykłady 2, 3 i 5.

2. Klasyfikacja ze względu na postać analityczną modelu:
a. Modele liniowe, postać analityczna jest zadana przez funkcję liniową. Przykłady 1, 2 i 3.
b. Modele nieliniowe, postać analityczna nie jest zadana przez funkcję liniową.
W klasie modeli nieliniowych wyróżniamy modele multiplikatywne, które można zlinearyzować poprzez zlogarytmowanie. Przykłady 4, 5 i 6.

2. Klasyfikacja ze względu na wymiar zmiennej objaśnianej:
a. Modele jednorównaniowe. Przykłady 1, 2, 4, 5 i 6.
b. Modele wielorównaniowe. Przykład 3.

Klasyfikacja ze względu na dynamikę wiąże się z planowanym wykorzystaniem modelu. Do prognozowania potrzebne są modele dynamiczne. Natomiast do badania wpływu zmian konkretnych czynników wystarczy model statyczny.

Klasyfikacja ze względu na postać analityczną modelu i wymiar określa złożoność kalibracji modelu. Jeśli model jest liniowy i jednorównaniowy to istnieją ogólne, w miarę proste, algorytmy (które omówimy na dalszych wykładach) pozwalające sprawnie wyestymować parametry modelu. W przeciwnym wypadku algorytm zależy od konkretnego przypadku i zwykle jest dużo bardziej skomplikowany.

2. Metoda najmniejszych kwadratów (MNK)

Metoda najmniejszych kwadratów (MNK). Sformułowanie zadania. Wyznaczanie optymalnych wartości parametrów. Oszacowanie błędu przybliżenia. Algebraiczne własności MNK. (1 wykład)

2.1. Wprowadzenie

Zadanie.
Dane jest $m+1$ ciągów n-elementowych o wyrazach rzeczywistych:

$$\begin{aligned} Y &= (Y_t)_{t=1,\dots,n} \\ X_1 &= (X_{t,1})_{t=1,\dots,n} \\ X_2 &= (X_{t,2})_{t=1,\dots,n} \\ &\dots \\ X_m &= (X_{t,m})_{t=1,\dots,n}. \end{aligned}$$

Wyznaczyć współczynniki $b_1, \dots, b_m \in \mathbb{R}$, które minimalizują błąd przybliżenia Y przez kombinację liniową $\widehat{Y}$

$$\widehat{Y}_t = b_1 X_{t,1} + \dots + b_m X_{t,m}.$$

Czyli mamy rozwiązać zadanie optymalizacyjne

$$\sum_{t=1}^{n} \xi_t^2 \longrightarrow min, \quad \text{gdzie} \quad \xi_t = Y_t - \widehat{Y}_t.$$

W zastosowaniach ekonometrycznych $\widehat{Y}$ nazywa się zmienną modelową w odróżnieniu od zmiennej empirycznej Y.

W dalszym ciągu będziemy stosować zapis macierzowy:
Y będzie zapisywać jako wektor kolumnowy czyli macierz $n \times 1$

$$Y = \begin{pmatrix} Y_1 \\ \dots \\ Y_n \end{pmatrix},$$

X jako macierz $n \times m$, której kolumnami są X_i

$$X = \begin{pmatrix} X_{1,1} & X_{1,2} & \dots & X_{1,m} \\ X_{2,1} & X_{2,2} & \dots & X_{2,m} \\ \dots & \dots & \dots & \dots \\ X_{n,1} & X_{n,2} & \dots & X_{n,m} \end{pmatrix},$$

szukane parametry b_i jako wektor kolumnowy $m \times 1$

$$B = \begin{pmatrix} b_1 \\ \dots \\ b_m \end{pmatrix},$$

podobnie składnik resztowy (residualny) ξ jako wektor kolumnowy $n \times 1$

$$\xi = \begin{pmatrix} \xi_1 \\ \dots \\ \xi_n \end{pmatrix}.$$

Wówczas możemy zapisać

$$\widehat{Y} = XB, \quad \xi = Y - \widehat{Y} = Y - XB.$$

Suma kwadratów reszt (SKR) wynosi

$$\sum_{t=1}^{n} \xi_t^2 = \|\xi\|^2 = \xi^T \xi = (Y^T - B^T X^T)(Y - XB) = SKR(b_1, \dots, b_m).$$

Zauważmy, że funkcja

$$SKR : \mathbb{R}^m \longrightarrow \mathbb{R}$$

jest funkcją kwadratową o wartościach nieujemnych, a zatem osiąga swoje minimum.

Twierdzenie 2.1. *Jeżeli ciągi X_1, ... , X_m są liniowo niezależne to SKR przyjmuje minimum dokładnie w jednym punkcie*

$$B_{min} = (X^T X)^{-1} X^T Y. \tag{2.1}$$

Minimum to wynosi

$$SKR_{min} = SKR(B_{min}) = Y^T Y - Y^T X (X^T X)^{-1} X^T Y.$$

Dowód.
Krok 1. Najpierw pokażemy, że macierz $X^T X$ jest odwracalna a zatem wzór 2.1 jest poprawny.

$$(X^T X)_{i,j} = \sum_{t=1}^{n} X_{t,i} X_{t,j} = X_i^T X_j.$$

$m \times m$ macierz $X^T X$ jest macierzą Grama wektorów X_i. Zatem jeżeli X_i są liniowo niezależne to macierz $X^T X$ jest nieujemnie określona, a zatem odwracalna (por. [1] §VI.11 Wniosek 11.4).

Krok 2. Pokażemy, że B_{min} to punkt w którym przyjmowane jest minimum globalne.

$$B = B_{min} + b, \qquad b \neq 0,$$
$$SKR(B_{min} + b) = (Y^T - B_{min}^T X^T - b^T X^T)(Y - XB_{min} - Xb) =$$
$$= (Y^T - B_{min}^T X^T)(Y - XB_{min}) - (Y^T - B_{min}^T X^T) Xb - b^T X^T (Y - XB_{min}) + b^T X^T Xb =$$
$$= SKR(B_{min}) - 2b^T X^T (Y - XB_{min}) + b^T X^T Xb.$$

Zauważmy, że drugi człon jest równy 0

$$X^T(Y - XB_{min}) = X^T(Y - X(X^TX)^{-1}X^TY) = X^TY - X^TX(X^TX)^{-1}X^TY = 0,$$

a trzeci jest nieujemny dla niezerowych b ponieważ macierz X^TX jest nieujemnie określona. Zatem dla $b \neq 0$

$$SKR(B_{min} + b) > SKR(B_{min}).$$

Krok 3. Wyznaczamy $SKR(B_{min})$.
Ponieważ jak pokazaliśmy powyżej $X^T(Y - XB_{min}) = 0$ to

$$SKR(B_{min}) = (Y^T - B_{min}^T X^T)(Y - XB_{min}) = Y^T(Y - XB_{min}) =$$
$$= Y^TY - Y^TXB_{min} = Y^TY - Y^TX(X^TX)^{-1}X^TY.$$

□

Wniosek 2.1. *Dla $B = B_{min}$ zachodzą następujące zależności:*
1. Wektor składników resztowych ξ jest prostopadły do wszystkich kolumn X_i

$$X^T\xi = 0.$$

2. Wektor składników resztowych ξ jest prostopadły do wektora $\widehat{Y}$

$$\widehat{Y}^T\xi = 0.$$

3. Uogólnione twierdzenie Pitagorasa

$$Y^TY = \widehat{Y}^T\widehat{Y} + \xi^T\xi \quad \textit{czyli} \quad \|Y\|^2 = \|\widehat{Y}\|^2 + \|\xi\|^2.$$

Dowód.
Ad 1. Z definicji ξ mamy

$$X^T\xi = X^T(Y - XB) = X^T(Y - X(X^TX)^{-1}X^TY) = X^TY - X^TX(X^TX)^{-1}X^TY = 0.$$

Ad 2. $\widehat{Y}$ jest kombinacją liniową X_i zatem

$$\widehat{Y}^T\xi = B^TX^T\xi = 0.$$

Ad 3. Ponieważ ξ i $\widehat{Y}$ są prostopadłe to

$$Y^TY = (\widehat{Y}^T + \xi^T)(\widehat{Y} + \xi) = \widehat{Y}^T\widehat{Y} + \xi^T\xi.$$

Uwaga 2.1. Gdy ciągi X_1, ... , X_m są liniowo zależne to wybieramy spośród nich maksymalny podzbiór liniowo niezależny X_{j_1}, ... , X_{j_k} ($k = rank\ X < m$). Niech $\widetilde{X}$ będzie $n \times k$ macierzą, której kolumnami są X_{j_i}.
Zmienna modelowa jest wyznaczona jednoznacznie (niezależnie od wyboru ciągów liniowo niezależnych)

$$\widehat{Y} = \widetilde{X}\widetilde{B}_{min},$$

gdzie

$$\widetilde{B}_{min} = (\widetilde{X}^T\widetilde{X})^{-1}\widetilde{X}^TY.$$

Natomiast SKR przyjmuje minimum na podprzestrzeni afinicznej złożonej z punktów postaci

$$B = B^* + b,$$

gdzie

$$B_j^* = \begin{cases} \widetilde{B}_{min,i} & \text{gdy} \quad j = j_i, \\ 0 & \text{gdy} \quad j \notin \{j_1, \ldots, j_k\}, \end{cases}$$

a wektory b opisują zależności między ciągami X_i

$$b \in ker(X) = \{v \in \mathbb{R}^m : Xv = 0\}.$$

Ponadto spełnione są punkty 1,2 i 3 z powyższego wniosku.

2.2. Odrobina algebry liniowej

Oznaczmy przez $\mathcal{X}$ podprzestrzeń liniową przestrzeni $\mathbb{R}^n$ rozpiętą przez kolumny macierzy X,

$$\mathcal{X} = lin(X_1, \ldots, X_m) = \{XV : V \in \mathbb{R}^m\}.$$

Lemat 2.1. *Macierz kwadratowa $n \times n$*

$$P = X(X^T X)^{-1} X^T$$

jest macierzą rzutu prostopadłego na podprzestrzeń $\mathcal{X}$, a macierz

$$M = Id_n - P$$

macierzą rzutu prostokątnego na podprzestrzeń $\mathcal{X}^\perp$ (dopełnienie ortogonalne $\mathcal{X}$).

Dowód.
Mnożenie przez macierz P zachowuje wektory z $\mathcal{X}$

$$P(XV) = X(X^T X)^{-1} X^T XV = X((X^T X)^{-1} X^T X)V = XV$$

i anihiluje wektory prostopadłe do $\mathcal{X}$

$$X^T W = 0 \Rightarrow PW = X(X^T X)^{-1} X^T W = X(X^T X)^{-1}(X^T W) = 0.$$

Natomiast mnożenie przez macierz M anihiluje wektory z $\mathcal{X}$ i zachowuje wektory prostopadłe do $\mathcal{X}$

$$M(XV) = XV - (PX)V = XV - XV = 0, \quad MW = W - PW = W.$$

□

Lemat 2.2. *1. Macierze P i M są symetryczne i idempotentne*

$$P^T = P, \quad M^T = M, \quad PP = P, \quad MM = M.$$

2. Rząd macierzy P wynosi m, a M $n - m$.

$$rk\, P = m, \quad rk\, M = n - m.$$

3. Ślad macierzy P wynosi m, a M $n - m$.

$$tr\, P = m, \quad tr\, M = n - m.$$

4. Istnieje taka $n \times n$ macierz unitarna U (tzn. $U^T U = Id$), że macierze $U^T PU$ i $U^T MU$ są diagonalne o wyrazach 0 lub 1. $U^T PU$ ma na przekątnej m jedynek, a $U^T MU$ $n - m$.

Dowód.
Ad.1. P i M są macierzami rzutów zatem $PP = P$ i $MM = M$. Symetria wynika z faktu, że transpozycja jest przemienna z odwracaniem macierzy

$$P^T = (X(X^T X)^{-1} X^T)^T = X((X^T X)^T)^{-1} X^T = X(X^T X)^{-1} X^T = P,$$

$$M^T = (Id - P)^T = Id^T - P^T = Id - P = M.$$

Ad.2. Rząd macierzy jest równy wymiarowi obrazu, zatem

$$rk\, P = dim\, \mathcal{X} = rk\, X = m,$$

$$rk\, M = dim\, \mathcal{X}^{\perp} = n - m.$$

Ad.3. P jest macierzą rzutu na podprzestrzeń m wymiarową, a zatem ma m wartości własnych równych 1 i $n - m$ równych 0. Natomiast M jest macierzą rzutu na podprzestrzeń $n - m$ wymiarową, a zatem ma $n - m$ wartości własnych równych 1 i m równych 0. Ponieważ ślad jest to suma wartości własnych to wynosi on odpowiednio m i $n - m$.
Ad.4. Niech wektory $U_1, \ldots, U_m$ tworzą bazę ortonormalną podprzestrzeni $\mathcal{X}$, a $U_{m+1}, \ldots U_n$ bazę podprzestrzeni $\mathcal{X}^{\perp}$. Niech U będzie macierzą o kolumnach U_i. Wówczas

$$U_i^T U_j = \begin{cases} 1 & \text{gdy} \quad i = j, \\ 0 & \text{gdy} \quad i \neq j, \end{cases}$$

$$U_i^T P U_j = \begin{cases} 1 & \text{gdy} \quad i = j \leqslant m, \\ 0 & \text{gdy} \quad i \neq j \vee i > m, \end{cases}$$

$$U_i^T M U_j = \begin{cases} 1 & \text{gdy} \quad i = j > m, \\ 0 & \text{gdy} \quad i \neq j \vee i \leqslant m. \end{cases}$$

Zatem wszystkie trzy macierze są diagonalne i zero-jedynkowe.

□

3. MNK w terminach statystyki opisowej

Metoda MNK dla modeli z wyrazem wolnym. Współczynnik determinacji. Przypadek $k = 2$. (1 wykład)

3.1. Notacja statystyki opisowej

Będziemy stosowali następującą notację:
Dla pojedynczej serii danych $X = (X_t)_{t=1}^n$:
• średnia

$$\overline{X} = \frac{1}{n}\sum_{t=1}^{n} X_t,$$

• wariancja empiryczna (wariancja z próby)

$$S^2(X) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})^2,$$

• empiryczne odchylenie standardowe (odchylenie standardowe z próby)

$$S(X) = \sqrt{S^2(X)}.$$

Dla dwóch serii danych $Y = (Y_t)_{t=1}^n$ i $X = (X_t)_{t=1}^n$:
• kowariancja empiryczna (kowariancja z próby)

$$Cov(X,Y) = \frac{1}{n}\sum_{t=1}^{n}(X_t - \overline{X})(Y_t - \overline{Y}).$$

Uwaga 3.1. Zachodzą następujące związki

$$Cov(X,Y) = Cov(Y,X), \quad Cov(X,X) = S^2(X),$$

$$Cov(X,Y) = \overline{XY} - \overline{X}\,\overline{Y}, \quad S^2(X) = \overline{X^2} - \overline{X}^2.$$

• współczynnik korelacji Pearsona (korelacja empiryczna)

$$r(X,Y) = \frac{Cov(X,Y)}{S(x)S(Y)} \quad \text{gdy } S(X) \neq 0 \neq S(Y).$$

Uwaga 3.2. Zachodzą następujące związki

$$r(X,Y) \in [0,1], \quad r(X,X) = 1, \quad r(X,-X) = -1.$$

Dla m serii danych $X_i = (X_{t,i})_{t=1}^n$, $i = 1, \ldots, m$:
• macierz kowariancji serii X_i:

$$C(X) = Var(X) = \frac{1}{n}(X - e\overline{X})^T(X - e\overline{X})$$

gdzie X jest $n \times m$ macierzą o współczynnikach $X_{t,i}$, a $\overline{X}$ wektorem horyzontalnym o m wyrazach (tzn. macierzą $1 \times m$) a e wektorem kolumnowym o n wyrazach (tzn. macierzą $n \times 1$)

$$\overline{X} = (\overline{X_1}, \ldots, \overline{X_m}), \quad e = (1, \ldots, 1)^T.$$

Uwaga 3.3. Macierz C jest symetryczna i nieujemnie określona. Ponadto

$$C(X)_{i,j} = Cov(X_i, X_j), \quad C(X)_{i,i} = S^2(X_i),$$

$$C(X) = \frac{1}{n}X^TX - \overline{X}^T\overline{X}.$$

Dla $m + 1$ serii danych $X_i = (X_{t,i})_{t=1}^n$, $i = 1, \ldots, m$ i $Y = (Y_t)_{t=1}^n$:
• macierz kowariancji serii Y i serii X_i, $i = 1, \ldots, m$:

$$Cov(X, Y) = \frac{1}{n}(X - e\overline{X})^T(Y - e\overline{Y}).$$

Uwaga 3.4. Zachodzą następujące związki

$$Cov(X, Y)_j = Cov(X_j, Y), \quad Cov(X, Y) = \frac{1}{n}X^TY - \overline{X}^T\overline{Y}.$$

3.2. MNK z wyrazem wolnym

Rozważmy przypadek gdy jeden z ciągów X_i, $i = 1, \ldots, m$ jest stały. Dla uproszczenia przyjmijmy $X_m = e$ (tzn. $\forall t\ X_{t,m} = 1$). Wówczas dla wszystkich $t \in \{1, \ldots, n\}$

$$\widehat{Y}_t = b_1X_{t,1} + \cdots + b_{m-1}X_{t,m-1} + d,$$

gdzie d nazywamy wyrazem wolnym. W zapisie macierzowym wygląda to następująco

$$\widehat{Y} = X'B' + de,$$

gdzie X' jest $n \times (m - 1)$ macierzą o kolumnach $X_1, \ldots, X_{m-1}$ a $B' = (b_1, \ldots, b_{m-1})^T$. Zatem suma kwadratów reszt wyniesie

$$SKR(b_1, \ldots, b_{m-1}, d) = (Y - X'B' - de)^T(Y - X'B' - de).$$

Twierdzenie 3.1. *Jeżeli ciągi $X_1, \ldots, X_{m-1}, X_m = e$ są liniowo niezależne to SKR przyjmuje minimum w punkcie*

$$B'_{min} = C(X')^{-1}Cov(X', Y), \quad d_{min} = \overline{Y} - \overline{X'}B'_{min}.$$

Ponadto

$$SKR_{min} = n(S^2(Y) - Cov(X', Y)^TC(X')^{-1}Cov(X', Y)).$$

Dowód.
Krok 1. Pokażemy, że macierz $C(X')$ jest dodatnio określona a zatem odwracalna.
Rozważmy dowolny niezerowy wektor B'. Wektor $Z = X'B'$ nie jest stały, zatem

$$0 < S^2(Z) = B'^T C B'.$$

Krok 2. Korzystając ze wzoru na B_{min} wyprowadzonego w twierdzeniu 2.1 wyznaczymy B'_{min} i d_{min}.
B_{min} spełnia zależność

$$X^T X B_{min} = X^T Y.$$

Korzystając z faktu, że $X = (X', e)$ (tzn. macierz X powstaje z X' przez dopisanie kolumny jedynek) a $B^T_{min} = (B'^T_{min}, d_{min})$, zapiszemy ją w terminach X', B'_{min} i d_{min}

$$\begin{pmatrix} X'^T X' & n\overline{X'}^T \\ n\overline{X'} & n \end{pmatrix} \circ \begin{pmatrix} B'_{min} \\ d_{min} \end{pmatrix} = \begin{pmatrix} X'^T Y \\ n\overline{Y} \end{pmatrix}.$$

Dzielimy obie strony przez n

$$\begin{aligned} \frac{1}{n} X'^T X' B'_{min} + \overline{X'}^T d_{min} &= \frac{1}{n} X'^T Y, \\ \overline{X'} B'_{min} + d_{min} &= \overline{Y}. \end{aligned}$$

Z drugiego równania otrzymujemy formułę na d_{min}, a następnie eliminujemy d_{min} z pierwszego równania. Po uporządkowaniu składników otrzymujemy

$$\left(\frac{1}{n} X'^T X' - \overline{X'}^T \overline{X'}\right) B' = \frac{1}{n} X'^T Y - \overline{X'}^T \overline{Y}.$$

Co możemy zapisać w postaci (patrz uwagi 3.3 i 3.4)

$$C(X') B'_{min} = Cov(X', Y).$$

Krok 3. Wyznaczamy SKR_{min}.

$$SKR_{min} = SKR(B'_{min}, d_{min}) = \sum_{t=1}^{n} (Y_t - \sum_{i=1}^{m-1} b_{min,i} X_{t,i} - d_{min})^2$$

Po podstawieniu $d_{min} = \overline{Y} - \overline{X'} B'_{min}$ otrzymujemy

$$SKR_{min} = \sum_{t=1}^{n} ((Y_t - \overline{Y}) - \sum_{i=1}^{m-1} b_{min,i} (X_{t,i} - \overline{X_i}))^2 =$$

$$= n(S^2(Y) - 2 \sum_{i=1}^{m-1} b_{min,i} Cov(X_i, Y) + S^2 \left(\sum_{i=1}^{m-1} b_{min,i} X_i\right)) =$$

$$= n(S^2(Y) - 2Cov(X', Y)^T B'_{min} + B'^T_{min} C(X') B'_{min}) = n(S^2(Y) -$$

$$-2Cov(X', Y)^T C(X')^{-1} Cov(X', Y) + Cov(X', Y)^T C(X')^{-1} C(X') C(X')^{-1} Cov(X', Y)) =$$

$$= n(S^2(Y) - Cov(X', Y)^T C(X')^{-1} Cov(X', Y)).$$

□

Uwaga 3.5. Dla $B' = B'_{min}$ i $d = d_{min}$ zachodzą następujące związki:

$$\begin{aligned} &1. \quad \overline{\xi} = 0, \quad \overline{\widehat{Y}} = \overline{Y} \\ &2. \quad \sum_{t=1}^{n}(Y_t - \overline{Y})^2 = \sum_{t=1}^{n}(\widehat{Y}_t - \overline{Y})^2 + \sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2. \end{aligned}$$

Dowód.
Ad.1. Mamy

$$Y_t = \sum_{i=1}^{m-1} b_i X_{t,i} + d + \xi_t.$$

Zatem

$$\overline{Y} = \sum_{i=1}^{m-1} b_i \overline{X_i} + d + \overline{\xi},$$

czyli

$$\overline{Y} - \overline{\widehat{Y}} = \overline{\xi} = \overline{Y} - \sum_{i=1}^{m-1} b_i \overline{X_i} - d = 0.$$

Ad.2. Z punktu 1 i z wniosku 2.1 wynika:

$$\sum_{t=1}^{n}(Y_t - \overline{Y})^2 - \sum_{t=1}^{n}(\widehat{Y}_t - \overline{Y})^2 - \sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2 =$$

$$= \sum_{t=1}^{n}(Y_t^2 - \overline{Y}^2) - \sum_{t=1}^{n}(\widehat{Y}_t^2 - \overline{Y}^2) - \sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2 =$$

$$= \sum_{t=1}^{n} Y_t^2 - \sum_{t=1}^{n} \widehat{Y}_t^2 - \sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2 = 0.$$

□

Definicja 3.1. Współczynnik determinacji zwany też współczynnikiem dopasowania i współczynnikiem regresji wielorakiej to

$$R^2 = 1 - \frac{\sum_{t=1}^{n} \xi_t^2}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}.$$

Uwaga 3.6.

$$R^2 = \frac{\sum_{t=1}^{n}(\widehat{Y}_t - \overline{Y})^2}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2} = \frac{Cov(X', Y)^T C(X')^{-1} Cov(X', Y)}{S^2(Y)}.$$

Definicja 3.2. Średni błąd kwadratowy

$$MSE = \frac{1}{n}\sum_{t=1}^{n} \xi_t^2 = \overline{\xi^2}.$$

Uwaga 3.7.

$$MSE = S^2(Y)(1 - R^2).$$

Podsumowanie.
R^2 i MSE określają dokładność aproksymacji przy zastosowaniu metody najmniejszych kwadratów (MNK).

3.3. Przypadek $m = 2$ i $X_2 = e$

$$\widehat{Y}_t = bX_t + d,$$
$$SKR = \sum_{t=1}^{n}(Y_t - bX_t - d)^2.$$

Twierdzenie 3.2. *Jeżeli ciąg X_t nie jest stały to SKR przyjmuje minimum w punkcie*

$$b_{min} = \frac{Cov(X,Y)}{S^2(X)}, \quad d_{min} = \overline{Y} - b_{min}\overline{X}.$$

$$SKR_{min} = nS^2(Y)(1 - r^2(X,Y).$$

Dowód.

$$SKR_{min} = n\left(S^2(Y) - \frac{cov^2(X,Y)}{S^2(X)}\right) = n\left(S^2(Y) - \frac{S^2(X)S^2(Y)r^2(X,Y)}{S^2(X)}\right) = nS^2(Y)(1-r^2(X,Y)).$$

□

Zamieniamy rolami Y i X.

$$\widehat{X}_t = fY_t + g, \quad SKR = \sum_{t=1}^{n}(X_t - fY_t - g)^2.$$

Otrzymujemy

$$f_{min} = \frac{cov(X,Y)}{S^2(Y)}, \quad g_{min} = \overline{X} - f_{min}\overline{Y}.$$

Okazuje się, że proste $Y = b_{min}X + d_{min}$ i $X = f_{min}Y + g_{min}$ na ogół nie pokrywają się. Przecinają się one w punkcie $(\overline{X}, \overline{Y})$ i iloczyn współczynników kierunkowych wynosi $r^2(X,Y)$

$$b_{min}f_{min} = \frac{cov(X,Y)}{S^2(X)}\frac{cov(X,Y)}{S^2(Y)} = r^2(X,Y).$$

4. Klasyczny model regresji

Klasyczny jednorównaniowy liniowy model ekonometryczny. Założenia modelu. Estymacja parametrów strukturalnych modelu metodą najmniejszych kwadratów (regresja wieloraka). (1 wykład)

4.1. Notacja macierzowa dla zmiennych losowych

Definicja 4.1. Niech X będzie $m \times n$ macierzą losową, której wyrazami są zmienne losowe $X_{i,j}$ określone na tej samej przestrzeni probabilistycznej. Wartością oczekiwaną X będziemy nazywać $m \times n$ macierz $E(X)$ taką, że

$$E(X)_{i,j} = E(X_{i,j}).$$

Uwaga 4.1. Wartość oczekiwana macierzy jest zgodna z transpozycją

$$E(X^T) = E(X)^T$$

oraz z mnożeniem przez macierze deterministyczne

$$E(AXB) = AE(X)B,$$

gdzie A i B macierze o współczynnikach rzeczywistych odpowiednio wymiaru $k \times m$ i $n \times p$.

Definicja 4.2. Niech X będzie $m \times 1$ macierzą losową (wektorem kolumnowym), której wyrazami są zmienne losowe X_i określone na tej samej przestrzeni probabilistycznej. Macierzą kowariancji X będziemy nazywać $m \times m$ macierz $Var(X)$

$$Var(X) = E((X - E(X))(X - E(X))^T).$$

Uwaga 4.2. Zachodzą następujące związki

$$Var(X) = E(XX^T) - E(X)E(X)^T,$$

$$Var(X)_{i,i} = D^2(X_i), \quad Var(X)_{i,j} = Cov(X_i, X_j).$$

$Var(X)$ jest macierzą symetryczną

$$Var(X)^T = Var(X).$$

Ponadto dla deterministycznej $k \times m$ macierzy A

$$Var(AX) = AVar(X)A^T.$$

Definicja 4.3. Niech X i Y będą wektorami kolumnowymi, których wyrazami są zmienne losowe X_i, $i = 1, \ldots, m_1$ i Y_j, $j = 1, \ldots, m_2$ określone na tej samej przestrzeni probabilistycznej. Macierzą kowariancji X i Y będziemy nazywać $m_1 \times m_2$ macierz $Cov(X, Y)$

$$Cov(X, Y) = E((X - E(X))(Y - E(Y))^T).$$

Uwaga 4.3. Zachodzą następujące związki

$$Cov(X, Y) = E(XY^T) - E(X)E(Y)^T,$$

$$Cov(X, Y)_{i,j} = Cov(X_i, Y_j),$$

$$Cov(Y, X) = Cov(X, Y)^T.$$

Ponadto dla deterministycznych $k \times m_1$ macierzy A i $p \times m_2$ B

$$Cov(AX, BY) = ACov(X, Y)B^T.$$

4.2. Warunkowa wartość oczekiwana

Niech $(\Omega, \mathcal{M}, P)$ będzie przestrzenią probabilistyczną, $\mathcal{F}$ σ-ciałem zawartym w $\mathcal{M}$ a Y zmienną losową określoną na $(\Omega, \mathcal{M}, P)$.

Definicja 4.4. Warunkową wartością oczekiwaną Y pod warunkiem $\mathcal{F}$ nazywamy każdą zmienną losową $E(Y|\mathcal{F})$ o wartościach w $\mathbb{R} \cup \{\pm\infty\}$ spełniającą warunki:
i) $E(Y|\mathcal{F})$ jest $\mathcal{F}$ mierzalna;
ii) Dla każdego $A \in \mathcal{F}$

$$\int_A Y dP = \int_A E(Y|\mathcal{F}) dP.$$

Lemat 4.1. *Każdy z poniższych warunków implikuje istnienie warunkowej wartości oczekiwanej $E(Y|\mathcal{F})$.*
1. $E(Y)$ jest określona (tzn. skończona lub nieskończona).
2. Y należy do $L^1(\Omega, \mathcal{M}, P)$.
3. $Y \geqslant 0$ p.n. lub $Y \leqslant 0$ p.n.

