



Joanna Małgorzata Landmesser

EKONOMETRIA

Literatura podstawowa:

1. Borkowski B., Dudek H., Szczesny W.: „Ekonometria. Wybrane zagadnienia”, PWN, Warszawa 2003.
2. Gruszczyński M., Kluza S., Winek D.: „Ekonometria”, WSHiFM, Warszawa 2003.
3. Kufel T.: „Ekonometria. Rozwiązywanie problemów z wykorzystaniem programu gretl”, PWN, Warszawa 2004.

Literatura uzupełniająca:

1. Maddala G. S.: „Ekonometria”, PWN, Warszawa 2006.
2. Welfe A.: „Ekonometria”, PWE, Warszawa 1995.
3. Nowak E.: „Zarys metod ekonometrii. Zbiór zadań”, PWN, Warszawa 2002.
4. Zeliaś A., Pawełek B., Wanat S.: „Prognozowanie ekonomiczne. Teoria, przykłady, zadania”, PWN, Warszawa 2003.
5. Jędrzejczyk Z., Kukuła K., Skrzypek J., Walkosz A.: „Badania operacyjne w przykładach i zadaniach”, PWN, Warszawa 2004.

Spis treści

I. Ekonometria	3
1. Przedmiot ekonometrii	3
2. Model ekonometryczny	3
3. Klasyfikacja zmiennych modelowych	4
4. Klasyfikacja modeli ekonometrycznych	4
5. Etapy modelowania ekonometrycznego	5
II. Dobór zmiennych objaśniających do modelu ekonometrycznego	6
1. Zasady doboru zmiennych	6
2. Metoda eliminacji zmiennych quasi-stałych	6
3. Metoda analizy macierzy współczynników korelacji	6
4. Metoda Hellwiga doboru zmiennych do modelu	8
III. Estymacja modelu regresji liniowej z jedną zmienną objaśniającą	9
1. Pojęcie estymacji i estymatora	9
2. Specyfikacja zależności	9
3. Metoda najmniejszych kwadratów	10
4. Metoda momentów	12
5. Interpretacja współczynników równania	13
IV. Weryfikacja modelu	14
1. Ocena jakości modelu	14
2. Badanie istotności zmiennych objaśniających modelu	16
3. Weryfikacja reszt	19
V. Analiza rozwoju zjawisk w czasie	24
1. Szeregi czasowe	24
2. Modele trendu	25
3. Wygładzanie szeregu czasowego metodą średniej ruchomej	26
VI. Prognozowanie ekonometryczne	29
1. Prognoza	29
2. Błędy prognozy	30
3. Prognozowanie na podstawie modeli tendencji rozwojowej	31
VII. Programowanie liniowe	33
1. Model procesu decyzyjnego	33
2. Programowanie liniowe	33
3. Dualność w programowaniu liniowym	38
VIII. Analiza przepływów międzygałęziowych	39
1. Tabele przepływów międzygałęziowych	39
2. Model Leontiefa	42



I. EKONOMETRIA

1. Przedmiot ekonometrii
2. Model ekonometryczny
3. Klasyfikacja zmiennych modelowych
4. Klasyfikacja modeli ekonometrycznych
5. Etapy modelowania ekonometrycznego

1. PRZEDMIOT EKONOMETRII

Termin **ekonometria** w szerszym rozumieniu oznacza zespół metod matematycznych i statystycznych stosowanych do badań ekonomicznych. W węższym znaczeniu to specyficzne metody statystyczne stosowane w badaniach nieeksperymentalnych. Oskar Lange podał następującą definicję ekonometrii: „nauka o matematyczno-statystycznych metodach modelowania i analizy zjawisk ekonomicznych”.

Ekonometria zajmuje się badaniem obserwacji empirycznych za pomocą statystycznych metod estymacji i testowania hipotez. Należy podkreślić, że badania empiryczne i teoretyczne wzajemnie się uzupełniają. Analizy teoretyczne poddawane są weryfikacji przy pomocy danych empirycznych. Z drugiej strony do badań statystycznych potrzebne są wskazówki płynące z teorii ekonomicznych.