Uwaga 4.4. Warunkowa wartość oczekiwana ma następujące własności:
1. $Y \geqslant 0$ p.n. to $E(Y|\mathcal{F}) \geqslant 0$ p.n.
2. $E(1|\mathcal{F}) = 1$ p.n.
3. $E(Y_1 + Y_2|\mathcal{F}) = E(Y_1|\mathcal{F}) + E(Y_2|\mathcal{F})$ o ile prawa strona jest określona (tzn. różna od $\infty - \infty$),
4. Jeżeli zmienna losowa ξ jest $\mathcal{F}$ mierzalna i wartość oczekiwana $E(\xi Y)$ jest określona to

$$E(\xi Y|\mathcal{F}) = \xi E(Y|\mathcal{F}).$$

5. Jeżeli wartość oczekiwana $E(Y)$ jest określona to dla dowolnego σ-ciała $\mathcal{G}$ zawartego w $\mathcal{F}$

$$E(Y|\mathcal{G}) = E(E(Y|\mathcal{F})|\mathcal{G}) \text{ p.n.}$$

W szczególności

$$E(Y) = E(E(Y|\mathcal{F})).$$

4.3. Założenia klasycznego modelu regresji

W modelu regresji rozważa się zmienną objaśnianą (zależną, zwaną też regressandem) - Y i zmienne objaśniające (zwane regressorami) - $X_1, \ldots, X_K$.

Dysponujemy próbką złożoną z n obserwacji. i-tą obserwację modelujemy jako realizację $K+1$ wymiarowej zmiennej losowej

$$(Y_i, X_{i,1}, \ldots, X_{i,K}), \quad i = 1, \ldots, n.$$

Przez model rozumie się łączny rozkład zmiennych losowych Y_i i $X_{i,k}$ spełniający pewne założenia.

Założenia modelu.

Z1. Liniowość.
Zmienne losowe Y_i i $X_{i,k}$ należą do L^2 i spełniają zależność

$$Y_i = \beta_1 X_{i,1} + \cdots + \beta_K X_{i,K} + \varepsilon_i, \quad i = 1, \ldots, n,$$

gdzie $\beta_i \in \mathbb{R}$ to deterministyczne choć na ogół nieznane parametry regresji zwane też parametrami strukturalnymi modelu, zaś zmienne losowe ε_i to składniki losowe. Funkcję

$$f(x) = \beta_1 x_1 + \beta_2 x_2 + \cdots + \beta_K x_K, \quad x \in \mathbb{R}^K,$$

nazywa się funkcją regresji. Warunek liniowości można zapisać w postaci macierzowej

$$Y = X\beta + \varepsilon,$$

gdzie X macierz o wyrazach $X_{i,k}$, Y, β i ε wektory kolumnowe o wyrazach odpowiednio Y_i, β_k i ε_i.

Uwaga 4.5. Założenie Z1 implikuje przynależność ε do L^2.

Z2. Ścisła egzogeniczność.

$$E(\epsilon|X) = 0.$$

Wniosek 4.1. *Przy założeniach Z1 i Z2 dla wszystkich $i, j \in \{1, \ldots, n\}$ i $k \in \{1, \ldots, K\}$ zachodzą następujące równości:*
1. $E(\varepsilon_i) = 0$;
2. $E(\varepsilon_i X_{j,k}) = 0$;
3. $Cov(\varepsilon_i, X_{j,k}) = 0$.

Dowód.
Ponieważ zarówno ε_i jak i $X_{j,k}$ należą do L^2 (to wynika z Z1) to możemy stosować twierdzenie o iterowanej wartości oczekiwanej.

$$\begin{array}{lrl}
\text{Ad1.} & E(\varepsilon_i) = & E(E(\varepsilon_i|X)) = E(0) = 0. \\
\text{Ad2.} & E(\varepsilon_i X_{j,k}) = & E(E(\varepsilon_i X_{j,k}|X)) = E(X_{j,k} E(\varepsilon_i|X)) = E(0) = 0. \\
\text{Ad3.} & Cov(\varepsilon_i, X_{j,k}) = & E(\varepsilon_i X_{j,k}) - E(\varepsilon_i)E(X_{j,k}) = 0 - 0 = 0.
\end{array}$$

□

Z3. Liniowa niezależność.

$$(P(rk(X) = K) = 1, \text{ i } (X^T X)^{-1} \in L^1.$$

Warunek Z3 oznacza, że kolumny macierzy X są prawie na pewno liniowo niezależne.

Z4. Sferyczność błędu

$$E(\varepsilon\varepsilon^T|X) = \sigma^2 Id_n,$$

gdzie $\sigma > 0$ deterministyczny parametr modelu.

Warunek Z4 można rozłożyć na dwa warunki:

Z4.1. Homoskedastyczność

$$E(\varepsilon_i^2|X) = \sigma^2.$$

Z4.2. Brak korelacji, dla $i \neq j$

$$E(\varepsilon_i\varepsilon_j|X) = 0.$$

Wniosek 4.2. *Przy założeniach Z1, Z2 i Z4 dla wszystkich $i, j \in \{1, \ldots, n\}$, $i \neq j$ zachodzą następujące równości:*
1. $D^2(\varepsilon_i) = \sigma^2$;
2. $Cov(\varepsilon_i, \varepsilon_j) = 0$.

Z5. Gaussowskość.
Łączny rozkład warunkowy ε względem X jest normalny.

Wniosek 4.3. *Przy założeniach Z1, Z2, Z4 i Z5:*
1. $\varepsilon|X \sim N(0, \sigma^2 Id_n)$;
2. $\varepsilon \sim N(0, \sigma^2 Id_n)$.

Dowód.
Punkt 1 wynika z założeń Z2 i Z5.
Punkt 2 wynika z faktu, że parametry warunkowego rozkładu ε nie zależy od X. Rzeczywiście, niech $F(e_1, \ldots, e_n)$ będzie dystrybuantą rozkładu $N(0, \sigma^2 Id_n)$, wówczas

$$P(\varepsilon_i \leqslant e_i, i = 1, \ldots, n) = E(\prod_{i=1}^{n} 1_{\varepsilon_i \leqslant e_i}) = E(E(\prod_{i=1}^{n} 1_{\varepsilon_i \leqslant e_i}|X)) = E(F(e_1, \ldots, e_n)) = F(e_1, \ldots, e_n).$$

□

4.4. Estymacja parametrów modelu metodą MNK

Estymatorem MNK wektora β jest wektor

$$B = (X^T X)^{-1} X^T Y.$$

Natomiast estmatorem MNK wariancji σ^2 jest

$$S_Y^2 = \frac{\xi^T \xi}{n - K} = \frac{SKR_{min}}{n - K},$$

gdzie $\xi = Y - XB$.

Twierdzenie 4.1. *Własności estymatorów B i S_y:*
a) nieobciążoność B. Jeśli zachodzą Z1,Z2 i Z3 to

$$B \in L^1 \quad i \quad E(B|X) = \beta.$$

b) skończona wariancja B. Jeśli zachodzą Z1,Z2,Z3 i Z4 to

$$Var(B|X) = \sigma^2(X^TX)^{-1} \quad i \quad B \in L^2.$$

c) efektywność (tw. Gaussa-Markowa). Jeśli zachodzą Z1,Z2,Z3 i Z4 to estymator MNK jest najefektywniejszy w klasie liniowych po Y, nieobciążonych estymatorów liniowych modeli.

$$\forall\beta \ \forall\widehat{\beta}\text{- lin. nieob. est.} \quad Var(\widehat{\beta}|X) - Var(B|X) \geqslant 0.$$

d) nieobciążoność S_Y. Jeśli zachodzą Z1,Z2,Z3 i Z4 to

$$\xi \in L^2 \quad i \quad E(S_Y^2|X) = \sigma^2.$$

e) ortogonalność B do składnika resztowego ξ. Jeśli zachodzą Z1,Z2,Z3 i Z4 to

$$Cov(B, \xi|X) = 0.$$

Dowód.
Ad.a. Najpierw pokażemy, że warunki Z3 i Z1 implikują przynależność B do L^1. Mamy

$$((X^TX)^{-1}X^T)((X^TX)^{-1}X^T)^T = (X^TX)^{-1}X^TX(X^TX)^{-1} = (X^TX)^{-1} \in L^1.$$

Zatem wszystkie wyrazy macierzy $(X^TX)^{-1}X^T$ należą do L^2. Ponieważ również $Y \in L^2$, to $B = ((X^TX)^{-1}X^T)Y$ należy do L^1.

Następnie pokażemy, że $E((B-\beta)|X) = 0$.
Mamy dwa równania opisujące zależność Y od X:

$$\begin{aligned} Y &= X\beta + \varepsilon, \\ Y &= XB + \xi. \end{aligned}$$

Po odjęciu stronami otrzymujemy:

$$X(B-\beta) = \varepsilon - \xi. \tag{4.1}$$

Mnożymy obie strony przez $(X^TX)^{-1}X^T$

$$(X^TX)^{-1}X^TX(B-\beta) = (X^TX)^{-1}X^T\varepsilon - (X^TX)^{-1}X^T\xi.$$

Biorąc pod uwagę, że $X^T\xi = 0$ (patrz wniosek 2.1) otrzymujemy:

$$B - \beta = (X^TX)^{-1}X^T\varepsilon. \tag{4.2}$$

Zatem

$$E((B-\beta)|X) = E((X^TX)^{-1}X^T\varepsilon|X) = (X^TX)^{-1}X^TE(\varepsilon|X) = 0.$$

Ad.b. Pokażemy, że dla każdego wektora kolumnowego $v \in \mathbb{R}^K \quad D^2(v^T B|X) = \sigma^2 v^T (X^T X)^{-1} v$. Skorzystamy z faktu, że wartości oczekiwane nieujemnych zmiennych losowych są zawsze określone.

$$\begin{aligned}
D^2(v^T B|X) &= D^2(v^T(B-\beta)|X) = D^2(v^T(X^TX)^{-1}X^T\varepsilon|X) = \\
&= E(v^T(X^TX)^{-1}X^T\varepsilon\varepsilon^T X(X^TX)^{-1}v|X) = \\
&= v^T(X^TX)^{-1}X^T E(\varepsilon\varepsilon^T|X)X(X^TX)^{-1}v = \\
&= v^T(X^TX)^{-1}X^T(\sigma^2 Id_n)X(X^TX)^{-1}v = \\
&= \sigma^2 v^T(X^TX)^{-1}X^TX(X^TX)^{-1}v = \\
&= \sigma^2 v^T(X^TX)^{-1}v.
\end{aligned}$$

Założenie, że $(X^TX)^{-1} \in L^1$ (Z3) implikuje skończoność wariancji $v^T B$ dla każdego v, a więc i wariancji B.

$$\begin{aligned}
D^2(v^T B) &= D^2(v^T(B-\beta)) = E(v^T(B-\beta)(B-\beta)^T v) = \\
&= E(E(v^T(B-\beta)(B-\beta)^T v|X)) = \sigma^2 v^T E(X^TX)^{-1})v < +\infty.
\end{aligned}$$

Ad.c. Niech $\widehat{\beta}$ będzie dowolnym nieobciążonym i liniowym po Y estymatorem dla modeli liniowych z K parametrami strukturalnymi i n obserwacjami. Wówczas istnieje funkcja macierzowa $C(\cdot)$ $(C \in K \times n)$, taka, że

$$\widehat{\beta} = C(X)Y.$$

Niech $G = C - (X^TX)^{-1}X^T$.

$$\widehat{\beta} = (G + (X^TX)^{-1}X^T)Y = GY + (X^TX)^{-1}X^TY = G(X\beta + \varepsilon) + B.$$

Ponieważ oba estymatory B i $\widehat{\beta}$ są nieobciążone to

$$\beta = E(\widehat{\beta}|X) = GX\beta + GE(\varepsilon|X) + E(B|X) = GX\beta + 0 + \beta.$$

Czyli dla dowolnego wektora $\beta \quad GX\beta = 0$, a zatem $GX = 0$. W efekcie otrzymujemy:

$$\widehat{\beta} = G\varepsilon + B,$$

$$\widehat{\beta} - \beta = G\varepsilon + B - \beta = (G + (X^TX)^{-1}X^T)\varepsilon.$$

Teraz możemy wyznaczyc warunkową wariancje $\widehat{\beta}$.

$$\begin{aligned}
Var(\widehat{\beta}|X) &= Var(\widehat{\beta} - \beta) = Var((G + (X^TX)^{-1}X^T)\varepsilon|X) = \\
&= E((G + (X^TX)^{-1}X^T)\varepsilon\varepsilon^T(G^T + X(X^TX)^{-1})|X) = \\
&= (G + (X^TX)^{-1}X^T)E(\varepsilon\varepsilon^T|X)(G^T + X(X^TX)^{-1}) = \\
&= (G + (X^TX)^{-1}X^T)(\sigma^2 Id_n)(G^T + X(X^TX)^{-1}) = \\
&= \sigma^2(GG^T + GX(X^TX)^{-1} + (X^TX)^{-1}X^TG^T + (X^TX)^{-1}).
\end{aligned}$$

Ponieważ $GX = 0$ a $Var(B|X) = \sigma^2(X^TX)^{-1}$, to

$$Var(\widehat{\beta}|X) - Var(B|X) = \sigma^2 GG^T \geqslant 0.$$

Ad.d. Z równań 4.1 i 4.2 otrzymujemy, że

$$\xi = \varepsilon - X(B-\beta) = (Id - X(X^TX)^{-1}X^T)\varepsilon = M\varepsilon.$$

Jak pokazaliśmy w lemacie 2.2 macierz M jest symetryczna i idempotentna, zatem

$$\xi^T\xi = \varepsilon^T M M\varepsilon = \varepsilon^T M\varepsilon. \tag{4.3}$$

Ponieważ, $\xi^T\xi$ to suma kwadratów ξ_t to jej wartości oczekiwane są zawsze określone. Otrzymujemy na mocy warunku Z3 i lematu 2.2

$$\begin{aligned} E(\xi^T\xi|X) &= E(\varepsilon^T M\varepsilon|X) = E(\sum_{i,j=1}^{n} \varepsilon_i M_{i,j}\varepsilon_j|X) = \\ &= \sum_{i,j=1}^{n} M_{i,j}E(\varepsilon_i\varepsilon_j|X) = \sum_{i=1}^{n} M_{i,i}\sigma^2 = \sigma^2 trM = (n-K)\sigma^2. \end{aligned}$$

Ponadto

$$E(\xi^T\xi) = E(E(\xi^T\xi|X)) = (n-K)\sigma^2.$$

Zatem ξ należy do L^2.

Ad.e. Biorąc pod uwagę, że $\xi = \varepsilon - X(B-\beta)$ (równanie 4.1) i $E(\xi|X) = 0$ to

$$\begin{aligned} Cov(B,\xi|X) &= E((B-\beta)(\varepsilon - X(B-\beta))^T|X) = \\ &= E(-(B-\beta)(B-\beta)^T X^T + (B-\beta)\varepsilon^T|X) = \\ &= -Var(B-\beta|X)X^T + E((X^TX)^{-1}X^T\varepsilon\varepsilon^T|X) = \\ &= -\sigma^2(X^TX)^{-1}X^T + (X^TX)^{-1}X^TE(\varepsilon\varepsilon^T|X) = \\ &= -\sigma^2(X^TX)^{-1}X^T + (X^TX)^{-1}X^T(\sigma^2 Id_n) = 0. \end{aligned}$$

□

Wniosek 4.4. *"Bezwarunkowe" własności estymatora MNK B.*
a. Warunki Z1,Z2 i Z4 implikują, że $E(B) = \beta$.
b. Warunki Z1,Z2,Z3 i Z4 implikują, że $Cov(B,\xi) = 0$.

Dowód.

$$E(B) = E(E(B|X)) = E(\beta) = \beta.$$

$$Cov(B,\xi) = Cov(B-\beta,\xi) = E((B-\beta)\xi^T) = E(E((B-\beta)\xi^T|X)) = E(Cov(B,\xi|X)) = 0.$$

□

Wniosek 4.5. *Estymacja warunkowej kowariancji estymatora B.*
1. $S_Y^2(X^TX)^{-1}$ jest naturalnym nieobciążonym estymatorem $Var(B|X)$.
2. $S_Y^2(X^TX)^{-1}_{k,k}$ jest naturalnym nieobciążonym estymatorem $D^2(b_k|X)$.

Dowód.
Pokażemy, że dla każdego wektora kolumnowego $v \in \mathbb{R}^K$ $\quad E(S_Y^2 v^T(X^TX)^{-1}v|X) = \sigma^2 v^T(X^TX)^{-1}v$.
Skorzystamy z faktu, że wartości oczekiwane nieujemnych zmiennych losowych są zawsze określone.

$$E(S_Y^2 v^T(X^TX)^{-1}v|X) = E(S_Y^2|X)v^T(X^TX)^{-1}v = \sigma^2 v^T(X^TX)^{-1}v = Var(v^TB|X).$$

□

Oznaczenie.

$$SE(b_k) = \sqrt{S_Y^2(X^TX)^{-1}}.$$

Uwaga 4.6. Związek wariancji estymatora B z wielkością próby.
Załóżmy, że poszczególne wiersze macierzy X (czyli obserwacje) są niezależne od siebie i o tym samym rozkładzie co pewien horyzontalny wektor losowy Z. Wówczas z prawa wielkich liczb otrzymujemy, że istnieje pewna macierz C taka, że

$$\lim_{n\to\infty} \frac{1}{n} X^T X = C = E(ZZ^T).$$

Warunek Z3 implikuje, że macierz C jest odwracalna. Zatem

$$\lim_{n\to\infty} n(X^T X)^{-1} = C^{-1}.$$

W efekcie

$$\lim_{n\to\infty} Var(B|X) = \lim_{n\to\infty} \frac{\sigma^2}{n}(n(X^T X)^{-1}) = 0.$$

5. Klasyczny model regresji z gaussowskim składnikiem losowym

Klasyczny jednorównaniowy liniowy model ekonometryczny -cd. Statystyczna weryfikacja modelu. (1 wykład)

Na tym wykładzie zajmiemy się "kompletnym" modelem regresji, tzn. przyjmiemy wszystkie pięć założeń Z1 – Z5.

Lemat 5.1. *Jeśli zachodzą Z1,Z2,Z3,Z4 i Z5 to estymator NMK B ma warunkowy rozkład normalny*

$$B - \beta|X \ \sim N(0, \sigma^2(X^TX)^{-1}),$$

$$b_k - \beta_k|X \ \sim N(0, \sigma^2(X^TX)^{-1}_{k,k}).$$

Dowód.

$$B - \beta = (X^TX)^{-1}X^T\varepsilon, \quad \text{gdzie} \quad \varepsilon|X \sim N(\sigma^2 Id_n).$$

Zatem $B - \beta|X$ ma rozkład normalny o wartości oczekiwanej 0 i wariancji

$$(X^TX)^{-1}X^T(\sigma^2 Id_n)X(X^TX)^{-1} = \sigma^2(X^TX)^{-1}.$$

□

5.1. Testowanie pojedynczego parametru strukturalnego β_k

Niech $\bar{\beta}_k$ pewna ustalona liczba rzeczywista.
Testujemy hipotezę $H_0 : \beta_k = \bar{\beta}_k$ wobec hipotezy alternatywnej $H_1 : \beta_k \neq \bar{\beta}_k$.

Twierdzenie 5.1. *Przy założeniach Z1–Z5 i H_0 statystyka T_k*

$$T_k = \frac{b_k - \bar{\beta}_k}{SE(b_k)}$$

ma rozkład t-Studenta z $n - K$ stopniami swobody.

Uwaga 5.1. Rozkład T_k nie zależy od X.

Dowód.

$$SE(b_k) = \sqrt{S_Y^2 (X^T X)^{-1}_{k,k}},$$

zatem

$$T_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\sigma^2 (X^T X)^{-1}_{k,k}}} \cdot \sqrt{\frac{\sigma^2}{S_Y^2}} = \frac{z_k}{\sqrt{\frac{S_Y^2}{\sigma^2}}} =$$

$$= \frac{z_k}{\sqrt{\frac{\xi^T \xi}{n-K} \frac{1}{\sigma^2}}} = \frac{z_k}{\sqrt{\frac{q}{n-K}}}.$$

gdzie

$$z_k = \frac{b_k - \bar{\beta}_k}{\sqrt{\sigma^2 (X^T X)^{-1}_{k,k}}}, \quad z_k | X \sim N(0,1),$$

$$q = \frac{\xi^T \xi}{\sigma^2}.$$

Lemat 5.2. *Przy założeniach twierdzenia 5.1:*
1. $q|X \sim \chi^2(n-K)$.
2. q *i* z_k *są warunkowo względem* X *niezależne.*

Dowód.
Ad.1.

$$\xi^T \xi = \varepsilon^T M \varepsilon,$$

zatem

$$q = \frac{\xi^T \xi}{\sigma^2} = \frac{1}{\sigma} \varepsilon^T M \frac{1}{\sigma} \varepsilon.$$

$$\frac{1}{\sigma} \varepsilon \sim N(0, Id_n),$$

zaś po odpowiednim obrocie układu współrzędnych M jest macierzą diagonalną mającą na przekątnej $n-K$ jedynek i K zer, zatem

$$\frac{1}{\sigma} \varepsilon^T M \frac{1}{\sigma} \varepsilon \,|\, X \sim \chi^2(n-K).$$

Ad.2.

$$B = \beta + (X^T X)^{-1} \varepsilon, \;\; \xi = M \varepsilon,$$

zatem warunkowy względem X rozkład B i ξ jest normalny. Ale są one warunkowo nieskorelowane a zatem warunkowo niezależne. Ponieważ z_k zależy od B i X a q od ξ i X to są one warunkowo względem X niezależne.

□

Cd. dowodu twierdzenia.
Z lematu wynika, że

$$T_k = \frac{z_k}{\sqrt{\frac{q}{n-K}}}$$

ma warunkowy względem X rozkład t-Studenta z $n-K$ stopniami swobody. Ponieważ rozkład warunkowy nie zależy od warunkowania to T_k ma "bezwarunkowy" rozkład t-Studenta z $n-K$ stopniami swobody.

□

Reguła decyzyjna testu t.
Przedstawimy trzy równoważne warianty reguły decyzyjnej dla zadanego poziomu istotności α.

Wariant 1.
1. Na podstawie próbki ω wyznaczamy realizację statystyki testowej $t_k = T_k(\omega)$.
2. Wyznaczamy wartość krytyczną $t^*_{\alpha/2}$

$$P(|T| < t^*_{\alpha/2}) = 1 - \alpha, \quad T \sim t(n-K).$$

3. Jeżeli $|t_k| < t^*_{\alpha/2}$ to nie ma podstaw do odrzucenia hipotezy H_0 (akceptujemy H_0).
Jeżeli $|t_k| \geqslant t^*_{\alpha/2}$ to odrzucamy H_0 na rzecz H_1.

Wariant 2.
1. Na podstawie próbki wyznaczamy etymator b_k i jego błąd $SE(b_k)$.
2. Wyznaczamy przedział ufności I_α

$$I_\alpha = (b_k - SE(b_k)t^*_{\alpha/2}, b_k + SE(b_k)t^*_{\alpha/2}).$$

3. Jeżeli $\bar{\beta}_k \in I_\alpha$ to nie ma podstaw do odrzucenia hipotezy H_0 (akceptujemy H_0).
Jeżeli $\bar{\beta}_k \notin I_\alpha$ to odrzucamy H_0 na rzecz H_1.

Wariant 3.
1. Na podstawie próbki ω wyznaczamy realizację statystyki testowej $t_k = T_k(\omega)$.
2. Wyznaczamy prawdopodobieństwo (tzw. p-value)

$$p = 2P(T \geqslant |t_k|), \quad \text{dla} \quad T \sim t(n-K).$$

3. Jeżeli $p > \alpha$ to nie ma podstaw do odrzucenia hipotezy H_0 (akceptujemy H_0).
Jeżeli $p \leqslant \alpha$ to odrzucamy H_0 na rzecz H_1.

Uwaga 5.2. Najczęściej testujemy przypadek $\bar{\beta}_k = 0$. Wówczas przyjecie H_0 oznacza, że zmienną objaśniającą X_k należy wykluczyć z naszago modelu. Tzn. jeżeli

$$|b_k| < SE(b_k)t^*_{\alpha/2},$$

to parametr β_k nie jest statystycznie istotny.

5.2. Testowanie hipotezy liniowości

Zajmiemy sie teraz testowaniem hipotezy, że nieznany parametr $\beta = (\beta_1, \ldots, \beta_K)^T$ spełnia m niezależnych warunków liniowych. Czyli, że należy do podprzestrzeni afinicznej kowymiaru m.

Niech r macierz o współczynnikach rzeczywistych wymiaru $m \times K$, rzędu m, gdzie $m = 1, \ldots, K$, a $\tilde{r}$ wektor kolumnowy wymiaru m. Testujemy hipotezę

$$H_0: \quad r\beta = \tilde{r},$$

wobec

$$H_1: \quad r\beta \neq \tilde{r}.$$

Twierdzenie 5.2. *Przy założeniach Z1–Z5 i H_0 statystka*

$$F = \frac{(rB - \tilde{r})^T (r(X^T X)^{-1} r^T)^{-1} (rB - \tilde{r})}{mS^2}$$

ma rozklad F-Snedecora $F(m, n-K)$ (rozkład F z m i $n-K$ stopniami swobody).

Uwaga 5.3. Jeśli X_1 i X_2 są niezależnymi zmiennymi losowymi o rozkładzie χ^2 o odpowiednio m_1 i m_2 stopniach swobody to zmienna losowa

$$F = \frac{X_1}{X_2} \frac{m_2}{m_1}$$

ma rozkład $F(m_1, m_2)$ ([12] s.44-46).

Dowód twierdzenia.

$$F = \frac{(rB - \tilde{r})^T (r(X^T X)^{-1} r^T)^{-1} (rB - \tilde{r})}{mS^2}.$$

Dzielimy licznik i mianownik przez σ^2 i podstawiamy $S^2 = \frac{\xi^T \xi}{n-K}$. Otrzymujemy

$$F = \frac{(rB - \tilde{r})^T (\sigma^2 r(X^T X)^{-1} r^T)^{-1} (rB - \tilde{r})/m}{\frac{\xi^T \xi}{\sigma^2 (n-K)}} = \frac{w/m}{q/(n-K)},$$

gdzie

$$w = (rB - \tilde{r})^T (\sigma^2 r(X^T X)^{-1} r^T)^{-1} (rB - \tilde{r}), \quad q = \frac{\xi^T \xi}{\sigma^2}.$$

Jak pokazaliśmy w lemacie 5.2 $q|X \sim \chi^2(n-K)$.

Lemat 5.3. *Przy założeniach twierdzenia 5.2:*
1. $w|X \sim \chi^2(m)$.
2. q i w są warunkowo względem X niezależne.

Dowód.
Ad.1. Przyjmijmy oznaczenie $v = rB - \tilde{r}$. Z H_0 wynika, że $\tilde{r} = r\beta$, zatem

$$v = rB - \tilde{r} = r(B - \beta).$$

Ponieważ warunkowy rozkład $B - \beta$ względem X jest normalny (lemat 5.1)to

$$v|X \sim N(0, \sigma^2 r(X^T X)^{-1} r^T).$$

Rzeczywiście

$$Var(v|X) = Var(r(B - \beta)|X) = rVar((B - \beta)|X)r^T = \sigma^2 r(X^T X)^{-1} r^T.$$

A więc

$$w = v^T Var(v|X)^{-1} v, \quad \text{i} \quad w|X \sim \chi^2(m).$$

B i ξ są warunkowo względem X niezależne. Ponieważ w zależy od B i X a q od ξ i X to również one są warunkowo względem X niezależne.

□

Cd. dowodu twierdzenia.
Z lematu wynika, że statystyka F ma warunkowy względem X rozkład F-Snedecora $F(m, n-K)$. Ponieważ rozkład warunkowy nie zależy od warunkowania to X ma "bezwarunkowy" rozkład $F(m, n-K)$.

□

Reguła decyzyjna testu F.
Przedstawimy dwa równoważne warianty reguły decyzyjnej dla zadanego poziomu istotności α.

Wariant 1.
1. Na podstawie próbki ω wyznaczamy realizację statystyki testowej $f = F(\omega)$.
2. Wyznaczamy wartość krytyczną f^*_α

$$P(X > f^*_\alpha) = \alpha, \quad X \sim F(m, n-K).$$

3. Jeżeli $f < f^*_\alpha$ to nie ma podstaw do odrzucenia hipotezy H_0 (akceptujemy H_0).
Jeżeli $f \geqslant f^*_\alpha$ to odrzucamy H_0 na rzecz H_1.

Wariant 2.
1. Na podstawie próbki ω wyznaczamy realizację statystyki testowej $f = F(\omega)$.
2. Wyznaczamy prawdopodobieństwo (tzw. p-value)

$$p = P(X \geqslant f), \quad \text{dla} \quad X \sim F(m, n-K).$$

3. Jeżeli $p > \alpha$ to nie ma podstaw do odrzucenia hipotezy H_0 (akceptujemy H_0).
Jeżeli $p \leqslant \alpha$ to odrzucamy H_0 na rzecz H_1.

Statystyka F w terminach sumy kwadratów reszt.
Statystykę F mozna wyrazić w prostszy sposób wykorzystując sumę kwadratów reszt modelu ograniczonego

$$SKR_o = \min\{SKR(B) : \ rB = \tilde{r}\} = \min\{\xi^T\xi : \ \xi = Y - XB_0, \ rB_0 = \tilde{r}\}.$$

Lemat 5.4. *Przy założeniach twierdzenia 5.2:*

$$F = \frac{SKR_o - SKR}{SKR}\frac{n-K}{m}.$$

Dowód.
Krok 1. Pokażemy, że estymator OMNK (metody najmniejszych kwadratów z ograniczeniami) wynosi

$$B_o = B - (X^TX)^{-1}r^T(r(X^TX)^{-1}r^T)^{-1}(rB - \tilde{r}).$$

Rozważamy funkcję Lagrange'a

$$L(B_o, \lambda) = \frac{1}{2}(Y - XB_o)^T(Y - XB_o) + \lambda^T(rB_o - \tilde{r}),$$

gdzie λ jest m-elementowym wektorem wierszowym.
Różniczkujemy L po współrzędnych B_o.

$$\frac{\partial L}{\partial B_{o,i}} = -(Y - XB_o)^TXe_i + \lambda^Tre_i = (\lambda^Tr - Y^TX + B_o^TX^TX)e_i,$$

gdzie e_i n-elementowym wektorem kolumnowym o współrzędnych 0 i 1

$$e_{i,j} = \begin{cases} 1 & \text{gdy} \quad j = i, \\ 0 & \text{gdy} \quad j \neq i. \end{cases}$$

Ponieważ wszystkie pochodne cząstkowe zerują się w punktach, w których funkcja przyjmuje minimum to

$$B_o^T = (-\lambda^T r + Y^T X)(X^T X)^{-1}.$$

Czyli po transpozycji mamy

$$B_o = (X^T X)^{-1}(X^T Y - r^T \lambda) = B - (X^T X)^{-1} r^T \lambda.$$

Po przemnożeniu przez macierz r otrzymujemy

$$\tilde{r} = rB_o = rB - r(X^T X)^{-1} r^T \lambda.$$

Ponieważ rząd $m \times n$ macierzy r wynosi m, a macierz $(X^T X)^{-1}$ jest prawie na pewno dodatnio określona, to macierz $r(X^T X)^{-1} r^T$ jest prawie na pewno odwracalna. Zatem

$$\lambda = (r(X^T X)^{-1} r^T)^{-1}(rB - \tilde{r})).$$

Czyli

$$B_0 = B - (X^T X)^{-1} r^T (r(X^T X)^{-1} r^T)^{-1}(rB - \tilde{r})).$$

Krok 2. Pokażemy, że $SKR_o - SKR = \sigma^2 w$, gdzie w takie jak w lemacie 5.3

$$w = (rB - \tilde{r})^T (\sigma^2 r(X^T X)^{-1} r^T)^{-1}(rB - \tilde{r}).$$

Mamy

$$SKR_o - SKR = \|Y - XB_o\|^2 - \|Y - XB\|^2 = \|Y - XB + X(B - B_o)\|^2 - \|Y - XB\|^2.$$

$$= 2(Y - XB)^T X(B - B_o) + \|X(B - B_o)\|^2.$$

Ponieważ $\xi = Y - XB$ jest ortogonalne do wszystkich kolumn macierzy X (wniosek 2.1) to

$$SKR_o - SKR = \|X(B - B_o)\|^2 = (B - B_o)^T X^T X(B - B_o) =$$

$$= (rB - \tilde{r})^T (r(X^T X)^{-1} r^T)^{-1} r(X^T X)^{-1} X^T X(X^T X)^{-1} r^T (r(X^T X)^{-1} r^T)^{-1}(rB - \tilde{r}) =$$

$$= (rB - \tilde{r})^T (r(X^T X)^{-1} r^T)^{-1}(rB - \tilde{r}) = \sigma^2 w.$$

Krok 3.

$$F = \frac{\sigma^2 w}{\sigma^2 q} \frac{n - K}{m}.$$

Ponieważ $\sigma^2 q = \xi^T \xi = SKR$, a $\sigma^2 w = SKR_o - SKR$ to otrzymujemy

$$F = \frac{SKR_o - SKR}{SKR} \frac{n - K}{m}.$$

□

Sprowadzanie modelu ograniczonego do modelu z mniejszą liczbą parametrów.

Rozwiązanie ogólne układu równań liniowych $r\beta = \tilde{r}$ mozna zapisać w postaci parametrycznej:

$$\beta = a_0 + a_1\gamma,$$

gdzie a_0 jest wektorem kolumnowym $K \times 1$, a_1 jest macierzą $K \times (K-m)$, a $(K-m)$ wektor kolumnowy γ jest wektorem nieznanych parametrów, które należy wyestymować. Zauważmy, że

$$ra_0 = \tilde{r}, \quad ra_1 = 0.$$

Model z ograniczeniami można zapisać w następujący sposób:

$$Y = X(a_0 + a_1\gamma) + \varepsilon.$$

Po podstawieniu $Y_o = Y - Xa_0$ i $X_o = Xa_1$ otrzymujemy równoważny mu model zredukowany

$$Y_o = X_o\gamma + \varepsilon.$$

Niech g będzie estymatorem MNK γ dla modelu zredukowanego. Wówczas $B_o = a_0 + a_1 g$ jest estymatorem β dla modelu z ograniczeniami. Zauważmy, że w obu wypadkach mamy ten sam składnik resztowy ξ_0.

$$\xi_0 = Y_o - X_o g = Y - Xa_0 - Xa_1 g = Y - XB_o.$$

Test istotności regresji dla regresji z wyrazem wolnym.

W przypadku gdy ostatni parametr jest wyrazem wolnym, czyli gdy $X_K = e$, stosuje się często następujący wariant testu liniowości:

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_{K-1} = 0, \quad H_1 : \exists i < K \;\; \beta_i \neq 0.$$

W tym przypadku r jest $(K-1) \times K$ wymiarową macierzą o wyrazach

$$r_{i,j} = \begin{cases} 1 & \text{gdy } i = j, \\ 0 & \text{gdy } i \neq j, \end{cases}$$

a $\tilde{r} = 0$.

Statystyka F wynosi wtedy

$$F = \frac{SKR_o - SKR}{SKR}\frac{n-K}{K-1} = \frac{\|Y - \overline{Y}e\|^2 - \|Y - \widehat{Y}\|^2}{\|Y - \widehat{Y}\|^2}\frac{n-K}{K-1} =$$

$$= \frac{\|\widehat{Y} - \overline{Y}e\|^2}{\|Y - \widehat{Y}\|^2}\frac{n-K}{K-1} = \frac{\sum_{t=1}^n (\widehat{Y}_t - \overline{Y})^2}{\sum_{t=1}^n (Y_t - \widehat{Y}_t)^2}\frac{n-K}{K-1}.$$

Uwaga 5.4. F można wyrazić za pomocą współczynnika determinacji R^2

$$F = \frac{R^2}{1-R^2}\frac{n-K}{K-1}.$$

Dowód.

$$R^2 = 1 - \frac{\sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2}{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}.$$

Dlatego też

$$\frac{R^2}{1-R^2}\,\frac{n-K}{K-1} = \left(\frac{1}{1-R^2} - 1\right)\frac{n-K}{K-1} =$$

$$= \left(\frac{\sum_{t=1}^{n}(Y_t - \overline{Y})^2}{\sum_{t=1}^{n}(Y_t - \widehat{Y}_t)^2} - 1\right)\frac{n-K}{K-1} = F.$$

□

6. Modele produkcji - funkcja Cobba-Douglasa

Klasyczny jednorównaniowy liniowy model ekonometryczny -cd. Przykład: Modele produkcji - funkcja Cobba-Douglasa. (1 wykład)

6.1. Funkcja Cobba-Douglasa

6.1.1. Wprowadzenie

Funkcja Cobba-Douglasa to funkcyjne przedstawienie zależności wielkości produkcji Q od nakładów na czynniki produkcji. W dalszym ciągu ograniczymy sie do trzech czynników pracy x_1, kapitału x_2 i paliwa x_3.

$$Q = Ax_1^{\alpha_1}x_2^{\alpha_2}x_3^{\alpha_3}, \quad 0 < \alpha_i < 1, \quad x_i > 0.$$

Współczynnik A zależy od efektywności konkretnej firmy.

Funkcja Cobba-Douglasa jest chętnie wykorzystywana w modelowaniu, gdyż dobrze przedstawia następujące fakty stylizowane:
• monotoniczność;
Q jest rosnąca ze względu na każdy x_i,

$$\frac{\partial Q}{\partial x_i} = \alpha_i \frac{Q}{x_i} > 0.$$

• wklęsłość;
Q jest wklęsła ze względu na każdy x_i,

$$\frac{\partial^2 Q}{\partial x_i^2} = \alpha_i(\alpha_i - 1)\frac{Q}{x_i^2} < 0.$$

Funkcja zachowuje zasadę malejących przychodów – każda kolejna jednostka jednego z zasobów bez wzrostu zasobu drugiego skutkuje mniejszym przyrostem produkcji.
• wzrost przychodów przy zwiększaniu nakładów na dwa czynniki produkcji;

$$\frac{\partial^2 Q}{\partial x_i \partial x_j} = \alpha_i \alpha_j \frac{Q}{x_i x_j} > 0 \quad \text{dla} \quad i \neq j.$$

• stała elastyczność ze względu na każdy czynnik produkcji;

$$E_{x_i}Q = \frac{x_i \frac{\partial Q}{\partial x_i}}{Q} = \frac{x_i \alpha_i \frac{Q}{x_i}}{Q} = \alpha_i.$$

Uwaga 6.1. Elastyczność mówi nam o ile wzrośnie produkcja gdy zwiększymy nakłady na czynnik produkcji

$$\frac{Q((1+h)x_i) - Q(x_i)}{Q(x_i)} \approx \frac{hx_i \frac{\partial Q}{\partial x_i}}{Q} = hE_{x_i}Q.$$

6.1.2. Efekt skali

Zmniejszamy albo zwiększamy proporcjonalnie wszystkie x_i

$$x_i' = hx_i, \quad h > 0, \quad i = 1, 2, 3.$$

Wówczas nowa wielkość produkcji wyniesie:

$$Q' = Q(x') = Ax_1^{\alpha_1}x_2^{\alpha_2}x_3^{\alpha_3}h^{\alpha_1+\alpha_2+\alpha_3}.$$

Czyli

$$\frac{Q'}{Q} = h^{\alpha_1+\alpha_2+\alpha_3}.$$

Zauważmy, że gdy $\alpha_1 + \alpha_2 + \alpha_3 > 1$ to

$$h > 1 \Rightarrow \frac{Q'}{Q} > h,$$

$$h < 1 \Rightarrow \frac{Q'}{Q} < h.$$

Wniosek: opłaca się zwiększyć nakłady i produkcję.

Gdy $\alpha_1 + \alpha_2 + \alpha_3 < 1$ to

$$h > 1 \Rightarrow \frac{Q'}{Q} < h,$$

$$h < 1 \Rightarrow \frac{Q'}{Q} > h.$$

Wniosek: opłaca się zmniejszyć nakłady i produkcję.

Podsumowując, jeśli obserwujemy „stan równowagi” to $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Mówimy wówczas o braku efektów skali.

6.1.3. Koszty produkcji

Koszty całkowite produkcji TC można wyrazić za pomocą kosztów jednostkowych dla poszczególnych czynników produkcji

$$TC = p_1x_1 + p_2x_2 + p_3x_3.$$

Zadanie: Zminimalizować koszty dla zadanego poziomu produkcji Q, $Q > 0$.

$$TC(x) \longrightarrow min, \;\; Q(x) = Q.$$

Lemat 6.1. *Powyższe zadanie optymalizacyjne posiada dokladnie jedno rozwiązanie.*

$$TC_{min} = \frac{r}{(A\alpha_1^{\alpha_1}\alpha_2^{\alpha_2}\alpha_3^{\alpha_3})^{\frac{1}{r}}} Q^{\frac{1}{r}} p_1^{\frac{\alpha_1}{r}} p_2^{\frac{\alpha_2}{r}} p_3^{\frac{\alpha_3}{r}},$$

gdzie $r = \alpha_1 + \alpha_2 + \alpha_3$.

Dowód. Połóżmy,

$$\widetilde{x}_i = Q^{\frac{1}{r}} A^{-\frac{1}{r}}, \quad i = 1, 2, 3.$$

Jak łatwo zauważyć

$$Q(\widetilde{x}) = Q.$$

Połóżmy

$$T = TC(\widetilde{x}).$$

Ponieważ zbiór

$$M = \{x \in \mathbb{R}^3 : x_i \geqslant 0, \;\; TC(x) \leqslant T, \quad Q(x) = Q\},$$

jest niepusty, domknięty i ograniczony, zatem badane zadanie optymalizacyjne posiada rozwiązanie.

Rozważmy warunek konieczny istnienia lokalnego ekstremum – $\nabla Q \parallel \nabla TC$.

$$\nabla Q = Q(x) \left(\frac{\alpha_1}{x_1}, \frac{\alpha_2}{x_2}, \frac{\alpha_3}{x_3} \right), \quad \nabla TC = (p_1, p_2, p_3).$$

Równoległość gradientów implikuje istnienie stałej λ takiej, że

$$\frac{\alpha_1}{p_1 x_1} = \frac{\alpha_2}{p_2 x_2} = \frac{\alpha_3}{p_3 x_3} = \lambda.$$

A zatem

$$x_i = \frac{\alpha_i}{\lambda p_i}, \quad i = 1, 2, 3.$$

Po podstawieniu do warunku $Q(x) = Q$ otrzymujemy

$$Q = A \left(\frac{\alpha_1}{\lambda p_1} \right)^{\alpha_1} \left(\frac{\alpha_2}{\lambda p_2} \right)^{\alpha_2} \left(\frac{\alpha_3}{\lambda p_3} \right)^{\alpha_3} = \lambda^{-r} A \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3} p_1^{-\alpha_1} p_2^{-\alpha_2} p_3^{-\alpha_3},$$

gdzie $r = \alpha_1 + \alpha_2 + \alpha_3$. Wyznaczamy λ

$$\lambda = Q^{-\frac{1}{r}} \left(A \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3} \right)^{\frac{1}{r}} p_1^{-\frac{\alpha_1}{r}} p_2^{-\frac{\alpha_2}{r}} p_3^{-\frac{\alpha_3}{r}}.$$

Teraz możemy wyznaczyć TC_{min}

$$TC_{min} = TC \left(\frac{\alpha_1}{\lambda p_1}, \frac{\alpha_2}{\lambda p_2}, \frac{\alpha_3}{\lambda p_3} \right) = p_1 \frac{\alpha_1}{\lambda p_1} + p_2 \frac{\alpha_2}{\lambda p_2} + p_3 \frac{\alpha_3}{\lambda p_3} = \frac{r}{\lambda} = \frac{r}{\left(A \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3} \right)^{\frac{1}{r}}} Q^{\frac{1}{r}} p_1^{\frac{\alpha_1}{r}} p_2^{\frac{\alpha_2}{r}} p_3^{\frac{\alpha_3}{r}}.$$

□

6.2. Przykład Nerlove'a

6.2.1. Charakterystyka danych

M.Nerlove przeprowadził badania dotyczące produkcji energii elektrycznej w USA w 1955 roku. Dane zostały zebrane dla 145 spółek w 44 stanach. Dotyczą one:

- całkowitych kosztów TC (mln USD),
- wielkości produkcji Q (mld kWh),
- średnich zarobków (koszt pracy) $PL = p_1$,
- ceny kapitału (stopy procentowe) $PK = p_2$,
- ceny paliwa $PF = p_3$.

Warunki działania spółek:
• dostawa energii zgodnie z zapotrzebowaniem,
• cena energii ustalana administracyjnie dla regionu,
• firmy nie mają bezpośredniego wpływu na PL, PK i PF. PF i PK kształtuje rynek, a PL długoterminowe umowy ze związkami zawodowymi.

6.2.2. Konstrukcja modelu

Model ekonometryczny:

$$TC_i = e^{\mu_i} Q_i^{\beta_2} p_{i,1}^{\beta_3} p_{i,2}^{\beta_4} p_{i,3}^{\beta_5}, \quad \beta_2 = \frac{1}{r}, \quad \beta_{2+j} = \frac{\alpha_j}{r}, \quad j = 1, 2, 3. \tag{6.1}$$

$$\mu_i = \ln\left(r\left(A_i \alpha_1^{\alpha_1} \alpha_2^{\alpha_2} \alpha_3^{\alpha_3}\right)^{-\frac{1}{r}}\right).$$

μ_i zawiera część losową zależną od firmy,

$$E(\mu_i) = \beta_1, \quad \mu_i = \beta_1 + \varepsilon_i.$$

Logarytmujemy równanie 6.1 i przechodzimy do modelu liniowego

$$\ln TC_i = \beta_1 + \beta_2 \ln Q_i + \beta_3 \ln p_{i,1} + \beta_4 \ln p_{i,2} + \beta_5 \ln p_{i,3} + \varepsilon_i. \tag{6.2}$$

Dodatkowo rozważamy model ograniczony, w którym spełniona jest zależność $\beta_3 + \beta_4 + \beta_5 = 1$. Podstawiamy $\beta_5 = 1 - \beta_3 - \beta_4$ i otrzymujemy

$$\ln\left(\frac{TC_i}{p_{i,3}}\right) = \beta_1 + \beta_2 \ln Q_i + \beta_3 \ln \frac{p_{i,1}}{p_{i,3}} + \beta_4 \ln \frac{p_{i,2}}{p_{i,3}} + \varepsilon_i. \tag{6.3}$$

6.2.3. Estymacja parametrów modelu 6.2

Estymujemy parametry strukturalne metodą MNK. Otrzymujemy następujące równanie regresji. W nawiasach podane są odchylenia standardowe estymatorów.

$$\begin{array}{lccccc} \ln TC = & -3{,}5 & +0{,}72 \ln Q & +0{,}44 \ln p_1 & -0{,}22 \ln p_2 & +0{,}43 \ln p_3 \\ & (1{,}8) & (0{,}017) & (0{,}29) & (0{,}34) & (0{,}10) \end{array}$$

Ponadto suma kwadratów składnika resztowego wyniosła $SKR = 21{,}552$.
Uwaga: Test t wskazuje na statystyczną nieistotność parametru β_4.

6.2.4. Estymacja parametrów modelu 6.3

Estymujemy parametry strukturalne metodą MNK. Otrzymujemy następujące równanie regresji. W nawiasach podane są odchylenia standardowe estymatorów.

$$\begin{array}{lcccc} \ln \frac{TC}{p_3} = & -4{,}7 & +0{,}72 \ln Q & +0{,}59 \ln \frac{p_1}{p_3} & -0{,}007 \ln \frac{p_2}{p_3} \\ & (0{,}88) & (0{,}017) & (0{,}20) & (0{,}19) \end{array}$$

Ponadto suma kwadratów składnika resztowego wyniosła $SKR_o = 21{,}640$.
Uwaga: Test t wskazuje na statystyczną nieistotność parametru β_4.

6.2.5. Test jednorodności modelu

Testujemy hipotezę $H_0 : \beta_3 + \beta_4 + \beta_5 = 1$ wobec $H_1 : \beta_3 + \beta_4 + \beta_5 \neq 1$ na poziomie istotności $\alpha = 0{,}05$.
Mamy $m = 1$, $n - K = 145 - 5 = 140$ stopni swobody. Wyznaczamy statystykę F.

$$F = \frac{SKR_o - SKR}{SKR} \cdot \frac{n - K}{m} = 0{,}57.$$

Wartość krytyczną F^* wyznaczamy z rozkładu Snedecora $F(1, 140)$ otrzymujemy

$$F^* = 3{,}9 >> F.$$

Zatem nie ma podstaw do odrzucenia hipotezy H_0.

6.2.6. Test braku efektów skali dla modelu ograniczonego 6.3

Testujemy hipotezę $H_0 : \beta_2 = 1$ wobec $H_1 : \beta_2 \neq 1$ na poziomie istotności $\alpha = 0{,}05$.
Mamy $n - K = 145 - 4 = 141$ stopni swobody. Wyznaczamy statystykę t

$$t = \frac{b_2 - 1}{sb_2} = \frac{0,72 - 1}{0,017} = -16.$$

Wartość krytyczną t^* wyznaczamy z rozkładu Studenta $t(141)$ otrzymujemy

$$t^* = 1{,}98 << |t|.$$

Zatem odrzucamy hipotezę H_0.

7. Modele nieliniowe

Zastosowanie metody najmniejszych kwadratów w modelach nieliniowych. Opis metody. Przykłady: Modelowanie popytu konsumpcyjnego - funkcje Tórnquista. (1 wykład)

7.1. Zadanie aproksymacyjne

Dane jest $m+1$ n-elementowych ciągów

$$Y = (Y_t), \quad X_i = (X_{t,i}), \quad i = 1, \ldots, m, \quad t = 1, \ldots, n,$$

oraz dana jest rodzina funkcji $F(\eta_1, \ldots, \eta_K, x_1, \ldots, x_m) : D \times \mathbb{R}^m \longrightarrow \mathbb{R}$, $D \subset \mathbb{R}^K$.

Zadanie:
Wyznaczyć współczynniki $b = (b_1, \ldots, b_K)$, $b \in D \subset \mathbb{R}^K$, które minimalizują błąd przybliżenia Y przez $\widehat{Y}$

$$\widehat{Y}_t = F_t(b) = F(b_1, \ldots b_K, X_{t,1}, \ldots, X_{t,m}), \quad t = 1, \ldots, n.$$

Czyli należy rozwiązać zadanie optymalizacyjne:

$$\|\xi\|^2 = \sum_{t=1}^{n} \xi_t^2 \longrightarrow \min, \quad \text{gdzie } \xi = Y - \widehat{Y}.$$

W ogólnym przypadku, gdy nie znamy własności funkcji F, to nie wiemy czy powyższe zadanie posiada rozwiązanie i czy jeśli posiada rozwiązanie to jest ono jedynym rozwiązaniem. Gdy F jest różniczkowalne to można sformułować warunki konieczne, takie jak na przykład poniższy.

Lemat 7.1. *Jeżeli F jest różniczkowalne ze względu na b i b_{min} należący do wnętrza zbioru D minimalizuje $\|\xi\|^2$ to wektor reszt $\xi = \xi(b_{min})$ jest ortogonalny do wektora pochodnych*

$$\sum_{t=1}^{n} \xi_t \frac{\partial F_t}{\partial b_i}(b_{min}) = 0 \quad dla \quad i = 1, \ldots, K.$$

Dowód.
Przy założeniach lematu suma kwadratów reszt $SKR(b) = \|\xi(b)\|^2$ jest różniczkowalna. Zatem w punkcie, w którym przyjmuje najmniejszą wartość jej pochodne cząstkowe muszą się zerować. Biorąc pod uwagę, że $\xi_t(b) = Y_t - F_t(b)$ otrzymujemy dla każdego i

$$0 = \frac{\partial SKR}{\partial b_i}(b_{min}) = 2\sum_{t=1}^{n} \xi_t \frac{\partial \xi_t}{\partial b_i} = -2\sum_{t=1}^{n} \xi_t \frac{\partial F_t}{\partial b_i}(b_{min}).$$

□

7.2. Założenia modelu i estymacja parametrów

Przyjmijmy dla modelu nieliniowego założenia wzorowane na modelu liniowym.
NZ1. Liniowa zależność od czynnika losowego

$$Y_t = F(\beta, X_t) + \varepsilon_t,$$

gdzie $\beta = (\beta_1, \ldots, \beta_K)^T$ wektor parametrów, a F funkcja różniczkowalna po β.
NZ2. Ścisła egzogeniczność.

$$E(\epsilon|X) = 0.$$

NZ3. Liniowa niezależność pochodnych.

$$\forall b \in D \quad (P(rkDF_*(b) = K) = 1,$$

gdzie DF_* oznacza macierz $n \times K$ pochodnych cząstkowych wektora $F_* = (F_1, \ldots, F_n)^T$, $F_t(b) = F(b, X_t)$,

$$DF_*(b)_{t,i} = \frac{\partial F_t}{\partial b_i}(b) = \frac{\partial F}{\partial b_i}(b, X_t).$$

NZ4. Sferyczność błędu

$$E(\varepsilon\varepsilon^T|X) = \sigma^2 Id_n,$$

gdzie $\sigma > 0$ deterministyczny parametr modelu.

Naturalne jest aby estymatorem NMNK wektora β nazywać rozwiązanie zadania aproksymacyjnego z poprzedniego podrozdziału.

Definicja 7.1. K-wymiarowa zmienna losowa B jest estymatorem NMNK wektora $\beta \in D$, gdy B przyjmuje wartości w D i prawie na pewno

$$\sum_{t=1}^{n}(Y_t(\omega) - F(B(\omega), X_t(\omega))^2 = \min\{\sum_{t=1}^{n}(Y_t(\omega) - F(b, X_t(\omega))^2 : b \in D\}.$$

Lemat 7.2. *Niech D będzie podzbiorem otwartym $\mathbb{R}^K$. Jeśli B jest estymatorem NMNK wektora β to*

$$B - \beta = (DF_*(B)^T DF_*(B))^{-1} DF_*(B)^T(\varepsilon + R_2(B)),$$

gdzie $R_2(B)$ wektor reszt rozwinięcia Taylora F_t w B

$$R_2(B) = F_*(B) - F_*(\beta) + DF_*(B)(\beta - B).$$

Dowód.
Mamy dwie równości wynikające z NZ1 i definicji składnika resztowego ξ

$$\begin{aligned} Y &= F_*(\beta) + \varepsilon, \\ Y &= F_*(B) + \xi. \end{aligned}$$

Zatem

$$F_*(B) - F_*(\beta) = \varepsilon - \xi.$$

Czyli

$$DF_*(B)(B - \beta) = \varepsilon - \xi + R_2(B).$$

Z NZ3 wynika, że macierz $DF_*(B)^T DF_*(B)$ jest prawie na pewno odwracalna, zatem macierz $K \times n$

$$P(B) = (DF_*(B)^T DF_*(B))^{-1} DF_*(B)^T$$

jest dobrze zdefiniowana poza, być może, zbiorem P-miary zero.
Zauważmy, że z definicji P otrzymujemy

$$P(B)DF_*(B) = (DF_*(B)^T DF_*(B))^{-1} DF_*(B)^T DF_*(B) = Id_K, \quad \text{pn.},$$

a z lematu 7.1

$$P(B)\xi = (DF_*(B)^T DF_*(B))^{-1} DF_*(B)^T \xi = 0.$$

Zatem

$$B - \beta = P(B)DF_*(B)(B - \beta) = P(B)(\varepsilon - \xi + R_2(B)) = P(B)(\varepsilon + R_2(B)).$$

□

Uwaga 7.1. Jak widać z powyższego warunek ścisłej egzogeniczności (NZ2) nie implikuje nieobciążoności estymatora NMNK.

7.3. Przykłady

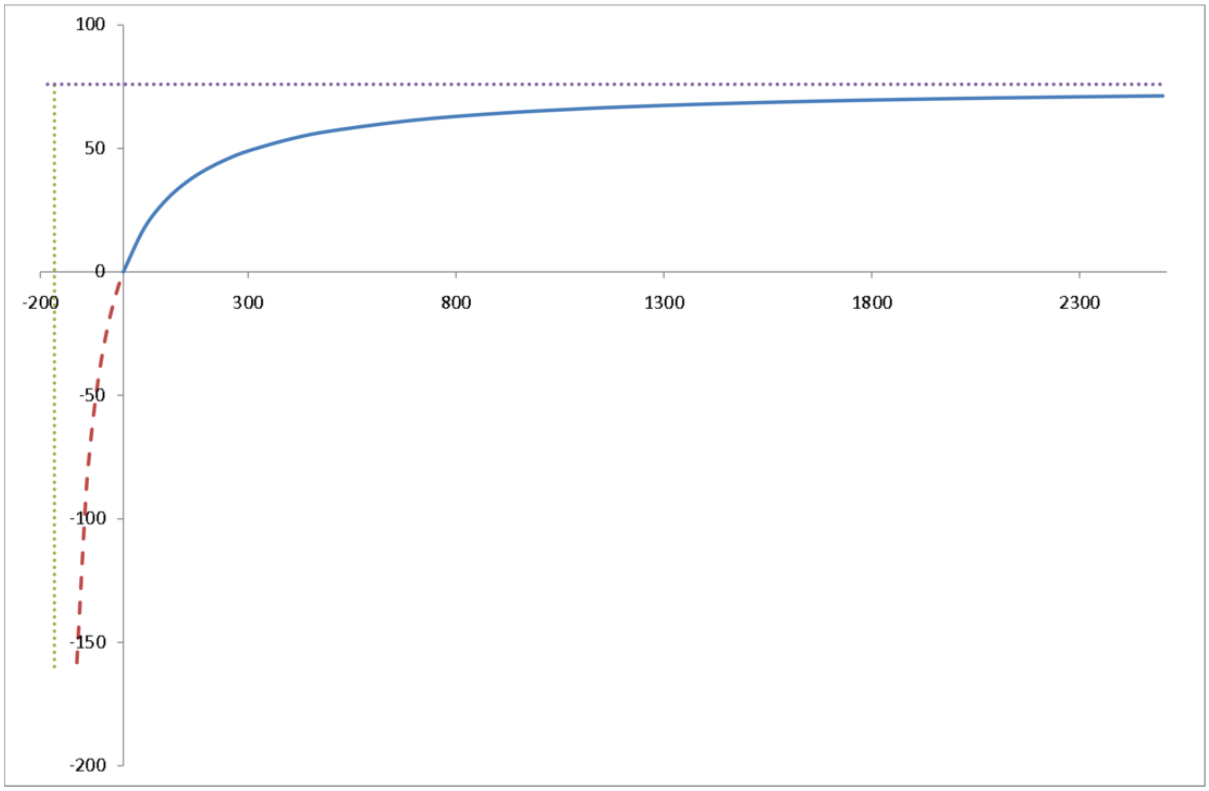
Opisane w poprzednich podrozdziałach trudności związane ze stosowaniem nieliniowej metody najmniejszych kwadratów powodują, że nieliniowe modele stosuje się tylko wtedy gdy dobrze opisują *fakty stylizowane* związane z modelowanym zjawiskiem. Tak jak na przykład funkcje Törnquista, których przebieg ilustruje hipotezy dotyczące reakcji konsumentów na zmiany dochodów, albo funkcja logistyczna, która znalazła zastosowanie w modelowaniu długookresowego wzrostu liczby ludności i w reprezentacji rozwoju sprzedaży nowych produktów na określonym rynku.

7.3.1. Funkcja Törnquista I typu

Funkcja Törnquista I typu wyraża zależność między popytem na dobra podstawowe (Y), a dochodami konsumentów (X)

$$Y = \frac{\alpha X}{X + \beta}, \quad \alpha, \beta > 0.$$

Popyt rośnie wraz ze wzrostem dochodów i stabilizuje się na poziomie równym α. Wykres ma asymptotę poziomą. Z drugiej strony popyt spada do zera wraz ze spadkiem dochodów ale nie znika dla niezerowych dochodów.



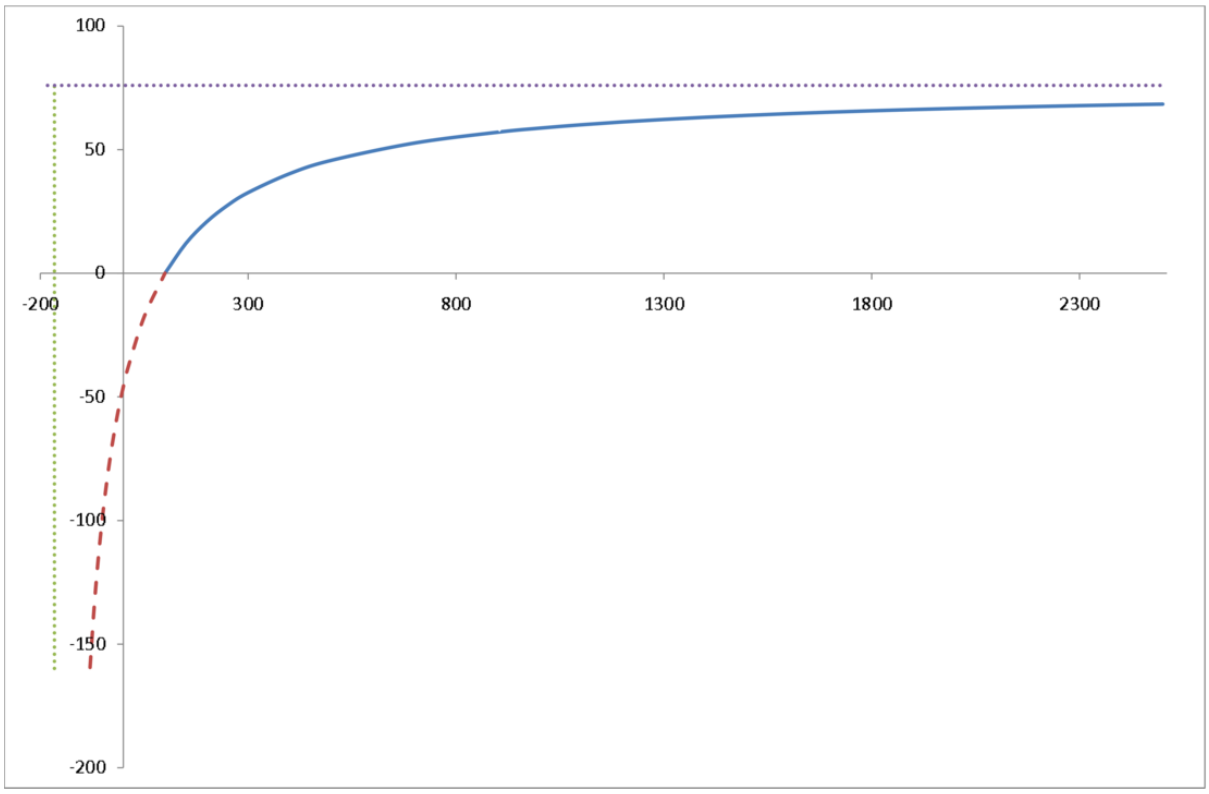
Rysunek 7.1. Wykres funkcji Törnquista I typu.

7.3.2. Funkcja Törnquista II typu

Funkcja Törnquista II typu wyraża zależność między popytem na dobra wyższego rzędu (Y), a dochodami konsumentów (X)

$$Y = \frac{\alpha(X-\gamma)}{X+\beta}, \quad \alpha, \beta, \gamma > 0.$$

Popyt rośnie wraz ze wzrostem dochodów i stabilizuje się na poziomie równym α. Wykres ma asymptotę poziomą. Z drugiej strony popyt spada do zera wraz ze spadkiem dochodów i zanika dla dochodów niższych od granicznej wartości γ.



Rysunek 7.2. Wykres funkcji Törnquista II typu.

7.3.3. Funkcja Törnquista III typu

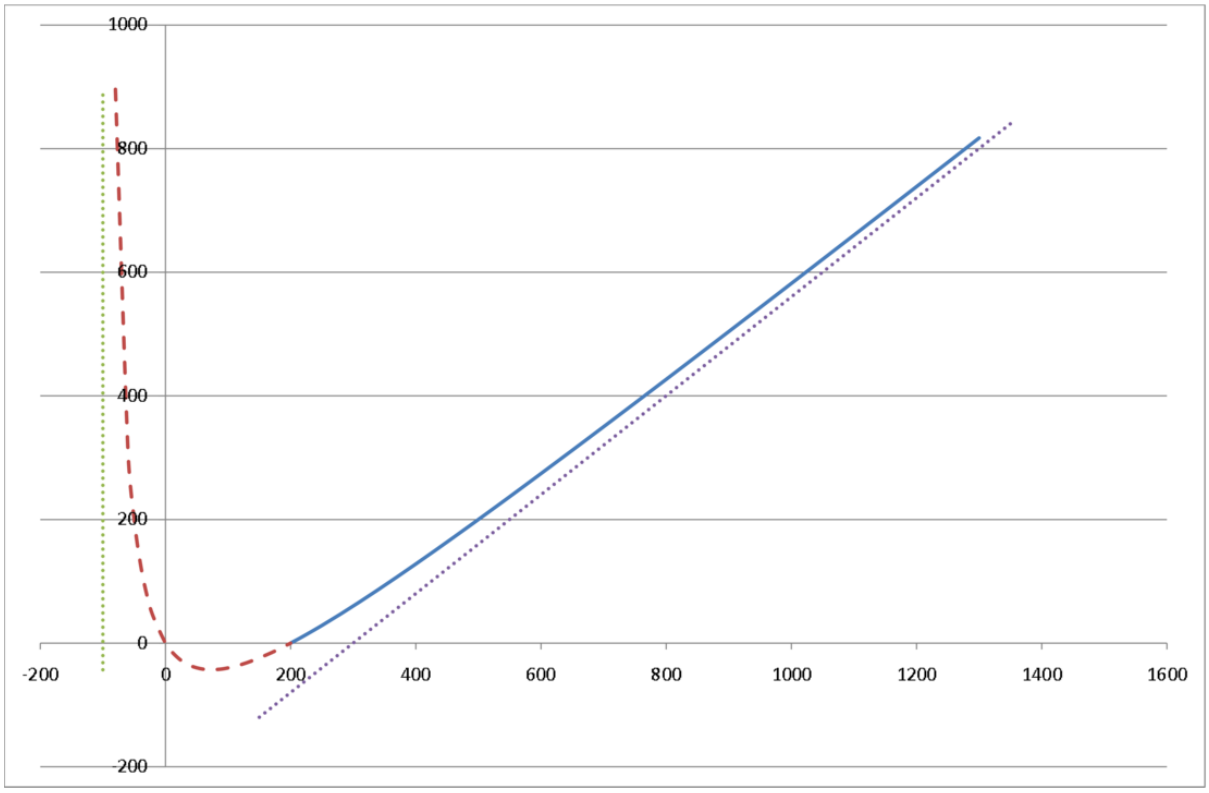
Funkcja Törnquista III typu wyraża zależność między popytem na dobra i usługi luksusowe (Y), a dochodami konsumentów (X)

$$Y = \frac{\alpha X(X-\gamma)}{X+\beta}, \quad \alpha, \beta, \gamma > 0.$$

Popyt rośnie w sposób nieograniczony wraz ze wzrostem dochodów i dla dużych X można go przybliżyć funkcją liniową

$$Y = \alpha(X - (\gamma + \beta)).$$

Wykres posiada asymptotę ukośną. Z drugiej strony popyt spada do zera wraz ze spadkiem dochodów i zanika dla dochodów niższych od granicznej wartości γ.



Rysunek 7.3. Wykres funkcji Törnquista III typu.

7.3.4. Funkcja logistyczna

Funkcja logistyczna wyraża zależność trendu wzrostowego Y od czasu t

$$Y(t) = \frac{\alpha}{1 + \beta \exp(-\gamma t)}, \quad \alpha, \beta, \gamma > 0.$$

Posiada ona następujące własności:
• Jest ściśle rosnąca i przyjmuje wartości z przedziału $(0, \alpha)$

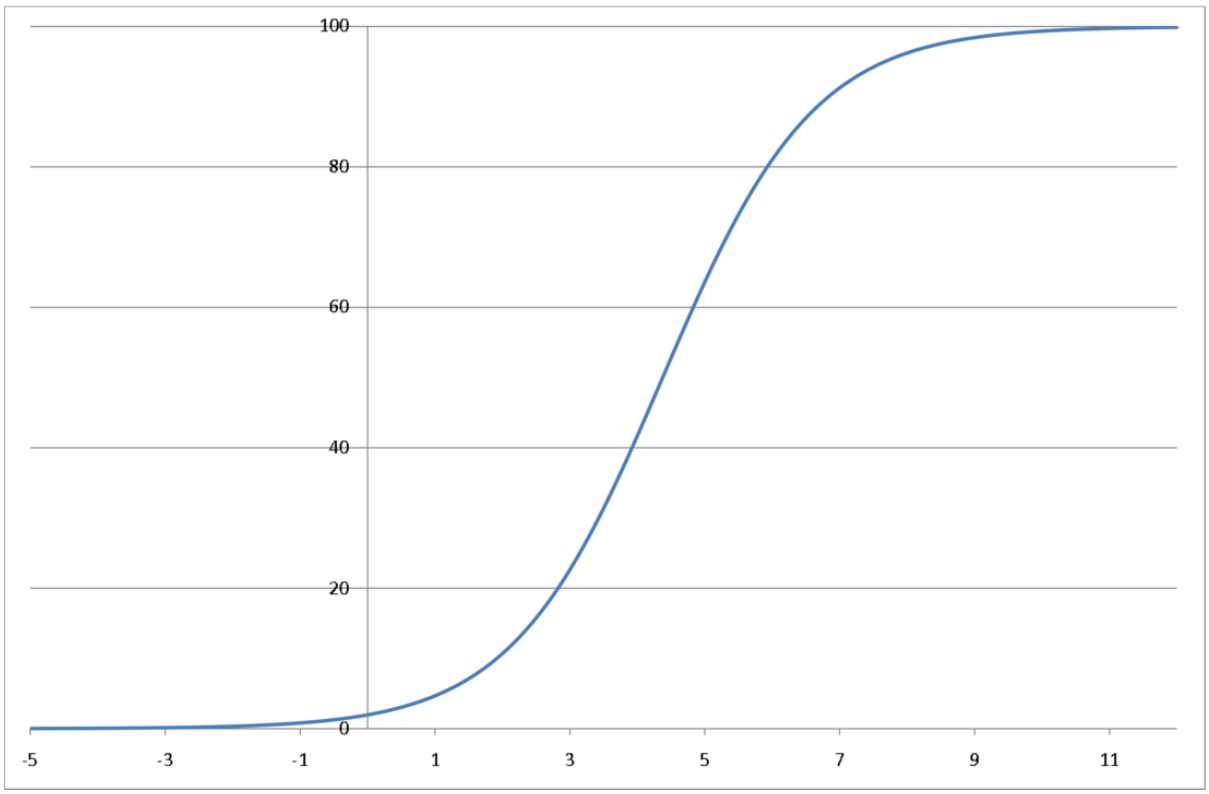
$$\lim_{t \to -\infty} Y(t) = 0, \qquad \lim_{t \to +\infty} Y(t) = \alpha.$$

• Dla $t = 0$ mamy $Y(0) = \frac{\alpha}{1+\beta}$.
• Ma punkt przegięcia w $t_* = \gamma^{-1} \ln \beta$. Trend ma w okolicach tego punktu największe przyrosty.
• Jest rozwiązaniem równania różniczkowego o zmiennych rozdzielonych zwanego równaniem Robertsona

$$Y' = \frac{\gamma}{\alpha} Y(\alpha - Y),$$

spełniającym warunek początkowy

$$Y(0) = \frac{\alpha}{1 + \beta}.$$



Rysunek 7.4. Wykres funkcji logistycznej.

8. Metody asymptotyczne

Zbieżność ciągów zmiennych losowych. Metoda delty. Asymptotyczna normalność estymatorów. Stacjonarność i ergodyczność. Ciągi przyrostów martyngałowych. Centralne Twierdzenie Graniczne dla przyrostów martyngałowych. (1 wykład)

8.1. Zbieżność zmiennych losowych

Omówimy pokrótce trzy pojęcia zbieżności zmiennych losowych.
Niech $(\Omega, \mathcal{M}, P)$ – przestrzeń probabilistyczna. Rozważmy ciąg zmiennych losowych $(Z_n)_{n=0}^{\infty}$ o wartościach w $\mathbb{R}^K$.

Definicja 8.1. Ciąg zmiennych losowych Z_n zbiega prawie napewno do zmiennej losowej Z gdy

$$P(\lim_{n\to\infty} Z_n = Z) = 1.$$

W zapisie skróconym będziemy pisali

$$Z_n \xrightarrow{as} Z \quad \text{lub} \quad Z_n \xrightarrow{pn} Z.$$

Definicja 8.2. Ciąg zmiennych losowych Z_n zbiega według prawdopodobieństwa do zmiennej losowej Z gdy

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P(\|Z_n - Z\| > \varepsilon) = 0,$$

gdzie $\|\cdot\|$ oznacza normę w $\mathbb{R}^K$.

W zapisie skróconym będziemy pisali

$$\operatorname*{plim}_{n\to\infty} Z_n = Z \quad \text{lub} \quad Z_n \xrightarrow{p} Z.$$

Definicja 8.3. Ciąg zmiennych losowych Z_n zbiega według rozkładu do zmiennej losowej Z gdy ciąg dystrybuant F_{Z_n} zbiega do dystrybuanty F_Z w każdym punkcie ciągłości F_Z.

W zapisie skróconym będziemy pisali

$$Z_n \xrightarrow{d} Z \quad \text{lub} \quad Z_n \xrightarrow{d} \mathcal{R},$$

gdzie R jest rozkładem zmiennej losowej Z.

Uwaga 8.1. Jeśli ciąg różnic $Z_n - Y_n$ zbiega do 0 według prawdopodobieństwa i ciąg Z_n zbiega według rozkładu do zmiennej losowej Z, to również Y_n zbiega według rozkładu do Z

$$Z_n \xrightarrow{d} Z, \quad Z_n - Y_n \xrightarrow{p} 0 \quad \Longrightarrow \quad Y_n \xrightarrow{d} Z.$$

Uwaga 8.2. Trzy powyższe zbieżności są od siebie zależne. Mamy

$$Z_n \xrightarrow{as} Z \quad \Longrightarrow \quad Z_n \xrightarrow{p} Z \quad \Longrightarrow \quad Z_n \xrightarrow{d} Z.$$

Aby zamknąć powyższy diagram należy dopuścić zmianę przestrzeni probabilistycznej i skorzystać z następującego twierdzenia o reprezentacji prawie napewno.

Twierdzenie 8.1. *(Skorochod)*
Jeśli ciąg zmiennych losowych Z_n zbiega według rozkładu do zmiennej losowej Z to istnieje przestrzeń probabilistyczna $(\Omega', \mathcal{M}', P')$ i określone na niej zmienne losowe Y i Y_n, $n = 0, 1, \ldots$ takie, że Y ma ten sam rozkład co Z, a Y_n co Z_n i ciąg zmiennych losowych Y_n zbiega prawie napewno do zmiennej losowej Y.

Sformułujemy teraz kilka przydatnych w ekonometrii twierdzeń o zbieżności zmiennych losowych.

Twierdzenie 8.2. *O odwzorowaniu ciągłym.[2, 16]*
Niech D_f oznacza zbiór punktów nieciągłości funkcji

$$f : \mathbb{R}^k \to \mathbb{R}^m.$$

Wówczas jeśli ciąg zmiennych losowych Z_n zbiega według rozkładu, prawdopodobieństwa lub prawie napewno do zmiennej losowej Z takiej, że

$$P(Z \in D_f) = 0,$$

to ciąg zmiennych losowych $f(Z_n)$ zbiega odpowiednio według rozkładu, prawdopodobieństwa lub prawie napewno do zmiennej losowej $f(Z)$.

Twierdzenie 8.3. *(Slutsky)*
Niech (X_n) i (Y_n) ciagi zmiennych losowych o wartościach macierzowych, X macierzowa zmienna losowa i C macierz (deterministyczna). Wówczas, jeśli

$$X_n \xrightarrow{d} X, \quad a \quad Y_n \xrightarrow{d} C,$$

to:

$$\begin{aligned} &a. \quad X_n + Y_n \xrightarrow{d} X + C, \\ &b. \quad Y_n X_n \xrightarrow{d} CX, \\ &c. \quad \exists C^{-1} \implies Y_n^{-1} X_n \xrightarrow{d} C^{-1} X, \end{aligned}$$

gdzie Y_n^{-1} to dowolna macierz losowa taka, że

$$Y_n(\omega) Y_n^{-1}(\omega) = Id$$

gdy $\det Y_n(\omega) \neq 0$.

Dowód.
Ponieważ Y_n zbiega do stałej, to

$$(X_n, Y_n) \xrightarrow{d} (X, C).$$

Działania $+, \cdot$ są ciągłe zatem z twierdzenia o odwzorowaniu ciągłym (8.2) otrzymujemy tezę twierdzenia.
W przypadku odwracania macierzy zbiorem punktów nieciągłości jest zbiór macierzy o wyznaczniku 0. Zatem wystarczy zauważyć, że prawdopodobieństwo, że odwracalna macierz deterministyczna ma zerowy wyznacznik jest równe 0.

□

Uwaga 8.3. Niech X_ni X to wektory kolumnowe, czyli macierze $K \times 1$, a Y_n i C macierze $K \times K$. Wówczas, jeśli

$$X_n \xrightarrow{d} N(0, \Sigma), \quad Y_n \xrightarrow{d} C,$$

to

$$Y_n X_n \xrightarrow{d} N(0, C\Sigma C^T).$$

Twierdzenie 8.4. *Metoda delty.*
Niech X_n ciąg zmiennych losowych o wartościach w $\mathbb{R}^m$, X zmienna losowa o wartościach w $\mathbb{R}^m$, C punkt z $\mathbb{R}^m$, a $f : \mathbb{R}^m \to \mathbb{R}^K$ funkcja różniczkowalna w C. Wówczas, jeśli

$$X_n \xrightarrow{d} C, \quad i \quad \sqrt{n}(X_n - C) \xrightarrow{d} X$$

to

$$\sqrt{n}(f(X_n) - f(C)) \xrightarrow{d} Df(C)X.$$

Dowód.
Korzystając z twierdzenia 8.1 o reprezentacji prawie napewno, możemy zastąpić ciąg X_n ciągiem X'_n takim, że

$$X'_n \overset{d}{\sim} X_n, \quad X'_n \xrightarrow{as} C, \quad \text{i} \quad \sqrt{n}(X'_n - C) \xrightarrow{d} X.$$

Wówczas

$$\sqrt{n}(f(X_n) - f(C)) \overset{d}{\sim} \sqrt{n}(f(X'_n) - f(C)) =$$
$$= \sqrt{n}(f(X'_n) - f(C) - Df(C)(X'_n - C)) + \sqrt{n}Df(C)(X'_n - C) =$$
$$= \frac{f(X'_n) - f(C) - Df(C)(X'_n - C)}{\|X'_n - C\|}\sqrt{n}\|X'_n - C\| + DF(C)(\sqrt{n}(X'_n - C)).$$

Z różniczkowalności f otrzymujemy

$$\frac{f(X'_n) - f(C) - Df(C)(X'_n - C)}{\|X'_n - C\|} \xrightarrow{as} 0.$$

Ponieważ

$$\sqrt{n}\|X'_n - C\| \xrightarrow{d} \|X\| \quad \text{i} \quad DF(C)(\sqrt{n}(X'_n - C)) \xrightarrow{d} Df(C)X,$$

to z twierdzenia Slutsky'ego (8.3) otrzymujemy

$$\sqrt{n}(f(X_n) - f(C)) \xrightarrow{d} 0\|X\| + Df(C)X.$$

□

Uwaga 8.4. Jeśli przy założeniach twierdzenia 8.4 X ma rozkład normalny $N(0, \Sigma)$ to

$$\sqrt{n}(f(X_n) - f(C)) \xrightarrow{d} N(0, Df(C)\,\Sigma Df(C)^T).$$

8.2. Estymatory jako ciągi zmiennych losowych

Załóżmy, że proces generujący dane Z_t, $t = 0, 1, \ldots$, pochodzi z parametrycznej rodziny procesów $\mathcal{SP}(\theta)$, gdzie zbiór parametrów Θ jest podzbiorem $\mathbb{R}^m$

$$\theta \in \Theta \subset \mathbb{R}^m.$$

Ustalmy funkcję $\nu : \Theta \longrightarrow \mathbb{R}^k$ i rodzinę estymatorów

$$\widehat{\nu} = (\widehat{\nu}_n)_{n=1}^{\infty},$$

wartości funkcji ν w punkcie θ_Z. Niech $\widehat{\nu}_n$ będzie estymatorem $\nu(\theta_Z)$, wyznaczonym na podstawie próbki rozmiaru n, tzn.

$$\widehat{\nu}_n = \widehat{\nu}_n(Z_0, \ldots, Z_{n-1}).$$

Definicja 8.4. Ciąg estymatorów $\widehat{\nu}$ nazywamy nazywamy zgodnym gdy

$$\operatorname*{plim}_{n\to\infty} \widehat{\nu}_n = \nu(\theta_Z).$$

Definicja 8.5. Zgodny ciąg estymatorów $\widehat{\nu}$ nazywamy nazywamy asymptotycznie normalnym gdy

$$\sqrt{n}(\widehat{\nu}_n - \nu(\theta_Z)) \xrightarrow{d} N(0, \Sigma).$$

Uwaga 8.5. Macierz Σ oznaczamy $Avar(\widehat{\nu})$ i nazywamy asymptotyczną wariancją estymatora $\widehat{\nu}_n$.

8.3. Stacjonarność i ergodyczność procesów stochastycznych

8.3.1. Definicje i podstawowe własności

Niech $Z = (Z_t)_{t=0}^{\infty}$ będzie K-wymiarowym procesem stochastycznym (ciągiem zmiennych losowych) określonym na przestrzeni probabilistycznej $(\Omega, \mathcal{M}, P)$.

Definicja 8.6. Proces stochastyczny Z jest (silnie) stacjonarny gdy dla dowolnych $p, q, r \in \mathbb{N}$ łączne rozkłady

$$\{Z_p, Z_{p+1}, \ldots, Z_{p+q}\} \quad \text{i} \quad \{Z_r, \ldots, Z_{r+q}\}$$

są identyczne.

Wniosek 8.1. *Jeśli proces stochastyczny Z jest stacjonarny i Z_t należą do L^2, to dla wszystkich $p, q, r \in \mathbb{N}$*

$$E(Z_p) = E(Z_r), \quad Cov(Z_p, Z_{p+q}) = Cov(Z_r, Z_{r+q}).$$

Definicja 8.7. Stacjonarny proces stochastyczny Z ma własność mieszania gdy dla dowolnych ograniczonych funkcji borelowskich f i g oraz indeksów $p, l, m \in \mathbb{N}$

$$\lim_{n\to\infty} E(f(Z_p, \ldots, Z_{p+m})g(Z_n, \ldots, Z_{n+l})) = E(f(Z_p, \ldots, Z_{p+m}))E(g(Z_p, \ldots, Z_{p+l})).$$

Definicja 8.8. Stacjonarny proces stochastyczny Z jest ergodyczny gdy

$$\forall A \in \mathcal{B}(R^K) \quad P(\forall t \;\; Z_t \in A) \in \{0, 1\}.$$

Twierdzenie 8.5. *Twierdzenie ergodyczne.*
Jeśli proces stochastyczny Z jest stacjonarny i Z_t należą do L^1, to zachodzdzą implikacje

$$a \Longrightarrow b \Longrightarrow c,$$

gdzie
a. Z ma własność mieszania,
b. Z jest ergodyczny,
c. średnie zbiegają do wartości oczekiwanej

$$\overline{Z_n} = \frac{1}{n}\sum_{t=0}^{n-1} Z_t \xrightarrow{as} E(Z_0).$$

Uwaga 8.6. Jeśli proces stochastyczny Z jest stacjonarny i ma własność mieszania a f jest funkcją borelowską to proces $Z' = (f(Z_t, \ldots, Z_{t+q}))_{t=0}^{\infty}$ też jest stacjonarny i ma własność mieszania.
Zatem jeśli Z_t należą do L^2 to

$$\frac{1}{n}\sum_{t=0}^{n-1} Z_t^2 \xrightarrow{as} E(Z_0^2),$$

$$\frac{1}{n}\sum_{t=0}^{n-1} Z_t Z_{t+p} \xrightarrow{as} E(Z_0 Z_p).$$

Wniosek 8.2. *Dla procesów stacjonarnych i ergodycznych średnie próbkowe są zgodnymi estymatorami.*

8.3.2. Przykłady

Definicja 8.9. Proces ε nazywamy gaussowskim białym szumem gdy

$$\varepsilon = (\varepsilon_t)_{t=0}^{\infty}, \quad \varepsilon_t \sim N(0,\sigma), \quad \sigma > 0, \quad \varepsilon_t \text{ niezależne.}$$

Lemat 8.1. *Biały szum jest stacjonarny i ergodyczny.*

Dowód.
Dla dowolnych indeksów p i q wektor $(\varepsilon_p, \ldots, \varepsilon_{p+q})$ ma rozkład $N(0, Id_{q+1})$. Rozkład ten nie zależy od p co implikuje stacjonarność.

Aby pokazać własność mieszania zauważmy, że dla $n > p+m$ zmienne losowe $f(\varepsilon_p, \ldots, \varepsilon_{p+m})$ i $g(\varepsilon_n, \ldots, \varepsilon_{n+l})$ są niezależne. Zatem dla n odpowiednio dużych

$$E(f(\varepsilon_p, \ldots, \varepsilon_{p+m})g(\varepsilon_n, \ldots, \varepsilon_{n+l})) = E(f(\varepsilon_p, \ldots, \varepsilon_{p+m}))E(g(\varepsilon_n, \ldots, \varepsilon_{n+l})) =$$

$$= E(f(\varepsilon_p, \ldots, \varepsilon_{p+m}))E(g(\varepsilon_p, \ldots, \varepsilon_{p+l})).$$

□

Uwaga 8.7. Powyższe rozumowanie można zastosować dla dowolnego procesu iid tzn. o wyrazach niezależnych i o jednakowym rozkładzie.

Proces autoregresyjny

Niech ε będzie gaussowskim białym szumem. Dodatkowo założymy, ze $\varepsilon_t \sim N(0,1)$. Proces $Z = (Z_t)$ zdefiniujemy rekurencyjnie:

$$Z_t = c + \rho Z_{t-1} + \sigma\varepsilon_t, \quad t = 1, 2, \ldots,$$
$$Z_0 = \frac{c}{1-\rho} + \frac{\sigma}{\sqrt{1-\rho^2}}\varepsilon_0,$$

gdzie c, ρ, σ rzeczywiste parametry, $|\rho| < 1$, $\sigma > 0$. Tak zdefiniowany proces nazywa się autoregresyjnym rzędu 1 ($AR(1)$).

Lemat 8.2. *Proces Z jest stacjonarny i ergodyczny.*

Dowód.

Krok 1. Pokażemy, że wszystkie Z_t mają rozkład normalny o parametrach $\frac{c}{1-\rho}$ i $\frac{\sigma^2}{1-\rho^2}$

$$Z_t \sim N\left(\frac{c}{1-\rho}, \frac{\sigma^2}{1-\rho^2}\right).$$

Zastosujemy indukcję po t.
Jak łatwo zauważyć Z_0 ma rozkład normalny oraz

$$E(Z_0) = \frac{c}{1-\rho}, \quad D^2(Z_0) = \frac{\sigma^2}{1-\rho^2}.$$

Załóżmy, że

$$Z_s \sim N\left(\frac{c}{1-\rho}, \frac{\sigma^2}{1-\rho^2}\right), \quad s < t.$$

Ponieważ ε_t i Z_{t-1} mają rozkłady normalne i są niezależne to Z_t ma rozkład normalny. Ponadto

$$E(Z_t) = c + \rho E(Z_{t-1}) + \sigma E(\varepsilon) = c + \rho \frac{c}{1-\rho} + 0 = \frac{c}{1-\rho}.$$

$$D^2(Z_t) = \rho^2 D^2(Z_{t-1}) + \sigma^2 D^2(\varepsilon_t) = \rho^2 \frac{\sigma^2}{1-\rho^2} + \sigma^2 = \frac{\sigma^2}{1-\rho^2}.$$

Krok 2. Pokażemy, że $Cov(Z_{t+k}, Z_t) = \rho^k \frac{\sigma^2}{1-\rho^2}$.
Zastosujemy indukcję po k.
Dla $k = 0$ mamy

$$Cov(Z_t, Z_t) = D^2(Z_t) = \frac{\sigma^2}{1-\rho^2}.$$

Załóżmy, że

$$Cov(Z_{t+k-1}, Z_t) = \rho^{k-1} \frac{\sigma^2}{1-\rho^2}.$$

Wówczas

$$Cov(Z_{t+k}, Z_t) = Cov(c + \rho Z_{t+k-1} + \sigma\varepsilon_t, Z_t) = \rho Cov(Z_{t+k-1}, Z_t) = \rho \cdot \rho^{k-1} \frac{\sigma^2}{1-\rho^2} = \rho^k \frac{\sigma^2}{1-\rho^2}.$$

Krok 3. Stacjonarność.
Z poprzednich dwóch "kroków" wynika, że rozkład wektora $(Z_t, \ldots, Z_{t+p})$ nie zależy od t. Rzeczywiście

$$(Z_t, \ldots, Z_{t+p}) \sim N(\frac{c}{1-\rho} e, \frac{\sigma^2}{1-\rho^2} R_p),$$

gdzie e jest wektorem o $p+1$ współrzędnych, które są wszystkie równe 1, $e = (1, \ldots, 1)$, a R_p jest macierzą $p+1 \times p+1$ o wyrazach $(R_p)_{i,j} = \rho^{|i-j|}$, czyli

$$R_p = \begin{pmatrix} 1 & \rho & \rho^2 & \ldots & \rho^p \\ \rho & 1 & \rho & \ldots & \rho^{p-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho^{p-1}, & \rho^{p-2} & \rho^{p-3} & \ldots & \rho \\ \rho^p, & \rho^{p-1} & \rho^{p-2} & \ldots & 1 \end{pmatrix}.$$

Krok 4. Ergodyczność.
Dla $n > p+m$ wektor $(Z_p, \ldots, Z_{p+m}, Z_n, \ldots, Z_{n+l})$ ma $m+l+2$-wymiarowy rozkład normalny

$$(Z_p, \ldots, Z_{p+m}, Z_n, \ldots, Z_{n+l}) \sim N(\frac{c}{1-\rho} e, \frac{\sigma^2}{1-\rho^2} C_n),$$

gdzie macierz C otrzymujemy z macierzy $R_{n+l+1-p}$ przez wycięcie kolumn i wierszy od $m+$ 2-giego do $n-1$-ego.

$$C_n = \begin{pmatrix} R_m & \rho^{n-p-m} A \\ \rho^{n-p-m} A^T & R_l \end{pmatrix},$$

gdzie $m+1 \times l+1$ macierz A nie zależy od n, $A_{i,j} = \rho^{m+j-i}$

$$A = \begin{pmatrix} \rho^m & \rho^{m+1} & \rho^{m+2} & \ldots & \rho^{m+l} \\ \rho^{m-1} & \rho^m & \rho^{m+1} & \ldots & \rho^{m+l-1} \\ \ldots & \ldots & \ldots & \ldots & \ldots \\ \rho^1, & \rho^2 & \rho^3 & \ldots & \rho^{l+1} \\ 1, & \rho^1 & \rho^2 & \ldots & \rho^l \end{pmatrix}.$$

Zatem

$$\lim_{n\to\infty} C_n^{-1} = C_\infty^{-1} = \begin{pmatrix} R_m^{-1} & 0 \\ 0 & R_l^{-1} \end{pmatrix}.$$

W oparciu o powyższą granicę pokażemy własność mieszania.

$$E(f(Z_p,\ldots,Z_{p+m})g(Z_n,\ldots,Z_{n+l})) =$$

$$= \frac{1}{\sqrt{2\pi}^{m+l+2}} \det(C_n)^{-1} \left(\frac{1-\rho^2}{\sigma}\right)^{m+l+2} \int f(x')g(x'')exp(-\frac{1}{2}(x-\mu e)^T C_n^{-1}(x-\mu e))$$

$$\longrightarrow \frac{1}{\sqrt{2\pi}^{m+l+2}} \det(C_\infty)^{-1} \left(\frac{1-\rho^2}{\sigma}\right)^{m+l+2} \int f(x')g(x'')exp(-\frac{1}{2}(x-\mu e)^T C_\infty^{-1}(x-\mu e))dx =$$

$$= \frac{1}{\sqrt{2\pi}^{m+l+2}} \det(R_m)^{-1} \det R_l^{-1} \left(\frac{1-\rho^2}{\sigma}\right)^{m+l+2} \int f(x')exp(-\frac{1}{2}(x'-\mu e')^T R_m^{-1}(x'-\mu e'))dx'$$

$$\int g(x'')exp(-\frac{1}{2}(x''-\mu e'')^T R_l^{-1}(x''-\mu e''))dx'' =$$

$$= E(f(Z_p,\ldots,Z_{p+m}))E(g(Z_p,\ldots,Z_{p+l})).$$

□

8.4. Martyngały i przyrosty martyngałowe

Definicja 8.10. K-wymiarowy proces stochastyczny $Z = (Z_t)_{t=0}^\infty$ nazywamy martyngałem gdy

1. $Z_t \in L^1$ dla $t = 0, 1, \ldots,$
2. $E(Z_t | Z_{t-1}, \ldots, Z_0) = Z_{t-1}$ dla $t = 1, 2, \ldots.$

Definicja 8.11. K-wymiarowy proces stochastyczny $g = (g_t)_{t=0}^\infty$ nazywamy ciągiem przyrostów martyngałowych gdy

1. $E(g_t) = 0,$ dla $t = 0, 1, \ldots,$
2. $E(g_t | g_{t-1}, \ldots, g_0) = 0$ dla $t = 1, 2, \ldots.$

Uwaga 8.8. Jeśli proces $Z = (Z_t)_{t=0}^\infty$ jest martyngałem, to proces $g = (g_t)_{t=0}^\infty$, gdzie

$$g_0 = Z_0 - E(Z_0), \quad g_t = Z_t - Z_{t-1}, \quad t > 0,$$

jest ciągiem przyrostów martyngałowych.

Uwaga 8.9. Jeśli proces $g = (g_t)_{t=0}^\infty$ jest ciągiem przyrostów martyngałowych a μ dowolną stałą, to proces $Z = (Z_t)_{t=0}^\infty$, gdzie

$$Z_0 = \mu + g_0, \quad Z_t = Z_{t-1} + g_t, \quad t > 0,$$

jest martyngałem.

Lemat 8.3. *Jeśli proces $g = (g_t)_{t=0}^\infty$ jest ciągiem przyrostów martyngałowych i g_t należą do L^2 to są one nieskorelowane*

$$Cov(g_t, g_s) = 0, \quad dla \quad t \neq s.$$

Dowód.
Zapiszemy wektory g_t i g_{t+j}, $j > 0$, jako wektory kolumnowe. $E(g_t) = E(g_{t+j}) = 0$ zatem

$$Cov(g_t, g_{t+j}) = E(g_t g_{t+j}^T) = E(E(g_t g_{t+j}^T | g_{t+j-1}, \ldots, g_0)) =$$

$$= E(g_t E(g_{t+j}^T | g_{t+j-1}, \ldots, g_0)) = E(g_t \cdot 0) = 0.$$

□

Przykład
Biały szum $\varepsilon = (\varepsilon_t)$ jest ciągiem przyrostów martyngałowych, a błądzenie przypadkowe czyli proces $X = (X_t)_{t=0}^{\infty}$

$$X_0 = \mu + \varepsilon_0, \quad X_t = X_{t-1} + \varepsilon_t, \quad t > 0,$$

jest martyngałem.

Twierdzenie 8.6. *Centralne Twierdzenie Graniczne ([2] Twierdzenie 23.1).*
Jeśli stacjonarny i ergodyczny proces $g = (g_t)_{t=0}^{\infty}$ jest ciągiem przyrostów martyngałowych i g_t należą do L^2, to

$$\sqrt{n}\overline{g_n} = \frac{\sqrt{n}}{n} \sum_{t=0}^{n-1} g_t \xrightarrow{d} N(0, \Sigma),$$

gdzie $\Sigma = E(g_0 g_0^T)$.

Uwaga 8.10. Powyższe twierdzenie jest uogólnieniem CTG Linderberga-Levy'ego, w którym pominięta została niezależnośc składników.

9. Teoria dużej próbki

Teoria dużej próbki. Założenia modelu. Asymptotyczne własności estymatorów MNK. Statystyczna weryfikacja modelu. (2 wykłady)

9.1. Założenia modelu

W teorii dużej próbki przez model rozumie się $K+1$ wymiarowy proces stochastyczny

$$\{(Y_t, X_t)\}_{t=0}^{\infty}$$

spełniający pewne założenia.

$\tilde{\text{Z}}$1. Liniowość.

$$\exists \beta \in \mathbb{R}^K \quad \exists \{\varepsilon_t\}_{t=0}^{\infty} \quad Y_t = X_t\beta + \varepsilon_t, \quad t = 0, 1, \ldots,$$

$\tilde{\text{Z}}$2. Stacjonarność i ergodyczność.
$K+1$ wymiarowy proces stochastyczny $\{(Y_t, X_t)\}_{t=0}^{\infty}$ jest stacjonarny i ergodyczny.

Uwaga.
Z warunków $\tilde{\text{Z}}$1 i $\tilde{\text{Z}}$2 wynika, że również proces $\{\varepsilon_t\}_{t=0}^{\infty}$ jest stacjonarny i ergodyczny.

$\tilde{\text{Z}}$3. Warunek maksymalnego rzędu.
Proces X_t jest klasy L^2 i $K \times K$ macierz

$$\Sigma_{xx} = E(X_t^T X_t)$$

jest odwracalna.

Uwaga.
Ze stacjonarności procesu X_t wynika, że macierz Σ_{xx} nie zależy od t.

$\tilde{\text{Z}}$4. Ortogonalność zmiennych objaśniających do składnika losowego.

$$E(\varepsilon_t X_t) = 0, \quad t = 0, 1, \ldots.$$

Oznaczenie.

$$g_t = \varepsilon_t X_t.$$

Uwaga.
Z warunków $\tilde{\text{Z}}$1 i $\tilde{\text{Z}}$2 wynika, że również proces $\{g_t\}_{t=0}^{\infty}$ jest stacjonarny i ergodyczny.

$\tilde{\mathbf{Z}}$5. Martyngałowość.
Proces $\{g_t\}_{t=0}^{\infty}$ jest ciągiem przyrostów martyngałowych, g_t jest klasy L^2 i $K \times K$ macierz

$$\Sigma_{gg} = E(g_t^T g_t)$$

jest odwracalna.

Uwaga.
1. Ze stacjonarności procesu g_t wynika, że macierz Σ_{gg} nie zależy od t.
2. Z centralnego twierdzenia granicznego (8.6) wynika , że Σ_{gg} jest równa asymptotycznej wariancji procesu średnich $\bar{g}_t$

$$\Sigma_{gg} = Avar(\bar{g}_t), \quad \bar{g}_t = \frac{1}{t}\sum_{s=0}^{t-1} g_s.$$

$\tilde{\mathbf{Z}}$6. Warunkowa homoskedastyczność.

$$E(\varepsilon_t^2|X_t) = \sigma^2 > 0, \quad t = 0, 1, \ldots.$$

Lemat 9.1. *Z warunków $\tilde{Z}2$ i $\tilde{Z}6$ wynika, że*

$$\Sigma_{gg} = \sigma^2 \Sigma_{xx}.$$

Dowód.
$g_t = \varepsilon_t X_t$, zatem

$$\Sigma_{gg} = E(g_t^T g_t) = E(\varepsilon^2 X_t^T X_t) = E(E(\varepsilon_t^2 X_t^T X_t|X_t)) =$$

$$= E(E(\varepsilon_t^2|X_t) X_t^T X_t) = E(\sigma^2 X_t^T X_t) = \sigma^2 E(X_t^T X_t) = \sigma^2 \Sigma_{xx}.$$

□

Uwaga 9.1. Aksjomaty $\tilde{\mathbf{Z}}$ modelu dużej próbki (poza $\tilde{\mathbf{Z}}$**2**) są słabsze od aksjomatów **Z** modelu klasycznego. Otóż, niech $\{(Y_t, X_t)\}_{t=0}^{\infty}$ będzie $K+1$ wymiarowym, klasy L^2, stacjonarnym i ergodycznym procesem stochastycznym, wówczas jeśli dla każdego $n > K$ jego początkowy fragment $\{(Y_t, X_t)\}_{t=0}^{n}$ spełnia aksjomaty **Z1**, **Z2**, **Z3**, **Z4** i **Z5** to spełnia on $\tilde{\mathbf{Z}}$**1**, $\tilde{\mathbf{Z}}$**2**, $\tilde{\mathbf{Z}}$**3**, $\tilde{\mathbf{Z}}$**4**, $\tilde{\mathbf{Z}}$**5** i $\tilde{\mathbf{Z}}$**6**.

9.2. Asymptotyka estymatorów MNK

Niech $\{(Y_t, X_t)\}_{t=0}^{\infty}$ będzie procesem generującym dane. W modelu dużej próbki, podobnie jak w modelu klasycznym, będziemy estymowali parametry modelu β i σ^2 w oparciu o metodę najmniejszym kwadratów. Niech n oznacza ilość obserwacji ($n >> K$), X macierz wymiaru $n \times K$, której wierszami są wektory losowe X_t a Y wektor kolumnowy wymiaru n o wyrazach Y_t. Estymatory MNK możemy zapisać na dwa sposoby jako iloczyn macierzy X i Y lub za pomocą średnich z iloczynów X_t i Y_t.

Gdy macierz X ma rząd maksymalny czyli K, to estymator MNK wektora β wynosi

$$B = (X^T X)^{-1} X^T Y = S_{xx}^{-1} S_{xy},$$

gdzie

$$S_{xx} = \frac{1}{n} X^T X = \frac{1}{n} \sum_{t=0}^{n-1} X_t^T X_t,$$

$$S_{xy} = \frac{1}{n} X^T Y = \frac{1}{n} \sum_{t=0}^{n-1} Y_t X_t^T.$$

W wyjątkowych przypadkach gdy rząd macierzy X jest mniejszy niż K to jako B bierzemy dowolny wektor minimalizujący sumę kwadratów reszt – patrz uwaga 2.1.

Natomiast estymator MNK parametru σ^2 wynosi

$$S^2 = \frac{\xi^T \xi}{n-K} = \frac{1}{n-K} \sum_{t=0}^{n-1} \xi_t^2,$$

gdzie

$$\xi_t = Y_t - X_t B.$$

Omówimy teraz podstawowe własności powyższych estymatorów w zależności od wielkości próbki n.

Twierdzenie 9.1. *Własności estymatorów B i S^2:*
a. Zgodność B

$$\tilde{Z}1 - \tilde{Z}4 \Longrightarrow \lim_{n\to\infty} B = \beta \quad p.n.$$

b. Asymptotyczna normalność B

$$\tilde{Z}1 - \tilde{Z}5 \Longrightarrow \sqrt{n}(B-\beta) \xrightarrow{d} N(0, Avar(B)),$$

gdzie

$$Avar(B) = \Sigma_{xx}^{-1} \Sigma_{gg} \Sigma_{xx}^{-1}.$$

Jeśli dodatkowo założymy $\tilde{Z}6$, to

$$Avar(B) = \sigma^2 \Sigma_{xx}^{-1}.$$

c. Zgodność S^2

$$\tilde{Z}1 - \tilde{Z}4 \Longrightarrow \lim_{n\to\infty} S^2 = E(\varepsilon_0^2) \quad p.n.$$

Jeśli dodatkowo założymy $\tilde{Z}6$, to

$$\operatorname*{plim}_{n\to\infty} S^2 = \sigma^2.$$

d. Zgodna estymacja $Avar(B)$

$$\tilde{Z}1 - \tilde{Z}6 \Longrightarrow \lim_{n\to\infty} S^2 S_{xx}^{-1} = Avar(B) \quad p.n.$$

Dowód.
Mamy dwa równania opisujące zależność Y_t od X_t:

$$\begin{aligned} Y_t &= X_t\beta + \varepsilon_t, \\ Y_t &= X_t B + \xi_t. \end{aligned}$$

Po odjęciu stronami otrzymujemy:

$$X_t(B-\beta) = \varepsilon_t - \xi_t. \tag{9.1}$$

Mnożymy obie strony przez X_t^T

$$X_t^T X_t(B-\beta) = \varepsilon_t X_t^T - \xi_t X_t^T.$$

Następnie liczymy średnią po t. Biorąc pod uwagę, że $\sum_{t=0}^{n-1} \xi_t X_t^T = X^T \xi = 0$ (patrz wniosek 2.1) otrzymujemy:

$$S_{xx}(B-\beta) = \frac{1}{n}\sum_{t=0}^{n-1} X_t^T X_t(B-\beta) = \frac{1}{n}\sum_{t=0}^{n-1} \varepsilon_t X_t^T = \bar{g}_n^T.$$

Ponieważ proces $X_t^T X_t$ jest stacjonarny i ergodyczny to

$$S_{xx} = \frac{1}{n}\sum_{t=0}^{n-1} X_t^T X_t \xrightarrow{as} E(X_t^T X_t) = \Sigma_{xx}.$$

Macierz Σ_{xx} jest odwracalna (warunek $\tilde{\text{Z}}3$), zatem dla dużych n również macierz $S_{xx}(\omega)$ jest odwracalna. W wyjątkowych przypadkach gdy $\det S_{xx}(\omega) = 0$ dookreślamy $S_{xx}^{-1}(\omega)$ w dowolny sposób.

Po przemnożeniu obu stron przez macierz S_{xx}^{-1} otrzymujemy równość analogiczną do 4.2

$$B - \beta = S_{xx}^{-1}\bar{g}_n + \gamma_n, \tag{9.2}$$

gdzie $\gamma_n(\omega) = 0$ dla n odpowiednio dużych.
Ad. a.
Ponieważ proces g_t jest stacjonarny i ergodyczny to z warunku $\tilde{\text{Z}}4$ otrzymujemy

$$\lim_{n\to\infty} \bar{g}_n = E(g_0) = 0 \quad p.n.$$

Zatem

$$\lim_{n\to\infty} B - \beta = \lim_{n\to\infty} S_{xx}^{-1}\bar{g}_n = \Sigma_{xx}^{-1} 0 = 0 \quad p.n.$$

Ad. b.
Proces g_t oprócz tego, że jest stacjonarny i ergodyczny to jest ciągiem przyrostów martyngałowych, a więc ((8.6))

$$\sqrt{n}\bar{g} \xrightarrow{d} N(0, \Sigma_{gg}).$$

Zatem

$$\sqrt{n}(B-\beta) \xrightarrow{d} N(0, \Sigma_{xx}^{-1}\Sigma_{gg}\Sigma_{xx}^{-1}).$$

Jeśli ponadto założymy $\tilde{\text{Z}}6$ to

$$\Sigma_{gg} = \sigma^2 \Sigma_{xx},$$

zatem

$$Avar(B) = \Sigma_{xx}^{-1}\Sigma_{gg}\Sigma_{xx}^{-1} = \Sigma_{xx}^{-1}(\sigma^2\Sigma_{xx})\Sigma_{xx}^{-1} = \sigma^2\Sigma_{xx}^{-1}.$$

Ad. c.
Przepisując odpowiednio równanie 9.1 otrzymujemy

$$\xi_t = \varepsilon_t - X_t(B-\beta).$$

Zatem

$$\xi_t^2 = \varepsilon_t^2 - 2g_t(B-\beta) + (B-\beta)^T X_t^T X_t(B-\beta).$$

Po uśrednieniu otrzymujemy

$$\frac{1}{n}\sum_{t=0}^{n-1}\xi_t^2 = \frac{1}{n}\sum_{t=0}^{n-1}\varepsilon_t^2 - 2\bar{g}_n(B-\beta) + (B-\beta)^T S_{xx}(B-\beta).$$

Przechodzimy do granicy. Ponieważ

$$\lim_{n\to\infty} B-\beta = 0 = \lim_{n\to\infty}\bar{g}_n, \quad \text{oraz} \quad \lim_{n\to\infty} S_{xx} = \Sigma_{xx} \quad p.n.,$$

to otrzymujemy, że

$$\lim_{n\to\infty}\frac{1}{n}\sum_{t=0}^{n-1}\xi_t^2 = E(\varepsilon^2) \quad p.n.$$

Zatem

$$\lim_{n\to\infty} S^2 = \lim_{n\to\infty}\frac{n}{n-K}\frac{1}{n}\sum_{t=0}^{n-1}\xi_t^2 = E(\varepsilon^2) \quad p.n.$$

Jeśli ponadto założymy Z̃6 to

$$E(\varepsilon^2) = E(E(\varepsilon^2|X_t)) = \sigma^2,$$

zatem

$$\lim_{n\to\infty} S^2 = \sigma^2 \quad p.n.$$

Ad. d.
Korzystając z warunku Z̃6 otrzymujemy

$$\lim_{n\to\infty} S^2 S_{xx}^{-1} = \sigma^2\Sigma_{xx}^{-1} = Avar(B) \quad p.n.$$

□

10. Teoria dużej próbki cd

10.1. Testy asymptotyczne

10.1.1. Testowanie pojedynczego parametru strukturalnego b_k

Przyjmujemy założenia $\tilde{Z}1$ – $\tilde{Z}6$. Wówczas dwa poniższe ciągi mają tą samą granicę:

$$\sqrt{n}(B-\beta) \overset{d}{\sim} N(0, S^2 S_{xx}^{-1}).$$

Zatem dla pojedynczej współrzędnej otrzymujemy

$$\sqrt{n}(B_k-\beta_k) \overset{d}{\sim} N(0, S^2 S_{xx}^{-1})_{k,k}.$$

Oznaczenie. Błąd standardowy estymatora B_k będziemy oznaczać przez $SE(B_k)$

$$SE(B_k) = \sqrt{\frac{1}{n} S^2 (S_{xx}^{-1})_{k,k}}.$$

Testujemy hipotezę $H_0 : \beta_k = \bar{\beta}_k$ wobec hipotezy alternatywnej $H_1 : \beta_k \neq \bar{\beta}_k$, gdzie $\bar{\beta}_k$ ustalona liczba rzeczywista.

Jako statystykę testową przyjmujemy stosunek odchylenia estymatora od wartości testowej i błędu standardowego estymatora

$$T_k = \frac{B_k - \bar{\beta}_k}{SE(B_k)}.$$

Twierdzenie 10.1. *Przy założeniach $\tilde{Z}1$ – $\tilde{Z}6$ i H_0*

$$T_k = \frac{B_k - \bar{\beta}_k}{SE(B_k)} \xrightarrow{d} N(0,1).$$

Dowód.
Ponieważ $\sqrt{n}(B_k-\beta_k)$ zbiega według rozkładu do $N(0, Avar(B_k))$ a $S^2(S_{xx}^{-1})_{k,k}$ prawie napewno do $Avar(B_k)$, to otrzymujemy

$$T_k = \frac{B_k - \bar{\beta}_k}{\sqrt{\frac{1}{n} S^2 (S_{xx}^{-1})_{k,k}}} = \frac{\sqrt{n}(B_k - \bar{\beta}_k)}{\sqrt{S^2 (S_{xx}^{-1})_{k,k}}} \xrightarrow{d} N(0,1).$$

□

Reguła decyzyjna dla zadanego poziomu istotności α.
1. Na podstawie próbki ω wyznaczamy realizację statystyki testowej $t_k = T_k(\omega)$.
2. Wyznaczamy wartość krytyczną $t^*_{\alpha/2}$ jako kwantyl rozkładu normalnego

$$F(t^*_{\alpha/2}) = 1 - \frac{\alpha}{2},$$

gdzie F oznacza dystrybuantę rozkładu $N(0,1)$.
3. Jeżeli $|t_k| < t^*_{\alpha/2}$ to nie ma podstaw do odrzucenia hipotezy H_0.
Jeżeli $|t_k| \geqslant t^*_{\alpha/2}$ to odrzucamy H_0 na rzecz H_1.

Uwaga.
Powyższa reguła jest zgodna z regułami przedstawionymi w rozdziale 5.1, gdyż rozkład t-Studenta wraz ze wzrostem liczby stopni swobody zbiega do standardowego rozkładu normalnego

$$t(n-K) \xrightarrow{d} N(0,1).$$

10.1.2. Testowanie hipotezy liniowości

Zajmiemy sie teraz testowaniem hipotezy, że nieznany parametr $\beta = (\beta_1, \ldots, \beta_K)^T$ spełnia m niezależnych warunków liniowych. Czyli, że należy do podprzestrzeni afinicznej kowymiaru m.

Niech r macierz o współczynnikach rzeczywistych wymiaru $m \times K$, rzędu m, gdzie $m = 1, \ldots, K$, a $\tilde{r}$ wektor kolumnowy wymiaru m. Testujemy hipotezę

$$H_0: \quad r\beta = \tilde{r},$$

wobec

$$H_1: \quad r\beta \neq \tilde{r}.$$

Jako statystykę testową przyjmujemy statystykę Walda

$$W = \frac{n}{S^2}(rB - \tilde{r})^T (rS_{xx}^{-1} r^T)^{-1} (rB - \tilde{r}).$$

Twierdzenie 10.2. *Przy założeniach $\tilde{Z}1$ – $\tilde{Z}6$ i H_0*

$$W = \frac{n}{S^2}(rB - \tilde{r})^T (rS_{xx}^{-1} r^T)^{-1} (rB - \tilde{r}) \xrightarrow{d} \chi^2(m),$$

gdzie $\chi^2(m)$ oznacza rozkład chi-kwadrat z m stopniami swobody.

Dowód.
Zapiszemy statystykę W jako

$$W = C_n^T Q_n^{-1} C_n,$$

gdzie

$$C_n = \sqrt{n}(rB - \tilde{r}), \quad Q_n = S^2 r S_{xx}^{-1} r^T.$$

Hipoteza zerowa H_0 implikuje, że $r\beta = \tilde{r}$ i

$$C_n = r(\sqrt{n}(B - \beta)).$$

Zatem (Tw. 9.1 b.)

$$C_n \xrightarrow{d} C \sim N(0, r\, Avar(B)\, r^T).$$

Natomiast Q_n zbiega do pewnej macierzy deterministycznej Q (Tw. 9.1 d.)

$$Q_n \xrightarrow{as} Q = r\, Avar(B)\, r^T = Var(C).$$

Ponieważ asymptotyczna macierz wariancji $Avar(B)$ jest dodatnio określona, a wiersze macierzy r są liniowo niezależne (macierz r ma rząd m), to $m \times m$ macierz Q jest odwracalna (jest to macierz Grama wierszy macierzy r). Zatem przechodząc do granicy otrzymujemy

$$W \xrightarrow{d} C^T Q^{-1} C = C^T Var(C)^{-1} C \sim \chi^2(m).$$

□

Uwaga 10.1. Zauważmy, że statystyka Walda daje się wyrazić przez F-statystykę stosowaną w modelu klasycznym (Tw. 5.2)

$$W = \frac{n}{n-K} mF.$$

W przypadku gdy spełnione są zarówno aksjomaty modelu klasycznego jak i modelu dużej próbki, to oba podejścia są zgodne. Rzeczywiście gdy $n \to \infty$, to

$$\frac{\chi^2(n-K)}{n-K} \xrightarrow{as} 1,$$

a zatem

$$F(m, n-K) \overset{d}{\sim} \frac{\frac{\chi^2(m)}{m}}{\frac{\chi^2(n-K)}{n-K}} \xrightarrow{d} \frac{\chi^2(m)}{m}.$$

Uwaga 10.2. Podobnie jak w modelu klasycznym możemy wyrazić statystykę W za pomocą sum kwadratów reszt. Niech SKR_o suma kwadratów reszt dla modelu spełniającego ograniczenie $r\beta = \tilde{r}$. Wówczas

$$W = m \frac{SKR_o - SKR}{SKR}.$$

Uwaga 10.3. W modelu z wyrazem stałym ($X_K = e$) współczynnik determinacji daje się wyrazić za pomocą statystyki W wyznaczonej dla $K-1$ warunków $\beta_1 = \cdots = \beta_{K-1} = 0$,

$$nR^2 = \frac{W}{1 + \frac{W}{n}}.$$

Zatem, po przejściu z n do granicy otrzymujemy

$$nR^2 \xrightarrow{d} \chi^2(m).$$

10.1.3. Testowanie nieliniowych zależności między parametrami modelu

Metoda delty (Tw. 8.4) pozwala uogólnić test Walda na przypadek zależności nieliniowych. Niech Φ oznacza submersję

$$\Phi : \mathbb{R}^K \longrightarrow \mathbb{R}^m, \quad m \leqslant K,$$

tzn. odwzorowanie klasy C^1 o maksymalnym rzędzie pochodnej

$$\forall p \in \mathbb{R}^K \quad rank(D\Phi(p)) = m.$$

Dla ustalenia uwagi przyjmiemy, że $\Phi(p)$ jest wektorem kolumnowym.

Będziemy testować hipotezę

$$H_0 : \Phi(\beta) = 0,$$

wobec hipotezy alternatywnej

$$H_1 : \Phi(\beta) \neq 0.$$

Jako statystykę testową przyjmujemy statystykę Walda

$$W = \frac{n}{S^2} \Phi(B)^T (D\Phi(B) S_{xx}^{-1} D\Phi(B)^T)^{-1} \Phi(B).$$

Twierdzenie 10.3. *Przy założeniach $\tilde{Z}1$ – $\tilde{Z}6$ i H_0*

$$W = \frac{n}{S^2}\Phi(B)^T(D\Phi(B)S_{xx}^{-1}D\Phi(B)^T)^{-1}\Phi(B) \xrightarrow{d} \chi^2(m),$$

gdzie $\chi^2(m)$ oznacza rozkład chi-kwadrat z m stopniami swobody.

Dowód.
Z twierdzenia o odwzorowaniu ciągłym (Tw. 8.2) i założenia H_0 otrzymujemy, że

$$\Phi(B) \xrightarrow{as} \Phi(\beta) = 0.$$

Zatem możemy zastosować metodę delty (Tw. 8.4)

$$\sqrt{n}\Phi(B) = \sqrt{n}(\Phi(B) - \Phi(\beta)) \xrightarrow{d} C \sim N(0, D\Phi(\beta)Avar(B)D\Phi(\beta)^T).$$

Z założeń modelowych wynika, że

$$S^2S_{xx}^{-1} \xrightarrow{as} Avar(B),$$

a z twierdzenia o odwzorowaniu ciągłym, że

$$D\Phi(B) \xrightarrow{as} D\Phi(\beta).$$

Z powyższego wynika, że

$$S^2D\Phi(B)S_{xx}^{-1}D\Phi(B)^T \xrightarrow{as} D\Phi(\beta)Avar(B)D\Phi(\beta)^T = Var(C).$$

Ponieważ macierz pochodnych ma rząd m to $m \times m$ macierz $Var(C)$ jest odwracalna. Zatem statystyka W jest dobrze zdefiniowana i

$$W = \frac{n}{S^2}\Phi(B)^T(D\Phi(B)S_{xx}^{-1}D\Phi(B)^T)^{-1}\Phi(B) \xrightarrow{d} C^TVar(C)^{-1}C \sim \chi^2(m).$$

□

10.1.4. Testowanie warunkowej homoskedastyczności – test White'a

Rozważamy model liniowy

$$Y_t = X_t\beta + \varepsilon_t, \quad t = 0, 1, \ldots,$$

gdzie X_t jest K wymiarowym wektorem wierszowym.
Niech

$$Z_t = (Z_{t,1}, \ldots, Z_{t,m}, 1)$$

wektor złożony z wybranych elementów $K \times K$ macierzy X_t^TX, $m \leqslant K^2$ i wyrazu stałego. Załóżmy, że proces $\{\varepsilon_t^2, Z_t\}$ spełnia warunki $\tilde{Z}1$ – $\tilde{Z}6$ modelu dużej próbki, w szczególności, że istnieją wektor $\tilde{\beta}$ i składnik losowy ε_1 takie, że

$$\varepsilon_t^2 = Z_t\tilde{\beta} + \varepsilon_{1,t}.$$

Testujemy hipotezę

$$H_0 : \tilde{\beta}_1 = \cdots = \tilde{\beta}_m = 0,$$

wobec hipotezy alternatywnej

$$H_1 : \exists i \quad \tilde{\beta}_i \neq 0.$$

W teście White reguła decyzyjna opiera się na fakcie, że przy założeniu H_0

$$nR^2 \xrightarrow{d} \chi^2(m),$$

gdzie R^2 współczynnik determinacji wyznaczony dla modelu pomocniczego

$$\xi_t^2 = Z_t\tilde{\beta} + \xi_{1,t}, \quad \xi_t = Y - X_t B,$$

w którym składnik losowy ε zastąpiono składnikiem resztowym ξ z metody MNK.

11. Testowanie autokorelacji składnika losowego i składnika resztowego

Rozkład współczynników autokowariancji i autokorelacji empirycznych. Statystyki Q Ljunga-Boxa i Boxa-Pierce'a. (1 wykład)

11.1. Autokorelacja składnika losowego

Niech $\{Z_t\}_{t=-\infty}^{+\infty}$ skalarny proces stochastyczny (szereg czasowy), stacjonarny, ergodyczny i klasy L^2. Dodatkowo założymy, że zmienne losowe Z_t nie są stałe (deterministyczne).

Ze stacjonarności wynika, że kowariancja Z_t i Z_s zależy tylko od różnicy $t - s$.

Definicja 11.1. j-tym współczynnikiem autokowariancji nazywamy kowariancję zmiennych losowych Z_* odległych o j

$$\gamma_j = cov(Z_i, Z_{i-j}).$$

Jak łatwo zauważyć

$$\gamma_0 = D^2(Z_i) > 0$$

oraz

$$\gamma_{-j} = \gamma_j.$$

Definicja 11.2. j-tym współczynnikiem autokorelacji nazywamy współczynnik korelacji zmiennych losowych Z_* odległych o j

$$\rho_j = \frac{\gamma_j}{\gamma_0}.$$

Naszym celem jest estymacja współczynników autokowariancji i autokorelacji na podstawie n-elementowej próbki Z_t, $t = 0, \ldots, n-1$. Będziemy korzystali z następujących estymatorów:

$$\widehat{\gamma_j} = \frac{1}{n}\sum_{t=j}^{n-1}(Z_t - \bar{Z})(Z_{t-j} - \bar{Z}), \quad \bar{Z} = \frac{1}{n}\sum_{t=0}^{n-1} Z_t,$$

$$\widehat{\rho_j} = \frac{\widehat{\gamma_j}}{\widehat{\gamma_0}}.$$

Uwaga 11.1. Z ergodyczności procesu Z_t wynika, że estymatory $\widehat{\gamma_j}$ i $\widehat{\rho_j}$ są zgodne

$$\widehat{\gamma_j} \xrightarrow{as} \gamma_j, \quad \widehat{\rho_j} \xrightarrow{as} \rho_j.$$

Przy pewnych dodatkowych założeniach można pokazać, że estymatory $\widehat{\gamma_j}$ i $\widehat{\rho_j}$ są asymptotycznie normalne.

Twierdzenie 11.1. *Niech*

$$Z_t = \mu + \varepsilon_t,$$

gdzie μ jest pewną stałą, a stacjonarny i ergodyczny proces $\{\varepsilon_t\}$ jest ciągiem przyrostów martyngałowych takim, że

$$E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) = \sigma^2 > 0.$$

Wówczas dla każdego $p \in \mathbb{N} \setminus \{0\}$

$$\sqrt{n}\,\widehat{\gamma} \xrightarrow{d} N(0, \sigma^4 Id_p),$$

$$\sqrt{n}\,\widehat{\rho} \xrightarrow{d} N(0, Id_p),$$

gdzie

$$\widehat{\gamma} = (\widehat{\gamma_1}, \dots, \widehat{\gamma_p}), \quad \widehat{\rho} = (\widehat{\rho_1}, \dots, \widehat{\rho_p}).$$

Dowód.
Zaczniemy od wyznaczenia pierwszego i drugiego momentu Z_t oraz kowariancji Z_t i Z_{t-j} dla $j > 0$

$$E(Z_t) = \mu,$$

$$D^2(Z_t) = E((Z_t - \mu)^2) = E(\varepsilon_t^2) = E(E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)) = E(\sigma^2) = \sigma^2,$$

$$cov(Z_t, Z_{t-j}) = E((Z_t - \mu)(Z_{t-j} - \mu)) = E(\varepsilon_t \varepsilon_{t-j}) = E(E(\varepsilon_t \varepsilon_{t-j} | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)) =$$
$$= E(\varepsilon_{t-j} E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)) = E(0) = 0.$$

Podsumowując

$$\gamma_0 = \sigma^2 > 0, \quad \gamma_j = 0 \text{ dla } j \neq 0,$$

$$\rho_0 = 1, \quad \rho_j = 0 \text{ dla } j \neq 0,$$

a zatem

$$\widehat{\gamma} \xrightarrow{as} 0, \quad \widehat{\rho} \xrightarrow{as} 0.$$

Następnie zajmiemy się badaniem procesów iloczynów

$$g_{j,t} = (Z_t - \mu)(Z_{t-j} - \mu) = \varepsilon_t \varepsilon_{t-j}, \quad j = 1, 2, 3, \dots.$$

$$E(g_{j,t}) = 0,$$

$$E(g_{j,t} | g_{j,t-1}, g_{j,t-2}, \dots) = E(E(\varepsilon_t \varepsilon_{t-j} | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) | g_{j,t-1}, g_{j,t-2}, \dots) =$$
$$= E(\varepsilon_{t-j} E(\varepsilon_t | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots) | g_{j,t-1}, g_{j,t-2}, \dots) = 0,$$

zatem wszystkie procesy $\{g_{j,t}\}$ są ciągami przyrostów martyngałowych. Wyznaczymy wariancję $g_{j,t}$ i kowariancję $g_{j,t}$ i $g_{k,t}$ dla $k > j$

$$E(g_{j,t}^2) = E(E(\varepsilon_t^2 \varepsilon_{t-j}^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)) = E(\varepsilon_{t-j}^2 E(\varepsilon_t^2 | \varepsilon_{t-1}, \dots)) = E(\varepsilon_{t-j}^2 \sigma^2) = \sigma^2 E(\varepsilon_{t-j}^2) = \sigma^4.$$

$$E(g_{j,t} g_{k,t}) = E(E(\varepsilon_t^2 \varepsilon_{t-j} \varepsilon_{t-k} | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)) = E(\varepsilon_{t-j} \varepsilon_{t-k} E(\varepsilon_t^2 | \varepsilon_{t-1}, \varepsilon_{t-2}, \dots)) =$$
$$= E(\varepsilon_{t-j} \varepsilon_{t-k} \sigma^2) = \sigma^2 E(\varepsilon_{t-j} \varepsilon_{t-k}) = 0.$$

Oznaczmy przez g_t wektor o wyrazach $g_{j,t}$ dla j od 1 do p

$$g_t = (g_{1,t}, \ldots, g_{p,t}).$$

Proces $\{g_t\}$ jest stacjonarnym i ergodycznym ciągiem przyrostów martyngałowych o skończonej sferycznej wariancji

$$Var(g_t) = \sigma^4 Id_p.$$

Zatem na mocy Centralnego Twierdzenia Granicznego (Tw. 8.6)

$$\sqrt{n}\frac{1}{n}\sum_{t=0}^{n-1} g_t \xrightarrow{d} N(0, \sigma^4 Id_p).$$

Zauważmy, że

$$\widehat{\gamma_j} = \frac{1}{n}\sum_{t=j}^{n-1} g_t,$$

zatem, również

$$\sqrt{n}\widehat{\gamma} \xrightarrow{d} N(0, \sigma^4 Id_p).$$

Natomiast

$$\widehat{\rho} = \frac{1}{\widehat{\gamma_0}}\widehat{\gamma}.$$

Ponieważ

$$\widehat{\gamma_0} \xrightarrow{as} \sigma^2,$$

to

$$\sqrt{n}\widehat{\rho} = \frac{1}{\widehat{\gamma_0}}\sqrt{n}\widehat{\gamma} \xrightarrow{d} N(0, Id_p).$$

□

W praktycznych zastosowaniach wygodniej jest stosować statystyki 1-wymiarowe. Przykładem są statystyki Q:

• Boxa-Pierce'a

$$Q = n\sum_{j=1}^{p} \widehat{\rho_j}^2$$

• i Ljunga-Boxa

$$Q = n(n+2)\sum_{j=1}^{p} \frac{\widehat{\rho_j}^2}{n-j}.$$

Wniosek 11.1. *Przy założeniach twierdzenia 11.1*

1. $\sqrt{n}\widehat{\rho_j} \xrightarrow{d} N(0,1), \quad j = 1, 2, \ldots,$
2. $Q = n\sum_{j=1}^{p} \widehat{\rho_j}^2 = \sum_{j=1}^{p} (\sqrt{n}\widehat{\rho_j})^2 \xrightarrow{d} \chi^2(p),$
3. $Q = n(n+2)\sum_{j=1}^{p} \frac{\widehat{\rho_j}^2}{n-j} = \sum_{j=1}^{p} \frac{n+2}{n-j}(\sqrt{n}\widehat{\rho_j})^2 \xrightarrow{d} \chi^2(p).$

11.2. Autokorelacja składnika resztowego

Rozważamy model liniowy z wyrazem wolnym

$$Y_t = X_t\beta + \varepsilon_t, \quad X_t = (X_{t,1}, \ldots, X_{t,K}), \quad X_{t,K} = 1,$$

spełniający warunki $\tilde{Z}1$ – $\tilde{Z}5$. Niech B_n będzie estymatorem MNK wyznaczonym na podstawie próbki n-elementowej, a $\xi_{n,t}$ składnikiem resztowym

$$Y_t = X_t B_n + \xi_{n,t}, \quad t = 0, \ldots, n-1.$$

Przyjmiemy następujące oznaczenia:
γ_* i $\rho_a st$ współczynniki autokowariancji i autokorelacji składnika losowego ε

$$\gamma_j = E(\varepsilon_j \varepsilon_0), \quad \rho_j = \frac{\gamma_j}{\gamma_0};$$

$\tilde{\gamma}_*$ i $\tilde{\rho}_*$ współczynniki "próbkowe" autokowariancji i autokorelacji składnika losowego ε

$$\tilde{\gamma}_j = \frac{1}{n}\sum_{t=j}^{n-1} \varepsilon_t \varepsilon_{t-j}, \quad \tilde{\rho}_j = \frac{\tilde{\gamma}_j}{\tilde{\gamma}_0};$$

$\widehat{\gamma}_*$ i $\widehat{\rho}_*$ współczynniki próbkowe autokowariancji i autokorelacji składnika losowego ξ

$$\widehat{\gamma}_j = \frac{1}{n}\sum_{t=j}^{n-1} \xi_{n,t}\xi_{n,t-j}, \quad \widehat{\rho}_j = \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0}.$$

Proces ε jest stacjonarny i ergodyczny zatem dla każdego j

$$\tilde{\gamma}_j \xrightarrow{as} \gamma_j, \quad \tilde{\rho}_j \xrightarrow{as} \rho_j.$$

Niestety, składnik $\{\varepsilon_t\}$ jest nieobserwowalny. Znamy tylko składnik resztowy $\{\xi_{n,t}\}$, zatem jako ewentualne estymatory można rozważać $\widehat{\gamma}_j$ i $\widehat{\rho}_j$.

Pytanie?
Czy można zastąpić w statystyce Q współczynniki "próbkowe" autokorelacji składnika losowego ε, przez współczynniki "próbkowe" autokorelacji składnika resztowego ξ, nie psując własności tej statystyki?

Okazuje się, że przy dodatkowych założeniach odpowiedź jest pozytywna.

Lemat 11.1.

$$\forall j \quad \widehat{\gamma}_j - \tilde{\gamma}_j \xrightarrow{as} 0 \quad i \quad \widehat{\rho}_j - \tilde{\rho}_j \xrightarrow{as} 0.$$

Dowód.
Skorzystamy z zależności

$$\xi_{n,t} = \varepsilon_t - X_t(B_n - \beta).$$

$$\widehat{\gamma}_j - \tilde{\gamma}_j = \frac{1}{n}\sum_{t=j}^{n-1} \xi_{n,t}\xi_{n,t-j} - \frac{1}{n}\sum_{t=j}^{n-1} \varepsilon_t\varepsilon_{t-j} =$$

$$= \frac{1}{n}\sum_{t=j}^{n-1} \left((\varepsilon_t - X_t(B_n - \beta))(\varepsilon_{t-j} - X_{t-j}(B_n - \beta)) - \varepsilon_t\varepsilon_{t-j}\right) =$$

$$= -\left(\frac{1}{n}\sum_{t=j}^{n-1}\varepsilon_t X_{t-j} + \frac{1}{n}\sum_{t=j}^{n-1}\varepsilon_{t-j}X_t\right)(B_n-\beta) + (B_n-\beta)^T\left(\frac{1}{n}\sum_{t=j}^{n-1}X_t^T X_{t-j}\right)(B_n-\beta)$$

$$\xrightarrow{as} -\left(E(\varepsilon_j X_0) + E(\varepsilon_0 X_j)\right)\cdot 0 + 0\cdot E(X_j^T X_0)\cdot 0 = 0.$$

Natomiast dla ρ mamy

$$\widehat{\rho}_j - \tilde{\rho}_j = \frac{\widehat{\gamma}_j}{\widehat{\gamma}_0} - \frac{\tilde{\gamma}_j}{\tilde{\gamma}_0} = \frac{\widehat{\gamma}_j\tilde{\gamma}_0 - \widehat{\gamma}_0\tilde{\gamma}_j}{\widehat{\gamma}_0\tilde{\gamma}_0} =$$

$$= \frac{(\widehat{\gamma}_j - \tilde{\gamma}_j)\tilde{\gamma}_0 + \tilde{\gamma}_j(\tilde{\gamma}_0 - \widehat{\gamma}_0)}{\widehat{\gamma}_0\tilde{\gamma}_0} \xrightarrow{as} \frac{0}{\gamma_0^2} = 0.$$

□

Lemat 11.2. *Jeśli dodatkowo spełniony jest warunek ścisłej egzogeniczności (Z2)*

$$\forall t \quad E(\varepsilon_t|X) = 0,$$

to

$$\forall j \quad \sqrt{n}\,(\widehat{\gamma}_j - \tilde{\gamma}_j) \xrightarrow{p} 0 \quad i \quad \sqrt{n}\,(\widehat{\rho}_j - \tilde{\rho}_j) \xrightarrow{p} 0.$$

Dowód.
Zauważmy, że ze ścisłej egzogeniczności wynika, że

$$E(\varepsilon_j X_0) = E(\varepsilon_0 X_j) = 0.$$

Niech Z będzie pewną zmienną losową o rozkładzie $N(0, Avar(B))$. Wówczas

$$\sqrt{n}\,(\widehat{\gamma}_j - \tilde{\gamma}_j) =$$

$$= -\left(\frac{1}{n}\sum_{t=j}^{n-1}\varepsilon_t X_{t-j} + \frac{1}{n}\sum_{t=j}^{n-1}\varepsilon_{t-j}X_t\right)\sqrt{n}(B_n-\beta) + \frac{1}{\sqrt{n}}(\sqrt{n}(B_n-\beta))^T\left(\frac{1}{n}\sum_{t=j}^{n-1}X_t^T X_{t-j}\right)(\sqrt{n}(B_n-\beta))$$

$$\xrightarrow{d} -(E(\varepsilon_j X_0) + E(\varepsilon_0 X_j))Z + 0\cdot Z^T E(X_j X_0)Z = 0.$$

Zbieżność do zera według rozkładu implikuje zbieżność do 0 według prawdopodobieństwa, zatem

$$\sqrt{n}\,(\widehat{\gamma}_j - \tilde{\gamma}_j) \xrightarrow{p} 0.$$

Dowód zbieżności dla ρ przebiega analogicznie jak w poprzednim lemacie.

□

Wniosek 11.2. *Przy założeniach lematu 11.2 statystyka Q policzona dla ξ jest asymptotycznie równoważna statystyce Q policzonej dla ε, zatem obie zbiegają według rozkładu do $\chi^2(p)$.*

12. Hipoteza efektywnego rynku - ekonometria racjonalnych oczekiwań

Teoria dużej próbki cd. Przykład: Teoria racjonalnych oczekiwań. (1 wykład)

12.1. Przykład E.Fama - konstrukcja modelu

Określenie.
Mówimy, że rynek jest efektywny, gdy w sposób efektywny wykorzytywana jest posiadana informacja. Ceny na efektywnym rynku "w pełni" odzwierciedlają dostępne informacje.

Przeanalizujemy przykład E.Fama, aby sprawdzić na ile rynek amerykańskich bonów skarbowych jest efektywny. Ograniczymy się do bonów jednomiesięcznych.

Charakterystyka instrumentu:
- czas życia 1 miesiąc,
- wypłata 100 USD,
- zakup z dyskontem.

Podstawą analizy będą dane miesięczne:
t – numer kolejny miesiąca;
V_t – cena bonu na początku miesiąca t;
P_t – indeks cen konsumenta (CPI) na początku miesiąca t (por. [10], §1.4).

Na ich podstawie wyznaczamy:
R_t – miesięczną stopę zwrotu dla bonów w miesiącu t

$$R_t = \frac{1 - V_t}{V_t}, \quad \text{czyli} \quad V_t = \frac{1}{1 + R_t};$$

π_t – miesięczną stopę inflacji od początku miesiąca $t-1$ do początku miesiąca t

$$\pi_t = \frac{P_t - P_{t-1}}{P_{t-1}} = \frac{P_t}{P_{t-1}} - 1;$$

r_t – realną stopę zwrotu w miesiącu $t-1$ (por.[10], §1.4)

$$r_t = \frac{\frac{1}{P_t} - \frac{V_{t-1}}{P_{t-1}}}{\frac{V_{t-1}}{P_{t-1}}} = \frac{P_{t-1}}{P_t V_{t-1}} - 1 = \frac{1 + R_{t-1}}{1 + \pi_t} - 1 = \frac{R_{t-1} - \pi_t}{1 + \pi_t}.$$

Ze względu na małą inflację będziemy stosowali wzór przybliżony

$$r_t \approx R_{t-1} - \pi_t.$$

Ponadto w analizie uwzględnimy wielkości nieobserwowalne:
$\widehat{\pi}_{t+1,t}$ – prognozowaną (oczekiwaną) na początku miesiąca t stopę inflacji w miesiącu t ($\widehat{\pi}_{t+1,t}$ jest prognozą π_{t+1});
η_t – błąd prognozy inflacji

$$\eta_t = \pi_t - \widehat{\pi}_{t,t-1};$$

$\widehat{r}_{t+1,t}$ – prognozowaną (oczekiwaną) na początku miesiąca t realną stopę zwrotu z bonów w miesiącu t ($\widehat{r}_{t+1,t}$ jest prognozą r_{t+1});

$$\widehat{r}_{t+1,t} = \frac{R_t - \widehat{\pi}_{t+1,t}}{1 + \widehat{\pi}_{t+1,t}} \approx R_t - \widehat{\pi}_{t+1,t};$$

$\mathcal{I}_t$ – zasób informacji dostępny dla inwestorów na początku miesiąca t

$$\begin{array}{ll} \text{a)} & \mathcal{I}_t \supset \{R_t, R_{t-1}, \dots, \pi_t, \pi_{t-1}, \dots\}, \\ \text{b)} & \mathcal{I}_t \supseteq \mathcal{I}_{t-1} \supseteq \mathcal{I}_{t-2} \supseteq \dots. \end{array}$$

Własność b) oznacza, że agenci nie zapominają informacji z poprzednich miesięcy.

$\mathcal{I}_t$ modelujemy jako σ-ciała, natomiast $V_t, P_t, R_t, \pi_t, r_t, \widehat{\pi}_{t+1,t}, \eta_t, \widehat{r}_{t+1,t}$ modelujemy jako zmienne losowe klasy L^2 mierzalne względem $\mathcal{I}_t$.

12.2. Hipoteza efektywnego rynku

Hipoteza efektywnego rynku opiera się na dwóch założeniach:
E1. Racjonalne oczekiwania inflacyjne

$$\widehat{\pi}_{t+1,t} = E(\pi_{t+1}|\mathcal{I}_t).$$

E2. Stała oczekiwana realna stopa zwrotu

$$\exists r \quad \forall t \quad \widehat{r}_{t+1,t} = r.$$

Przeanalizujemy wnioski jakie wynikają z założeń E1 i E2.

Lemat 12.1.

a) $E(\eta_{t+1}|\mathcal{I}_t) = 0$;
b) $\mathcal{I}_t \supseteq \sigma\{\eta_t, \eta_{t-1}, \dots\}$;
c) *proces* $\{\eta_t\}$ *jest ciągiem przyrostów martyngałowych.*

Dowód.
Ad a.

$$E(\eta_{t+1}|\mathcal{I}_t) = E(\pi_{t+1} - \widehat{\pi}_{t+1,t}|\mathcal{I}_t) = E(\pi_{t+1}|\mathcal{I}_t) - \widehat{\pi}_{t+1,t} = \widehat{\pi}_{t+1,t} - \widehat{\pi}_{t+1,t} = 0.$$

Punkt b wynika z faktu, że dla $s \leqslant t$ zmienna losowa η_s jest $\mathcal{I}_s$ mierzalna zatem

$$\sigma(\eta_s) \subset \mathcal{I}_s \subset \mathcal{I}_t.$$

Punkt c wynika z a i b.

$$E(\eta_{t+1}) = E(E(\eta_{t+1}|\mathcal{I}_t)) = E(0) = 0.$$

$$E(\eta_{t+1}|\eta_t, \eta_{t-1}, \dots) = E(E(\eta_{t+1}|\mathcal{I}_t)|\eta_t, \eta_{t-1}, \dots) = E(0|\dots) = 0.$$

□

Szczególnie ważne są wnioski dotyczące wielkości obserwowalnych gdyż można je przetestować.

Wniosek 12.1.

a) $R_{t-1} - \pi_t = r - \eta_t$;

b) $E(r_t) = r$;

c) *proces* $\{r_t\}$ *nie jest autoskorelowany;*

d) $E(\pi_{t+1}|\mathcal{I}_t) = -r + R_t$.

Dowód.
Zauważmy, że z przyjętych założeń wynika co następuje

$$R_{t-1} - \pi_t = (R_{t-1} - \widehat{\pi}_{t,t-1}) - (\pi_t - \widehat{\pi}_{t,t-1}) = \widehat{r}_{t,t-1} - \eta_t = r - \eta_t.$$

Z drugiej strony

$$R_{t-1} - \pi_t = r_t,$$

zatem

$$r_t = r - \eta_t.$$

Ponieważ stopa r jest deterministyczna to

$$E(r_t) = r - E(\eta_t) = r.$$

Ponadto proces $\{\eta_t\}$ jest ciągiem przyrostów martyngałowych, zatem nie jest on autoskorelowany i to samo dotyczy procesu $\{r_t\}$. Punkt d wynika z faktu, że

$$\pi_{t+1} = R_t - r + \eta_{t+1}.$$

Zatem

$$E(\pi_{t+1}|\mathcal{I}_t) = E(R_t - r + \eta_{t+1}|\mathcal{I}_t) = R_t - r + E(\eta_{t+1}|\mathcal{I}_t) = R_t - r.$$

□

12.3. Analiza danych empirycznych

Fam poddał analizie dane z okresu styczeń 1953 – lipiec 1971 obejmującego 223 miesiące.

12.3.1. Test na autokorelację realnych stóp zwrotu

W oparciu o próbę z 223 miesięcy wyznaczamy współczynniki autokorelacji szeregu czasowego r_t, a następnie statystyki Q Ljunga-Boxa dla $j = 1, 2, \ldots, 12$ współczynników autokorelacji. Wyniki przedstawione są w tabeli 12.1. W ostatniej kolumnie są przedstawione "asymptotyczne" p-wartości (p-value) wyznaczone według wzoru

$$p_j = 1 - \chi_j^2(Q_j)$$

gdzie χ_j^2 dystrybuanta rozkładu chi kwadrat z j stopniami swobody (granicznego rozkładu Q_j gdy wielkość próbki rośnie do nieskończoności).

Otrzymane p-wartości p_j należą do przedziału $(0{,}9\%, 12{,}8\%)$ co można uznać za potwierdzenie hipotezy o braku autokorelacji.

j	$\widehat{\rho}_p$	Q	p-value
1	-0,101	2,3	0,128
2	0,172	9,1	0,011
3	-0,019	9,1	0,027
4	-0,004	9,1	0,058
5	-0,064	10,1	0,073
6	-0,021	10,2	0,117
7	-0,091	12,1	0,096
8	0,094	14,2	0,076
9	0,094	16,3	0,061
10	0,019	16,4	0,089
11	0,004	16,4	0,128
12	0,207	26,5	0,009

Tabela 12.1. Współczynniki autokorelacji i statystyki Q Ljunga-Boxa

12.3.2. Test predykcji stopy inflacji π w oparciu o nominalną stopę zwrotu R

Z założeń modelu wynika, że stopa inflacji π_{t+1} i nominalna stopa zwrotu R_t są związane zależnością liniową

$$\pi_{t+1} = R_t - r + \eta_{t+1},$$

gdzie r jest stałą, a η można interpretować jako składnik losowy.

Aby zweryfikować powyższą równość wyznaczymy parametry strukturalne modelu regresji z wyrazem wolnym

$$\pi_{t+1} = \beta_1 R_t + \beta_2 + \eta_{t+1}, \tag{12.1}$$

a następnie przetestujemy prawdziwość hipotezy

$$H_0 : \beta_1 = 1 \quad \text{wobec} \quad H_1 : \beta_1 \neq 1.$$

Sprawdzamy czy dla modelu opisanego równaniem 12.1 spełnione są założenia modelu dużej próbki $\tilde{\text{Z}}1$-$\tilde{\text{Z}}6$.

Warunek $\tilde{\text{Z}}1$.
Kładziemy $Y_t = \pi_{t+1}$, $X_t = (R_t, 1)$ i $\varepsilon_t = \eta_{t+1}$, otrzymujemy

$$Y_t = X_t\beta + \varepsilon_t.$$

Warunek $\tilde{\text{Z}}2$.
Stacjonarność i ergodyczność procesu $\{(Y_t, X_t)\}$ przyjmujemy („na wiarę") po analizie wykresu.

Warunek $\tilde{\text{Z}}3$.

$$X_t^T X_t = \begin{pmatrix} R_t \\ 1 \end{pmatrix} \cdot (R_t, 1) = \begin{pmatrix} R_t^2 & R_t \\ R_t & 1 \end{pmatrix}.$$

Zatem

$$\Sigma_{xx} = E(X_t^T X_t) = \begin{pmatrix} E(R_t^2) & E(R_t) \\ E(R_t) & 1 \end{pmatrix}.$$

Warunek maksymalnego rzędu jest spełniony gdyż

$$det(\Sigma_{xx}) = E(R_t^2) - E(R_t)^2 = D^2(R_t) > 0.$$

Warunek $\tilde{Z}4$.
Kładziemy

$$g_t = \varepsilon_t(R_t, 1) = (R_t\eta_{t+1}, \eta_{t+1}).$$

$E(\eta_{t+1}) = 0$ podobnie

$$E(R_t\eta_{t+1}) = E(E(R_t\eta_{t+1}|\mathcal{I}_t)) = E(R_tE(\eta_{t+1}|\mathcal{I}_t)) = E(0) = 0.$$

Zatem $E(g_t) = 0$ czyli spełniony jest warunek ortogonalności.
Warunek $\tilde{Z}5$.
Zauważmy, że dla każdego s zmienna losowa g_s jest $\mathcal{I}_{s+1}$ mierzalna, a więc

$$\mathcal{I}_t \supset \sigma(g_{t-1}, g_{t-2}, \dots).$$

Zatem

$$E(g_t|g_{t-1}, \dots) = E(E(g_t|\mathcal{I}_t)|g_{t-1}, \dots)) = E(0) = 0,$$

czyli proces $\{g_t\}$ jest ciągiem przyrostów martyngałowych.
Warunek $\tilde{Z}6$.
Warunkową homoskedastyczność przyjmujemy dla uproszczenia analizy modelu. Tym samym przyjmujemy odwracalność macierzy Σ_{gg}.

Na podstawie badanej próbki wyznaczamy estymator MNK parametrów β_1 i β_2

$$B \approx (1{,}0147, -0{,}8678).$$

Błąd standardowy B_1 wynosi 0,1227 zatem statystyka testowa

$$t_1 = \frac{B_1 - 1}{SB_1} \approx 0{,}12.$$

Zatem p-wartość wynosi około 90%, czyli bez problemu możemy zaakceptować hipotezę H_0.

12.3.3. Dyskusja wyników

Omówione powyżej testy potwierdzają hipotezę o efektywności rynku bonów w latach 1953 – 1971. Obserwacje w latach następnych nie potwiedziły już tej własności rynku. Można to tłumaczyć
- dużymi skokami inflacji
- i zdecydowaną interwencją FED na rynku.

Zauważmy ponadto, że powyższy przykład nie może być modelowany w terminach klasycznego modelu regresji gdyż nie jest spełniony warunek ścisłej egzogeniczności Z2.

13. Regresja względem czasu.

Teoria dużej próbki cd. Model tendencji rozwojowej z liniowym trendem. Estymatory hiper-zgodne.

13.1. Model tendencji rozwojowej z liniowym trendem

13.1.1. Założenia modelu

Rozważamy następujący model liniowy

$$Y_t = \beta_1 t + \beta_2 + \varepsilon_t, \tag{13.1}$$

gdzie
t – czas (kolejny moment), $t \in \mathbb{N}$ (lub $t \in \mathbb{Z}$),
$\{\varepsilon_t\}$ – niezależny, biały szum; tzn. ε_t nie zależą od historii i mają ten sam rozkład. Ponadto zakładamy, że $\varepsilon \in L^4$, $E(\varepsilon) = 0$ oraz $E(\varepsilon_t^2) = \sigma^2 > 0$.

Przyjmiemy następujące oznaczenia

$$X_t = (t, 1), \quad \beta = (\beta_1, \beta_2)^T.$$

Wówczas model 13.1 można zapisać w następujący sposób

$$Y_t = X_t \beta + \varepsilon_t.$$

W zapisie macierzowym otrzymamy

$$Y = X\beta + \varepsilon.$$

Powyższy model spełnia założenia modelu klasycznego Z1 – Z4 (bez założenia o normalności składnika losowego) i nie spełnia założeń modelu "dużej próbki", bo proces $\{t\}$ nie jest stacjonarny.

Problem.
Co można powiedzieć o asymptotyce estymatorów MNK dla modelu opisanego równaniem 13.1?

13.1.2. Estymacja parametrów modelu

Rozważmy proces generujący dane $\{Y_t, X_t\}_{t=0}^{\infty}$, z którego bierzemy n-elementową próbkę dla $t = 0, \ldots, n-1$. MNK estymator wektora β wyznaczamy ze wzoru

$$B = (X^T X)^{-1} X^T Y = S_{xx}^{-1} S_{xy},$$

gdzie

$$S_{xx} = \frac{1}{n} X^T X = \frac{1}{n} \sum_{t=0}^{n-1} X_t^T X_t, \quad S_{xy} = \frac{1}{n} X^T Y = \frac{1}{n} \sum_{t=0}^{n-1} X_t^T Y_t.$$

Natomiast MNK estymator wariancji σ^2 wynosi

$$S^2 = \frac{1}{n-2} \xi^t \xi = \frac{1}{n-2} \sum_{t=0}^{n-1} \xi_t^2, \quad \xi_t = Y_t - X_t B.$$

Twierdzenie 13.1.

$$\begin{pmatrix} \sqrt{n}^3 (B_1 - \beta_1) \\ \sqrt{n}(B_2 - \beta_2) \end{pmatrix} \xrightarrow{d} N(0, \sigma^2 \Sigma), \quad \Sigma = \begin{pmatrix} 12 & -6 \\ -6 & 4 \end{pmatrix}.$$

Dowód.
Zauważmy, że

$$B - \beta = S_{xx}^{-1} \bar{g}^T,$$

gdzie

$$g_t = \varepsilon_t X_t = (t\varepsilon_t, \varepsilon_t), \quad \bar{g} = \frac{1}{n} \sum_{t=0}^{n-1} g_t.$$

Macierz S_{xx} można łatwo wyliczyć

$$S_{xx} = \frac{1}{n} \sum_{t=0}^{n-1} X_t^T X_t = \frac{1}{n} \sum_{t=0}^{n-1} \begin{pmatrix} t \\ 1 \end{pmatrix} \circ (t, 1) =$$

$$= \frac{1}{n} \sum_{t=0}^{n-1} \begin{pmatrix} t^2 & t \\ t & 1 \end{pmatrix} = \begin{pmatrix} \frac{(n-1)(2n-1)}{6} & \frac{n-1}{2} \\ \frac{n-1}{2} & 1 \end{pmatrix}.$$

Gdy n rośnie do nieskończoności to 3 wyrazy macierzy S_{xx} zbiegają do nieskończoności

$$\begin{pmatrix} \frac{(n-1)(2n-1)}{6} & \frac{n-1}{2} \\ \frac{n-1}{2} & 1 \end{pmatrix} \longrightarrow \begin{pmatrix} \infty & \infty \\ \infty & 1 \end{pmatrix}.$$

Aby uzyskać rodzine macierzy o skończonej granicy pomnożymy macierz S_{xx} z obu stron przez macierz diagonalną

$$\Phi_n = \begin{pmatrix} n^{-1} & 0 \\ 0 & 1 \end{pmatrix}.$$

Otrzymujemy

$$\Phi_n S_{xx} \Phi_n = \begin{pmatrix} \frac{(n-1)(2n-1)}{6n^2} & \frac{n-1}{2n} \\ \frac{n-1}{2n} & 1 \end{pmatrix} \longrightarrow Q,$$

gdzie

$$Q = \begin{pmatrix} \frac{1}{3} & \frac{1}{2} \\ \frac{1}{2} & 1 \end{pmatrix}.$$

Z drugiej strony z centralnego twierdzenia granicznego Linderberga-Levy'ego ([9] §10.2 Twierdzenie 1) otrzymujemy asymptotyczną normalność $\Phi_n \bar{g}$

$$\sqrt{n} \Phi_n \bar{g}^T = \left(\sqrt{n}^{-3} \sum_{t=0}^{n-1} t\varepsilon_t, \sqrt{n}^{-1} \sum_{t=0}^{n-1} \varepsilon_t \right)^T \xrightarrow{d} N(0, \sigma^2 Q).$$

Podsumowując

$$\begin{pmatrix} \sqrt{n}^3(B_1-\beta_1) \\ \sqrt{n}(B_2-\beta_2) \end{pmatrix} = \sqrt{n}\Phi_n^{-1}(B-\beta) = \sqrt{n}\Phi_n^{-1}S_{xx}^{-1}\bar{g}^T =$$

$$= \sqrt{n}\Phi_n^{-1}S_{xx}^{-1}\Phi_n^{-1}\Phi_n\bar{g}^T = (\Phi_n S_{xx}\Phi_n)^{-1}) \cdot \sqrt{n}\Phi_n\bar{g}^T.$$

Ponieważ macierze $(\Phi_n S_{xx}\Phi_n)^{-1}$ zbiegają do Q^{-1}, a proces $\sqrt{n}\Phi_n\bar{g}^T$ zbiega według rozkładu do $N(0,\sigma^2 Q)$ to

$$\begin{pmatrix} \sqrt{n}^3(B_1-\beta_1) \\ \sqrt{n}(B_2-\beta_2) \end{pmatrix} \xrightarrow{d} N(0,\sigma^2 Q^{-1}QQ^{-1}) = N(0,\sigma^2 Q^{-1}).$$

Aby zakończyć dowód wystarczy zauważyć, że

$$Q^{-1} = \Sigma.$$

□

Uwaga 13.1. Estymator B_1 nazywa się estymatorem $n^{3/2}$-zgodnym albo hiper-zgodnym (hyper-consistent).

Twierdzenie 13.2. *Estymator S^2 jest zgodny*

$$S^2 \xrightarrow{p} \sigma^2.$$

Dowód.
Jak pokazaliśmy w rozdziale 4 (patrz równanie 4.3)

$$\xi^T\xi = \varepsilon^T M\varepsilon,$$

gdzie M macierz rzutu na dopełnienie ortogonalne podprzestrzeni rozpiętej przez kolumny macierzy X

$$M = Id - X(X^TX)^{-1}X^T.$$

Zatem

$$S^2 = \frac{1}{n-2}\xi^t\xi = \frac{1}{n-2}(\varepsilon^T\varepsilon - \varepsilon^T X(X^TX)^{-1}X^T\varepsilon) = \frac{1}{n-2}(\varepsilon^T\varepsilon - \varepsilon^T X(B-\beta)).$$

Stosując notację i oszacowania z dowodu poprzedniego twierdzenia otrzymujemy

$$S^2 = \frac{n}{n-2}\left(\frac{1}{n}\varepsilon^T\varepsilon - \bar{g}\Phi_n\Phi_n^{-1}(B-\beta)\right) \xrightarrow{p} \sigma^2 - 0.$$

□

13.1.3. Testowanie parametrów strukturalnych

Dla $k = 0, 1$ testujemy hipotezę

$$H_{k,0} : \beta_k = \bar{\beta}_k,$$

wobec hipotezy alternatywnej

$$H_{k,1} : \beta_k \neq \bar{\beta}_k.$$

Analogicznie jak w modelu klasycznym przyjmujemy

$$T_1 = \frac{B_1 - \bar{\beta}_1}{SE(B_1)}, \quad T_2 = \frac{B_2 - \bar{\beta}_2}{SE(B_2)},$$

gdzie

$$SE(B_k) = \sqrt{\frac{1}{n} S^2 (S_{xx}^{-1})_{k,k}}, \quad S^2 = \frac{1}{n-2} \sum_{t=0}^{n-1} \xi_t^2.$$

Twierdzenie 13.3. *Przy założeniu hipotezy zerowej $H_{k,0}$ rozkład statystyki T_k zbiega do N(0,1).*

Dowód.
Korzystamy z faktu, że S^2 zbiega do σ^2, $n^2(S_{xx}^{-1})_{1,1}$ do $\Sigma_{1,1}$, a $(S_{xx}^{-1})_{2,2}$ do $\Sigma_{2,2}$.

$$T_1 = \frac{B_1 - \bar{\beta}_1}{\sqrt{\frac{1}{n} S^2 (S_{xx}^{-1})_{1,1}}} = \frac{\sqrt{n}^3 (B_1 - \bar{\beta}_1)}{\sqrt{S^2 n^2 (S_{xx}^{-1})_{1,1}}} \xrightarrow{d} N(0, \Sigma_{1,1}^{-1} \Sigma_{1,1}) = N(0,1).$$

$$T_2 = \frac{B_2 - \bar{\beta}_2}{\sqrt{\frac{1}{n} S^2 (S_{xx}^{-1})_{2,2}}} = \frac{\sqrt{n} (B_2 - \bar{\beta}_2)}{\sqrt{S^2 (S_{xx}^{-1})_{2,2}}} \xrightarrow{d} N(0, \Sigma_{2,2}^{-1} \Sigma_{2,2}) = N(0,1).$$

□

14. Liniowe szeregi czasowe

Stacjonarność rzędu 2. Funkcje tworzące. Klasyczne modele liniowe MA, AR, ARMA i ARIMA. Ułamkowe ruchy Browna. (1 wykład)

14.1. Szeregi czasowe stacjonarne rzędu 2

Na początek ustalmy przestrzeń probabilistyczną

$$(\Omega, \mathcal{F}, P).$$

Przez $L^2(\Omega)$ będziemy oznaczać zbiór zmiennych losowych X takich, że $E(X^2) < \infty$. Jak łatwo zauważyć jest to przestrzeń liniowa nad $\mathbb{R}$. Ponadto, gdy utożsamimy zmienne losowe, które są prawie wszędzie równe to funkcja

$$\|X\|_{L^2} = \sqrt{E(X^2)}$$

jest normą, a $L^2(\Omega)/\sim$ jest przestrzenią Banacha (por. [9] dodatek C.5 lub [3] §2.10).

Dla ustalenia uwagi przyjmiemy następującą definicję szeregu czasowego.

Definicja 14.1. Szeregiem czasowym nazywamy ciąg zmiennych losowych X_t o wartościach rzeczywistych, $t \in \mathbb{Z}$

$$\ldots, X_{-n}, \ldots, X_{-1}, X_0, X_1, \ldots, X_n, \ldots.$$

Definicja 14.2. Szereg czasowy X_t nazywamy stacjonarnym rzędu 2 gdy

a. $\forall t \quad X_t \in L^2(\Omega)$;

b. $\forall t_1, t_2 \quad E(X_{t_1}) = E(X_{t_2})$;

c. $\forall t_1, t_2, n \quad cov(X_{t_1+n}, X_{t_2+n}) = cov(X_{t_1}, X_{t_2})$.

Zauważmy, że czasami stacjonarność rzędu 2 implikuje silną stacjonarność.

Lemat 14.1. *Gaussowski szereg czasowy stacjonarny rzędu 2 jest silnie stacjonarny w sensie definicji 8.6.*

Dowód.
Jeśli X_t jest procesem Gaussowskim to dla kazdego $s > 0$ łączny rozkład $(X_1, \ldots, X_s)$ jest rozkładem normalnym $N(\mu, \Sigma)$, gdzie

$$\mu_1 = \cdots = \mu_s = E(X_1), \quad \Sigma_{i,j} = cov(X_i, X_j).$$

Podobnie dla dowolnego n łączny rozkład $(X_{n+1}, \ldots, X_{n+s})$ jest rozkładem normalnym $N(\mu', \Sigma')$, gdzie

$$\mu'_1 = \cdots = \mu'_s = E(X_{n+1}), \quad \Sigma'_{i,j} = cov(X_{n+i}, X_{n+j}).$$

Ponieważ

$$E(X_{n+1}) = E(X_1), \quad \text{i} \quad cov(X_{n+i}, X_{n+j}) = cov(X_i, X_j),$$

to oba rozkłady są identyczne.

□

Podstawowymi narzędziami służącymi do opisu stacjonarnych szeregów czasowych są funkcje autokowariancji. Dla każdego stacjonarnego rzędu 2 szeregu czasowego $\{X_t\}_{t=-\infty}^{+\infty}$ definiujemy funkcję γ

$$\gamma : \mathbb{Z} \longrightarrow \mathbb{R},$$

przyporządkowującą liczbie całkowitej k - k-ty współczynnik autokowariancji (patrz definicja 11.1)

$$\gamma(k) = cov(X_0, X_k).$$

Zauważmy, że ze stacjonarności wynika, że dla każdego n

$$cov(X_n, X_{n+k}) = \gamma(k),$$

a w szczególności

$$D^2(X_n) = \gamma(0).$$

Twierdzenie 14.1. *Niech $k \geqslant 0$, wówczas funkcja autokowariancji stacjonarnego szeregu czasowego spełnia następujące warunki*

1. $\gamma(-k) = \gamma(k)$;
2. $\gamma(0) \geqslant 0$;
3. $|\gamma(k)| \leqslant \gamma(0)$;
4. $\forall a_{-k}, \ldots, a_k \in \mathbb{R} \quad \sum_{i,j=-k}^{k} a_i a_j \gamma(i-j) \geqslant 0.$

Uwaga 14.1. Funkcje spełniające warunek 4 dla każdego k nazywa się dodatnio określonymi.

Dowód.

Punkt 1 wynika z symetrii kowariancji.

$$\gamma(-k) = cov(X_0, X_{-k}) = cov(X_{-k}, X_0) = cov(X_0, X_k) = \gamma(k).$$

Punkt 2 jest oczywisty.

$$\gamma(0) = cov(X_0, X_0) = D^2(X_0) \geqslant 0.$$

Punkt 3 wynika ze związków między korelacją i kowariancją.

$$\gamma(k)^2 = cov(X_0, X_k)^2 \leqslant D^2(X_0) D^2(X_k) = \gamma(0)^2.$$

Punkt 4 wynika z nieujemności wariancji zmiennej losowej. Rozważmy zmienną losową

$$X_a = \sum_{i=-k}^{k} a_i X_i.$$

Korzystając z dwuliniowości kowariancji otrzymujemy

$$0 \leqslant D^2\left(\sum_{i=-k}^{k} a_i X_i\right) = \sum_{i,j=-k}^{k} a_i a_j cov(X_i, X_j) = \sum_{i,j=-k}^{k} a_i a_j \gamma(i-j).$$

□

Prawdziwe jest również twierdzenie odwrotne.

Twierdzenie 14.2. *Jeśli funkcja $\gamma : \mathbb{Z} \to \mathbb{R}$ spełnia warunki 1,2,3 i 4 z powyższego twierdzenia to jest ona funkcją autokowariancji pewnego stacjonarnego szeregu czasowego.*

Dowód
Patrz – [3] Theorem 1.5.1.

Zachowanie się funkcji autokowariancji dla dużych n ma istotne znaczenie w praktycznych zastosowaniach. Wyróżnia się dwie klasy szeregów czasowych stacjonarnych rzędu 2.

Definicja 14.3. Niech X będzie szeregiem czasowym stacjonarnym rzędu 2, a γ_X jego funkcją autokowariancji.
Gdy $\gamma_X(n)$ zbiega do zera dla dużych n w sposób wykładniczy

$$\exists C > 0 \quad \exists \delta \in (0,1) \quad |\gamma_X(n)| < C\delta^n,$$

to mówimy, że X jest procesem o krótkiej pamięci.
W przeciwnym przypadku mówimy, że X jest procesem o długiej pamięci.

Gdy szereg czasowy nie jest prawie na pewno stały czyli gdy $\gamma(0) > 0$ to definiujemy dodatkowo funkcję autokorelacji

$$\rho : \mathbb{Z} \longrightarrow [-1,1], \quad \rho(k) = \frac{\gamma(k)}{\gamma(0)}.$$

Kluczowym obiektem w teorii szeregów czasowych stacjonarnych rzędu 2 jest biały szum. Jest to przykład procesu o (bardzo!) krótkiej pamięci.

Definicja 14.4. Stacjonarny rzędu 2 szereg czasowy ε nazywamy białym szumem gdy

$$E(X_0) = 0 \quad \gamma(0) > 0 \quad \text{i} \ \gamma(t) = 0 \quad \text{dla } t \neq 0.$$

Gdy dodatkowo $\gamma(0) = 1$ to ε nazywamy unormowanym białym szumem.

Zbiór szeregów czasowych będących białym szumem o wariancji $\sigma^2 = \gamma(0)$ będziemy oznaczać $WN(\sigma^2)$.

Uwaga 14.2. Kolejne wyrazy białego szumu są nieskorelowane ale nie muszą być niezależne, chyba, że jest to gaussowski biały szum.

14.2. Sploty vel filtry

Na początek przypomnimy, kiedy ciąg liczb rzeczywistych jest klasy l^p.

Definicja 14.5. Ciąg $(a_n)_{n=0}^{\infty}$ należy do klasy l^p gdy jest sumowalny w p-tej potędze.

$$(a_n)_{n=0}^{\infty} \in l^p \Longleftrightarrow \sum_{n=0}^{\infty} |a_n|^p < \infty.$$

Uwaga 14.3. Zauważmy, że l^2 zawiera l^1. Ponadto dla $p \geqslant 1$ klasy l^p są unormowanymi przestrzeniami liniowymi

$$\|(a_n)\|_{l^p} = \left(\sum_{n=0}^{\infty} |a_n|^p\right)^{1/p}.$$

Definicja 14.6. Niech $X = (X_t)_{t=-\infty}^{+\infty}$ będzie szeregiem czasowym, a $a = (a_n)_{n=0}^{\infty}$ ciągiem o wyrazach rzeczywistych. Wówczas szereg czasowy $Y = (Y_t)_{t=-\infty}^{+\infty}$, którego wyrazy dają się przedstawić jako sumy nieskończone w $L^2(\Omega)$

$$Y_t = \sum_{i=0}^{\infty} a_i X_{t-i},$$

nazywamy splotem ciągu a i szeregu X.

Operację splotu będziemy oznaczać przez "$\star$"

$$Y = a \star X.$$

Uwaga 14.4. W niektórych źródłach operację splotu z ciągiem (a_n) nazywa się filtrem o współczynnikach a_n.

Podamy teraz dwa warunki gwarantujące istnienie splotu (Y_t) ciągu $a = (a_n)$ i szeregu (X_t).

Twierdzenie 14.3. *Jeżeli zachodzi jeden z poniższych warunków*

1. $(a_n) \in l^1$ *i* $\forall t \in \mathbb{Z}$ $|E(X_t)| \leqslant \mu$, $D^2(X_t) \leqslant \sigma^2$.
2. $(a_n) \in l^2$, $\forall t \in \mathbb{Z}$ $E(X_t) = 0$, $D^2(X_t) \leqslant \sigma^2$, *oraz* X_t *nieskorelowane ze sobą.*

to istnieje splot ciągu $a = (a_n)$ *i szeregu* (X_t).

Dowód.
Skorzystamy z zasady majoryzacji.
Ad 1.

$$\|Y_t\|_{L^2} \leqslant \sum_{i=0}^{\infty} \|a_i X_{t-i}\|_{L^2} = \sum_{i=0}^{\infty} |a_i| \|X_{t-i}\|_{L^2} \leqslant \sum_{i=0}^{\infty} |a_i| \sqrt{\sigma^2 + \mu^2} = \sqrt{\sigma^2 + \mu^2} \|a\|_{l^1}.$$

Ad 2.

$$\|Y_t\|_{L^2}^2 = D^2\left(\sum_{i=0}^{\infty} a_i X_{t-i}\right) = \sum_{i=0}^{\infty} a_i^2 D^2(X_{t-i}) \leqslant \sigma^2 \sum_{i=0}^{\infty} a_i^2 = \sigma^2 \|a\|_{l^2}^2.$$

□

Twierdzenie 14.4. *Gdy szereg czasowy* (X_t) *jest stacjonarny rzędu 2, a ciąg* $a = (a_n)$ *należy do* l^1 *to ich splot* (Y_t) *jest stacjonarny rzędu 2.*

Dowód.
Pokażemy, że wartość oczekiwana i autokowariancje szeregu (Y_t) nie zmieniają się przy przesunięciu. Oznaczmy przez μ wartość oczekiwaną X_t a przez γ_X funkcję autokowariancji.

$$E(Y_k) = E(\sum_{i=0}^{\infty} a_i X_{k-i}) = \sum_{i=0}^{\infty} a_i E(X_{k-i}) = \sum_{i=0}^{\infty} \mu a_i = \sum_{i=0}^{\infty} a_i E(X_{-i}) = E(Y_0).$$

$$cov(Y_n, Y_{k+n}) = \sum_{i,j=0}^{\infty} a_i a_j cov(X_{n-i}, X_{n+k-j}) = \sum_{i,j=0}^{\infty} a_i a_j \gamma_X(k+i-j) =$$

$$= \sum_{i,j=0}^{\infty} a_i a_j cov(X_{-i}, X_{k-j}) = cov(Y_0, Y_k).$$

□

Twierdzenie 14.5. *Gdy szereg czasowy (ε_t) jest białym szumem, a ciąg $a = (a_n)$ należy do l^2 to ich splot (Y_t) jest stacjonarny rzędu 2. Ponadto*

$$E(Y_t) = 0, \quad \gamma_Y(k) = \sigma^2 \sum_{i=0}^{\infty} a_i a_{k+i},$$

gdzie σ^2 wariancja ε_0.

Dowód.
Powtarzamy rozumowanie z poprzedniego dowodu i wstawiamy $E(\varepsilon_t) = 0$ oraz $\gamma_\varepsilon(0) = \sigma^2$ i $\gamma_\varepsilon(t) = 0$ dla $t \neq 0$.

□

Okazuje się, że z dokładnością do pewnego "nieistotnego" składnika, wszystkie szeregi czasowe stacjonarne rzędu 2 można przedstawić w postaci splotu z białym szumem.

Twierdzenie 14.6. *Niech X będzie szeregiem czasowym stacjonarnym rzędu 2 o zerowej wartości oczekiwanej, $E(X_0) = 0$. Wówczas istnieją ciąg a klasy l^2 i unormowany biały szum ε takie, że*

$$X = a \star \epsilon + V,$$

gdzie szereg czasowy V jest nieskorelowany z ε

$$\forall s, t \quad E(V_s \varepsilon_t) = 0$$

i ponadto jest zawarty w przecięciu domknięć (w L^2) podprzestrzeni generowanych przez początkowe X_s

$$\forall t \quad V_t \in \bigcap_{n=-\infty}^{+\infty} Cl_{L^2}(lin\{X_s; s \leqslant n\}).$$

Dowód.
Powyższe twierdzenie jest wnioskiem z "rozkładu Wolda" - [3] Theorem 5.7.1.

Operacja splotu przeprowadza szeregi czasowe stacjonarne rzędu 2 na szeregi czasowe stacjonarne rzędu 2. Zatem sploty można iterować.

Twierdzenie 14.7. *Niech ciągi a i b należą do przestrzeni l^1, a X będzie szeregiem czasowym stacjonarnym rzędu 2. Wówczas*

$$a \star (b \star X) = c \star X,$$

gdzie c jest ciągiem o wyrazach

$$c_n = \sum_{i=0}^{n} a_i b_{n-i}.$$

Uwaga 14.5. Ciąg c z powyższego twierdzenia nazywa się iloczynem Cauchy'ego ciągów a i b. Jak łatwo sprawdzić c też należy do l^1. Zatem splot definiuje działanie algebry l^1 na zbiorze szeregów czasowych, które są stacjonarne rzędu 2. Ponadto mnożenie Cauchy'ego jest przemienne gdyż

$$\sum_{i=0}^{n} a_i b_{n-i} = \sum_{j=0}^{n} a_{n-j} b_j,$$

zatem dla $a, b \in l^1$

$$b \star (a \star X) = a \star (b \star X).$$

Dowód twierdzenia.
Oznaczmy przez Y splot b i X, a przez Z splot a i Y. Wówczas

$$Y_t = \sum_{i=0}^{\infty} b_i X_{t-i}, \quad Z_t = \sum_{j=0}^{\infty} a_j Y_{t-j}.$$

Przedstawiamy szereg Z w zależności od szeregu X

$$Z_t = \sum_{j=0}^{\infty} a_j \left(\sum_{i=0}^{\infty} b_i X_{t-j-i} \right) = \sum_{i,j=0}^{\infty} a_j b_i X_{t-j-i}.$$

Podstawiamy $k = i + j$ i porządkujemy powyższą sumę względem X_{t-k}.

$$Z_t = \sum_{k=0}^{\infty} \sum_{j=0}^{k} a_j b_{k-j} X_{t-k} = \sum_{k=0}^{\infty} c_k X_{t-k}.$$

□

Gdy szereg X z twierdzenia 14.7 jest białym szumem to możemy osłabić założenia dotyczące ciągu b. W analogiczny sposób jak twierdzenie 14.7 dowodzi się następujące twierdzenie.

Twierdzenie 14.8. *Niech ciąg a należy do przestrzeni l^1, ciąg b do l^2, a ε będzie białym szumem. Wówczas*

$$a \star (b \star \varepsilon) = c \star \varepsilon,$$

gdzie c jest ciągiem o wyrazach

$$c_n = \sum_{i=0}^{n} a_i b_{n-i}.$$

14.3. Funkcje tworzące

Mnożenie Cauchy'ego ciągów jest ściśle związane z mnożeniem szeregów potęgowych. Otóż

$$\left(\sum_{n=0}^{\infty} a_n z^n\right)\left(\sum_{n=0}^{\infty} b_n z^n\right) = \left(\sum_{n=0}^{\infty} c_n z^n\right),$$

gdzie ciąg c jest iloczynem Cauchy'ego ciągów a i b

$$c_n = \sum_{i=0}^{n} a_i b_{n-i}.$$

Dlatego przyporządkujemy ciągom funkcje tworzące czyli sumy szeregów potęgowych. Pozwala to wykorzystać aparat analizy zespolonej do badania szeregów czasowych.

Definicja 14.7. Funkcję holomorficzną

$$A(z) = \sum_{n=0}^{\infty} a_n z^n$$

nazywamy funkcją tworzącą ciągu liczbowego $a = (a_n)_{n=0}^{\infty}$.

Zauważmy, że istnieje zależność między promieniem zbieżności R_A szeregu potęgowego $A(z)$ a klasą ciągu a.

Lemat 14.2.

$$\begin{aligned} R_A > 1 &\implies a \in l^1, \\ R_A < 1 &\implies a \notin l^2. \end{aligned}$$

Dowód.
Gdy promień zbieżności szeregu potęgowego A jest większy od 1 to jest on zbieżny bezwzględnie w punkcie $z_* = 1$. Zatem

$$\sum_{n=0}^{\infty} |a_n| = \sum_{n=0}^{\infty} |a_n| \cdot |z_*|^n < \infty.$$

Natomiast gdy ciąg a jest klasy l^2 to musi on zbiegać do 0

$$\lim_{n\to\infty} a_n = 0.$$

Zatem

$$\limsup_{n\to\infty} \sqrt[n]{a_n} \leqslant 1.$$

Z czego wynika, że promień zbieżności szeregu potęgowego A jest nie mniejszy niż 1

$$R_A = \left(\limsup_{n\to\infty} \sqrt[n]{a_n}\right)^{-1} \geqslant 1.$$

Więc jeśli promień zbieżności R_A jest mniejszy od 1, to ciąg a nie należy do l^2.

□

Funkcje tworzące dla funkcji autokowariancji określamy jako sumy szeregów Laurenta (por. [15] Cz. I, §6).

Definicja 14.8. Funkcję holomorficzną

$$\Gamma(z) = \sum_{-\infty}^{+\infty} \gamma(n) z^n$$

określoną na pierścieniu

$$\frac{1}{r} < |z| < r, \quad r > 1,$$

nazywamy funkcją tworzącą ciągu liczbowego $\gamma = (\gamma(n))_{-\infty}^{+\infty}$.

Uwaga 14.6. Współczynniki rozwinięcia funkcji Γ w szereg Laurenta na pierścieniu

$$\frac{1}{r} < |z| < r, \quad r > 1,$$

są wyznaczone jednoznacznie (por. [15] Cz. I, §6, Twierdzenie 2).

Lemat 14.3. *Niech γ będzie funkcją autokowariancji szeregu czasowego X stacjonarnego rzędu 2. Wówczas następujące warunki są równoważne:*
1. Szereg Laurenta $\Gamma(z)$ jest zbiezny w pewnym pierścieniu

$$\frac{1}{r} < |z| < r, \quad r > 1.$$

2. Szereg czasowy X jest procesem o krótkiej pamięci.

Dowód.
$1 \Rightarrow 2$.
Z nierówności Cauchy'ego dla współczynników szeregu Laurenta ([15] s.120) otrzymujemy, że dla każdego $\rho \in (1, r)$

$$\forall n \quad |\gamma(n)| \leqslant \frac{M}{\rho^n},$$

gdzie M to maksimum modułu funkcji Γ na okręgu $|z| = \rho$.

$2 \Rightarrow 1$.
Gdy szereg czasowy X jest procesem o krótkiej pamięci to istnieją stałe C i δ, $C > 0$, $\delta \in (0, 1)$, takie, że

$$\forall n \quad |\gamma(n)| < C\delta^{|n|}.$$

Zatem

$$\limsup_{n\to\infty} \sqrt[n]{\gamma(-n)} = \limsup_{n\to\infty} \sqrt[n]{\gamma(n)} \leqslant \delta < 1.$$

Zatem szereg Laurenta $\Gamma(z)$ jest zbieżny na pierścieniu

$$\delta < |z| < \frac{1}{\delta}$$

(por. [15] s.117).

□

14.4. Operator przesunięcia

Oznaczymy przez L operator przesunięcia szeregu czasowego o 1 w prawo

$$Y = L(X) \iff \forall t \in \mathbb{Z} \quad Y_t = X_{t-1}.$$

Pozwoli to nam zapisać splot $Y = a \star X$ jako szereg potęgowy iterowanych operatorów przesunięcia

$$\forall t \in \mathbb{Z} \quad Y_t = \sum_{i=0}^{\infty} a_i X_{t-i} \implies Y = \sum_{i=0}^{\infty} a_i L^i(X) = A(L)X.$$

Z twierdzeń 14.7 i 14.8 wynika następujący wniosek.

Wniosek 14.1. *Funkcja tworząca złożenia operatorów $A_1(L)$ i $A_2(L)$ jest iloczynem funkcji $A_1(z)$ i $A_2(z)$*

$$A_1(L) \circ A_2(L) = (A_1 \cdot A_2)(L).$$

Funkcja tworząca funkcji autokowariancji szeregu czasowego X będącego obrazem unormowanego białego szumu ε, $X = A(L)\varepsilon$ wynosi

$$\Gamma(z) = A(z)A(z^{-1}).$$

Uwaga 14.7. Gdy promień zbieżności szeregu potęgowego $A(z)$ jest większy od 1 to szereg $\Gamma(z) = A(z)A(z^{-1})$ jest zbieżny na pierścieniu

$$\{z : R_A^{-1} < |z| < R_A\}.$$

Z powyższej uwagi i lematu 14.3 otrzymujemy:

Wniosek 14.2. *Gdy promień zbieżności szeregu potęgowego $A(z)$ jest większy od 1 to szereg czasowy $X = A(L)\varepsilon$, $\varepsilon \in WN(\sigma^2)$, jest procesem o krótkiej pamięci.*

14.5. Przykłady

Przedstawimy teraz kilka najpopularniejszych szeregów czasowych stacjonarnych rzędu 2. Niech $\varepsilon = (\varepsilon_t)_{-\infty}^{+\infty}$ będzie unormowanym białym szumem.

1. Wielomianowa funkcja tworząca.
Szereg czasowy X postaci

$$X = A(L)\varepsilon,$$

gdzie $A(z)$ wielomian stopnia p nazywa się szeregiem średnich ruchomych rzędu p – $MA(p)$.

$$\forall t \quad X_t = a_0\varepsilon_t + a_1\varepsilon_{t-1} + \cdots + a_p\varepsilon_{t-p}.$$

2. Funkcja tworząca jest odwrotnością funkcji wielomianowej.
Szereg czasowy X postaci

$$X = C(L)\varepsilon, \quad C(z) = \frac{1}{B(z)},$$

gdzie $B(z)$ wielomian stopnia q, który nie zeruje się na kole jednostkowym

$$|z| \leqslant 1 \implies B(z) \neq 0,$$

nazywa się szeregiem autoregresyjnym rzędu q – $AR(q)$.
Szereg X spełnia równanie

$$B(L)X = \varepsilon,$$

zatem

$$\forall t \quad b_0X_t + b_1X_{t-1} + \cdots + b_qX_{t-q} = \varepsilon_t.$$

3. Funkcja tworząca jest ilorazem funkcji wielomianowych.
Szereg czasowy X postaci

$$X = D(L)\varepsilon, \quad D(z) = \frac{A(z)}{B(z)},$$

gdzie $A(z)$ wielomian stopnia p a $B(z)$ wielomian stopnia q, który nie zeruje się na kole jednostkowym

$$|z| \leqslant 1 \quad \Longrightarrow \quad B(z) \neq 0,$$

nazywa się autoregresyjnym szeregiem średnich ruchomych rzędu (q, p) – $ARMA(q, p)$.
Szereg X spełnia równanie

$$B(L)X = A(L)\varepsilon,$$

zatem

$$\forall t \quad b_0X_t + b_1X_{t-1} + \cdots + b_qX_{t-q} = a_0\varepsilon_t + a_1\varepsilon_{t-1} + \cdots + a_p\varepsilon_{t-p}.$$

Biorąc pod uwagę, że funkcje tworzące w trzech powyższych przykładach mają promień zbieżności większy od 1, to na mocy wniosku 14.2 otrzymujemy:

Lemat 14.4. *Szeregi czasowe MA, AR i $ARMA$ są procesami o krótkiej pamięci.*

Uwaga 14.8. Niech γ będzie funkcją autokowariancji szeregu czasowego $X = A(L)\varepsilon$, $\varepsilon \in WN(1)$. Gdy X jest klasy AR to operator A jest wyznaczony przez funkcje autokowariancji γ z dokładnością do znaku. Natomiast dla szeregów czasowych MA i $ARMA$ taka jednoznaczność zachodzi tylko przy dodatkowym warunku, że funkcje tworzące nie zerują się w kole jednostkowym.

Przykład $ARMA(1, 1)$**.**
Rozważmy szereg czasowy

$$X = D(L)\varepsilon, \quad D(z) = \frac{3z+1}{z+3}, \quad \varepsilon \in WN(1).$$

Funkcja $D(z)$ jest holomorficzna na całej płaszczyźnie zespolonej poza punktem $z = -3$ gdzie ma biegun. Zatem promień zbieżności jej szeregu Taylora w 0 wynosi 3.
Okazuje się, że funkcja tworząca funkcji autokowariancji jest stała

$$\Gamma(z) = D(z)D(z^{-1}) = \frac{3z+1}{z+3}\frac{3z^{-1}+1}{z^{-1}+3} = \frac{3z+1}{z+3}\frac{3+1z}{1+3z} = 1.$$

Czyli szereg X jest białym szumem. Zatem reprezentacja $ARMA$ szeregu czasowego o zadanej funkcji autokowariancji nie jest jednoznaczna.

14.6. Procesy o przyrostach stacjonarnych

Niech Y będzie dowolnym szeregiem czasowym. Szereg X taki, że

$$\forall t \quad X_t = Y_t - Y_{t-1},$$

czyli

$$X = (1 - L)Y,$$

nazywamy szeregiem przyrostów szeregu Y.

Operację brania przyrostów można iterować. Szereg

$$X = (1 - L)^k Y = P(L)Y, \quad P(z) = (1 - z)^k,$$

nazywamy szeregiem k-tych przyrostów szeregu Y.

Gdy k-te przyrosty szeregu czasowego Y są stacjonarne rzędu 2 i należą do klasy $ARMA(q, p)$ to mówimy, że szereg Y jest klasy $ARIMA(q, k, p)$.

Przykłady

• Błądzenie przypadkowe bez dryfu

$$X_t = X_{t-1} + \sigma\varepsilon_t$$

jest procesem klasy $ARIMA(0, 1, 0)$.

• Błądzenie przypadkowe z dryfem

$$X_t = X_{t-1} + \mu + \sigma\varepsilon_t$$

jest procesem klasy $ARIMA(0, 2, 1)$.

• Trend liniowy

$$X_t = at + b + \sigma\varepsilon_t$$

jest procesem klasy $ARIMA(0, 2, 2)$.

• Trend wielomianowy stopnia k

$$X_t = a_k t^k + a_{k-1} t^{k-1} + \ldots a_0 + \sigma\varepsilon_t$$

jest procesem klasy $ARIMA(0, k + 1, k + 1)$.

14.7. Ułamkowy ruch Browna

Definicja 14.9. Gaussowski proces B_t^H, $t \in \mathbb{Z}$, $H \in (0, 1)$, taki, że

$$\forall t \quad E(X_t) = 0;$$

$$\forall t, s \quad cov(B_t^H, B_s^H) = \frac{\sigma^2}{2}\left(|t|^{2H} + |s|^{2H} - |t - s|^{2H}\right), \quad \sigma > 0,$$

nazywamy ułamkowym ruchem Browna.

Pokażemy, że szereg B^H jest procesem o przyrostach stacjonarnych. Niech

$$X_t = B_t^H - B_{t-1}^H.$$

Lemat 14.5. *Szereg czasowy X_t^H jest stacjonarny o zerowej wartości oczekiwanej, $E(X_t^H) = 0$, oraz funkcji autokowariancji*

$$\gamma(n) = \frac{\sigma^2}{2}\left(|n + 1|^{2H} + |n - 1|^{2H} - 2|n|^{2H}\right).$$

Dowód.
Ponieważ B^H jest szeregiem o zerowej wartości oczekiwanej to to samo zachodzi dla szeregu przyrostów. Natomiast

$$cov(X_t^H, X_{t+n}^H) = cov(B_t^H - B_{t-1}^H, B_{t+n}^H - B_{t+n-1}^H) =$$

$$= \frac{\sigma^2}{2}\Big((|t|^{2H} + |t+n|^{2H} - |n|^{2H}) + (|t-1|^{2H} + |t+n-1|^{2H} - |n|^{2H}) -$$

$$-(|t+n|^{2H} + |t-1|^{2H} - |n+1|^{2H}) - (|t+n-1|^{2H} + |t|^{2H} - |n-1|^{2H})\Big) =$$

$$= \frac{\sigma^2}{2}\Big(|n+1|^{2H} + |n-1|^{2H} - 2|n|^{2H}\Big).$$

Zatem X^H jest stacjonarny z funkcją autokowariancji

$$\gamma(n) = \frac{\sigma^2}{2}\Big(|n+1|^{2H} + |n-1|^{2H} - 2|n|^{2H}\Big).$$

□

Uwaga 14.9. Dla wszystkich X^H $\gamma(0) = \sigma^2$. Dodatkowo dla $H = \frac{1}{2}$ i $n \neq 0$ $\gamma(n) = 0$. Zatem $X^{1/2}$ jest gaussowskim białym szumem, a $B^{1/2}$ jest gaussowskim błądzeniem przypadkowym.

Lemat 14.6. *Dla* $H \neq \frac{1}{2}$

$$\lim_{n\to\infty} \gamma(n) n^{2-2H} = H(2H-1)\sigma^2.$$

Dowód.
Dla $n > 1$

$$\frac{\gamma(n)}{\sigma^2} = \frac{1}{2}\Big((n+1)^{2H} + (n-1)^{2H} - 2n^{2H}\Big) = \frac{n^{2H}}{2}\Big((1+\frac{1}{n})^{2H} + (1-\frac{1}{n})^{2H} - 2\Big) =$$

$$= \frac{n^{2H}}{2}\Big(1 + \frac{2H}{n} + \frac{2H(2H-1)}{2n^2} + 1 - \frac{2H}{n} + \frac{2H(2H-1)}{2n^2} - 2 + O(n^{-3})\Big) =$$

$$= H(2H-1)n^{2H-2} + O(n^{2H-3}).$$

□

Uwaga 14.10. Gdy $H \neq \frac{1}{2}$, to szereg czasowy X^H jest procesem stacjonarnym o długiej pamięci.

15. Nieliniowe szeregi czasowe

Modele uwzględniające heteroskedastyczność - GARCH. (1 wykład)

15.1. Wstęp.

Zmienność odchyleń standardowych (heteroscedasticity) zwrotów finansowych powoduje, że do ich opisu należy stosować bardziej skomplikowane modele stochastyczne niż model błądzenia przypadkowego, na przykład modele z rodziny GARCH. Są to modele *nieliniowe*. Ponadto, oprócz badanej wielkości, np. przyrostów logarytmicznych kursów walutowych, wprowadza się zmienne pomocnicze, których nie można bezpośrednio mierzyć.

Najprostszy model z tej rodziny, GARCH(1,1), jest opisany następująco:
Rozważamy dwa ciągi zmiennych losowych r_t i h_t. Wartość r_t poznajemy w momencie t, a h_t jest zmienną pomocniczą. Są one związane wzorami

$$r_t = \sqrt{h_t} \cdot \varepsilon_t,$$

$$h_t = a + bh_{t-1} + cr_{t-1}^2, \quad t \in \mathbb{Z},$$

gdzie a, b, c są parametrami modelu a ε_t są niezależnymi zmiennymi losowymi o tym samym rozkładzie o wartości oczekiwanej 0 i wariancji 1.
$\sqrt{h_t}$ można interpretować jako *zmienne* odchylenie standardowe zmiennych losowych r_t.

Indeksy 1,1 w nazwie modelu oznaczają, że h_t zależy liniowo od h_{t-1} i r_{t-1}^2. W literaturze są badane również modele GARCH(p,q)

$$r_t = \sqrt{h_t} \cdot \varepsilon_t, \quad t \in \mathbb{Z},$$

$$h_t = a + b_1 h_{t-1} + \cdots + b_p h_{t-p} + c_1 r_{t-1}^2 + \cdots + c_q r_{t-q}^2.$$

15.2. Ogólne własności modelu GARCH(1,1).

Model GARCH pozwala w prosty sposób wyliczać warunkową wartość oczekiwaną i warunkowe momenty zmiennej r_t. Dla uproszczenia przyjmiemy, że inowacje ε_t mają rozkład normalny

$$\forall t \quad \varepsilon_t \sim N(0, 1).$$

Oznaczmy przez $E_t(\cdot)$ wartość oczekiwaną wyznaczoną gdy znane są już wartości r_i dla $i \leqslant t$. Zauważmy, że dla każdego t h_t i ε_t są niezależne zatem

$$E_t(r_{t+k}^{2m-1}) = E_t(h_{t+k}^{m-1/2}) \cdot E_t(\varepsilon_{t+k}^{2m-1}) = 0,$$

$$E_t(r_{t+k}^{2m}) = E_t(h_{t+k}^m) \cdot E_t(\varepsilon_{t+k}^{2m}) = 1 \cdot 3 \cdot \dots \cdot (2m-1) \cdot E_t(h_{t+k}^m).$$

Wynika, to z założenia, ε_t nie zależą od *historii* i mają rozkład normalny $N(0,1)$

$$E_t(\varepsilon_{t+k}^m) = E(\varepsilon_{t+k}^m) = \begin{cases} 0 & m = 1,3,5,7,\dots \\ 1 \cdot 3 \cdot 5 \cdot \dots \cdot (m-1) & m = 2,4,6,\dots \end{cases}$$

(patrz [12] str.60 lub [11] §18.8-3.)

Zauważmy, że h_{t+1} jest wyznaczone przez h_1 i wartości *historyczne* r_i, $i \leqslant t$. Zatem przy prognozowaniu o jeden krok naprzód, mamy

$$E_t(r_{t+1}^{2m}) = h_{t+1}^m \cdot 1 \cdot 3 \cdot \dots (2m-1)$$

gdzie $h_{t+1} = a + bh_t + cr_t^2$.

Dla $k > 1$ korzystamy ze wzoru

$$h_{t+1} = a + bh_t + ch_t\varepsilon_t^2.$$

Co daje nam

$$E_t(h_{t+k}) = a + (b+c)E_t(h_{t+k-1}).$$

Oznaczmy przez A_k warunkową wartość oczekiwaną $E_t(h_{t+k})$. Zależność rekurencyjna

$$A_k = a + (b+c)A_{k-1}, \quad k = 2,3,\dots, \quad A_1 = h_{t+1} > 0$$

wyznacza jednoznacznie A_k.
Jeśli $b + c = 1$ to

$$A_k = (k-1)a + A_1;$$

gdy $b + c \neq 1$ to

$$A_k = \frac{a}{1-b-c} + (b+c)^{k-1}(A_1 - \frac{a}{1-b-c}).$$

Zauważmy, że w przypadku gdy $b + c = 1$ i $a = 0$ ciąg A_k jest stały, a gdy $b + c = 1$ i $a > 0$ rozbieżny liniowo do nieskończoności.
Gdy $|b + c| < 1$ to ciąg A_k jest zbieżny do

$$A_\infty = \frac{a}{1-b-c},$$

a gdy $|b + c| > 1$ to ciąg A_k jest rozbieżny wykładniczo.

Aby wyznaczyć czwarty moment r_t, czyli drugi h_t, korzystamy ze wzoru

$$h_{t+1}^2 = a^2 + 2a(b + c\varepsilon_t^2) \cdot h_t + (b + c\varepsilon_t^2)^2 h_t^2.$$

Otrzymujemy

$$E_t(h_{t+k}^2) = a^2 + 2a(b + cE_t(\varepsilon_{t+k-1}^2)) \cdot E_t(h_{t+k-1}) + (b^2 + 2bcE_t(\varepsilon_{t+k-1}^2) + c^2E_t(\varepsilon_{t+k-1}^4) \cdot E_t(h_{t+k-1}^2).$$

Oznaczmy przez B_k warunkowy moment $E_t(h_{t+k}^2)$. Wówczas mamy zależność

$$B_k = a^2 + 2a(b+c)A_{k-1} + ((b+c)^2 + 2c^2)B_{k-1}, \quad B_1 = h_{t+1}^2 \neq 0.$$

Gdy $b + c = 1$ to

$$B_k = a^2 + 2a((k-1)a + A_1) + (1 + 2c^2)B_{k-1}.$$

Zatem gdy $c \neq 0$ to ciąg B_k jest rozbieżny. W szczególności model wzorowany na Risk Metrics ($a = 0$, $b = 0,94$ i $c = 0,06$) jest rozbieżny.

Zauważmy, że ograniczność czwartego momentu jest zagwarantowana gdy $(b+c)^2 + 2c^2 < 1$. Wówczas B_k zbiega do

$$B_\infty = \frac{a^2(1+b+c)}{(1-b-c)(1-(b+c)^2-2c^2)}.$$

Otrzymujemy w ten sposób następujący wzór na *asymptotyczną* kurtozę:

$$\frac{3B_\infty}{A_\infty^2} = \frac{3(1-(b+c)^2)}{1-(b+c)^2-2c^2} = 3 + \frac{6c^2}{1-(b+c)^2-2c^2}.$$

15.3. Ograniczenia na parametry modelu GARCH(1,1).

Ograniczenia na parametry modelu wynikają z naturalnych założeń dotyczących ograniczoności procesu. Zakładamy, że istnieją granice warunkowych wartości oczekiwanych

$$\lim_{k\to\infty} E_t(h_{t+k}), \quad \text{i} \quad \lim_{k\to\infty} E_t(h_{t+k}^2).$$

Pierwsza granica istnieje gdy $b + c < 1$ i zachodzi wówczas

$$\lim_{k\to\infty} E_t(h_{t+k}) = \frac{a}{1-b-c}.$$

Druga gdy $(b+c)^2 + 2c^2 < 1$, wówczas

$$\lim_{k\to\infty} E_t(h_{t+k}^2) = \frac{a^2(1+b+c)}{(1-b-c)(1-(b+c)^2-2c^2)}.$$

Natomiast założenie o rozkładzie ε_t możemy osłabić. Istotne są tylko następujące warunki

$$E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = 1, \quad E(\varepsilon_t^3) = 0, \quad E(\varepsilon_t^4) = 3.$$

15.4. Stacjonarność modeli GARCH.

Nie dla wszystkich modeli GARCH istnieją rozwiązania stacjonarne. Potrzebne są dodatkowe warunki na parametry (patrz [14] §3.3.1). Przykładowo dla modelu GARCH(1,1) zachodzi:

Twierdzenie 15.1. *Następujące warunki są równoważne:*
1. Model GARCH(1,1) z parametrami a, b, c ma dokładnie jedno nieujemne rozwiązanie stacjonarne (r_t, h_t).
2. Parametry a, b, c są nieujemne i spełnione jest oszacowanie

$$E(ln(b\varepsilon_t^2 + c)) < 0.$$

Momenty rozwiązania stacjonarnego można stosunkowo łatwo wyznaczyć.

Twierdzenie 15.2. *Niech (r_t, h_t) będzie stacjonarnym procesem GARCH(1,1) z parametrami a, b, c. Przy założeniu*

$$E(\varepsilon_t) = 0, \quad E(\varepsilon_t^2) = 1, \quad E(\varepsilon_t^3) = 0, \quad E(\varepsilon_t^4) = 3,$$

otrzymujemy:
A. Gdy $b + c < 1$ to

1. $E(h_t) = \dfrac{a}{1-b-c}$;
2. $E(r_t) = 0$;
3. $D^2(r_t) = E(r_t^2) = \dfrac{a}{1-b-c}$;
4. $cov(r_t, r_{t+k}) = 0 \quad k = 1, 2, \ldots$.

B. Gdy ponadto $(b + c)^2 + 2c^2 < 1$ to

5. $E(h_t^2) = \dfrac{a^2(1+b+c)}{(1-b-c)(1-(b+c)^2-2c^2)}$;
6. $D^2(h_t) = \dfrac{2a^2c^2}{(1-b-c)^2(1-(b+c)^2-2c^2)}$;
7. $cov(h_t, h_{t+k}) = (b+c)^k D^2(h_t) \quad k = 1, 2, \ldots$;
8. $E(r_t^4) = 3E(h_t^2)$;
9. $D^2(r_t^2) = \dfrac{2a^2(1-b^2-bc)}{(1-b-c)^2(1-(b+c)^2-2c^2)}$;
10. $cov(r_t^2, r_{t+1}^2) = \dfrac{c(1-b^2-bc)}{1-b^2-2bc} D^2(r_t^2)$;
11. $cov(r_t^2, r_{t+k}^2) = (b+c)^{k-1} cov(r_t^2, r_{t+1}^2) \quad k = 2, 3, \ldots$.

Literatura

[1] A.Białynicki-Birula, *Algebra liniowa z geometrią.* PWN 1976.
[2] P.Bilingsley, *Convergence of probability measures.* Wiley 1968.
[3] P.J.Brockwell, R.A.Davis, *Time Series. Theory and Methods.* Springer, New York 1991.
[4] M.Doman, R.Doman, *Modelowanie zmienności i ryzyka*, Wolters Kluwer, Kraków 2009.
[5] A.Goryl, Z.Jędrzejczyk, K.Kukuła, J.Osiewalski, A.Walkosz, *Wprowadzenie do ekonometrii.* PWN 2000.
[6] W.H.Greene, *Econometric Analysis.* Prentice Hall 2000.
[7] M.Gruszczyński, M.Podgórska (red.), *Ekonometria.* Oficyna Wydawnicza SGH 1996.
[8] F.Hayashi, *Econometrics.* Princeton University Press 2000.
[9] J.Jakubowski, R.Sztencel, *Wstęp do teorii prawdopodobieństwa.* SCRIPT, Warszawa 2000.
[10] P.Jaworski, J.Micał, *Modelowanie matematyczne w finansach i ubezpieczeniach,* Poltext, Warszawa 2005.
[11] G.Korn, T.Korn, Mathematical Handbook, McGraw-Hill Book Company 1968.
[12] R.Magiera, *Modele i metody statystyki matematycznej.* Oficyna Wydawnicza GiS, Wrocław 2002.
[13] P.A.Ruud, *An Introduction to Classical Econometric Theory.* Oxford University Press 2000.
[14] D.Straumann, *Estimation in Conditionally Heteroscedastic Time Series Models.* Springer, Berlin, Heidelberg, New York 2005.
[15] B.W.Szabat, *Wstęp do analizy zespolonej.* PWN, Warszawa 1974.
[16] A.W. van der Vaart, *Asymptotic statistics.* Cambridge University Press, New York 1998.
[17] W.Welfe, A.Welfe, *Ekonometria stosowana.* PWE 1996.

Literatura