2. MODEL EKONOMETRYCZNY

Model to uproszczone odwzorowanie związków zachodzących w wyróżnionym fragmencie rzeczywistości. Podstawowym zaś obiektem rozpatrywanym w dziedzinie ekonometrii jest model ekonometryczny.

Zgodnie z definicją Władysława Welfego **model ekonometryczny** opisuje zależności zachodzące w sferze zjawisk ekonomicznych, przy czym odwzorowanie następuje w języku matematycznym, a wśród zmiennych występuje przynajmniej jedna zmienna, będąca zmienną losową.

Taki formalny zapis procesu lub zjawiska ekonomicznego przyjmuje formę równania lub kilku równań, ewentualnie nierówności, wiążących rozważane zmienne. Pojedyncze równanie stochastyczne zbudowane jest ze zmiennej objaśnianej, zmiennych objaśniających, parametrów strukturalnych oraz składnika losowego.

Przykład 1.

Wysokość wynagrodzeń ustalana jest w wyniku negocjacji między pracodawcami a pracownikami. Istotną rolę odgrywa w nich stopa inflacji, osiągnięty przyrost wydajności pracy, skala napięć na rynku pracy (np. stopa bezrobocia) oraz inne czynniki. Model ekonometryczny dla rozważanego zjawiska mógłby przyjąć postać:

$$y_t = f(x_{1t}, x_{2t}, x_{3t}, \varepsilon_t)$$

- y_t - wynagrodzenie przeciętne,
- f - dowolna funkcja liniowa lub nieliniowa,
- x_{1t} - wskaźnik cen detalicznych dóbr konsumpcyjnych,
- x_{2t} - wydajność pracy,
- x_{3t} - stopa bezrobocia,
- ε_t - składnik losowy, będący m.in. sumaryczną charakterystyką efektów pozostałych czynników wpływających na wynagrodzenie,
- t - czas.

3. KLASYFIKACJA ZMIENNYCH MODELOWYCH

W literaturze przedmiotu używa się kilku różnych nazw dla zmiennych y, x_1, x_2, \dots, x_k .

Ze względu na rolę przez nie pełnioną w ekonometrycznym modelu regresji, można mówić o zmiennej **objaśnianej** (y) i zmiennych **objaśniających** (x_1, x_2, \dots, x_k). Często używa się również nazw zmienna **zależna** i zmienne **niezależne** lub mówi się o **regresancie** i **regresorach**. W wypadku analizy przyczynowości wskazuje się na zmienną **skutkową** i zmienne **przyczynowe**. Do celów predykcyjnych stosuje się terminy: zmienna **prognozowana** i **predykatory**.

W ekonometrii modeli wielorównaniowych ze względu na źródło zmienności wyróżniamy zmienne **endogeniczne**, czyli zmienne wyjaśniane na podstawie modelu, oraz zmienne **egzogeniczne**, których wartości są dane, ustalone, nie wyjaśniane przez model. Można wtedy również wskazać zmienne **łącznie współzależne** i zmienne **z góry ustalone**. W wypadku modelowania problemów związanych z zarządzaniem pojawiają się zmienna **celowa** i zmienne **kontrolowane**.

Generalnie zmienne występujące w modelach ekonometrycznych mogą mieć charakter zarówno zmiennych **ilościowych**, jak i **jakościowych** (np. zmienne zero-jedynkowe). Ponadto w dynamicznych modelach ekonometrycznych mamy do czynienia ze zmiennymi **bieżącymi** oraz **opóźnionymi**.

4. KLASYFIKACJA MODELI EKONOMETRYCZNYCH

1. Ze względu na liczbę równań modele ekonometryczne dzielimy na:

- modele **jednorównaniowe** – opisują pojedyncze procesy ekonomiczne, zbudowane są z pojedynczego równania,
- modele **wielorównaniowe** – opisują złożone procesy ekonomiczne, składają się z wielu równań (lub nierówności).

2. Ze względu na postać analityczną modele dzielimy na:

- modele **liniowe** – są z natury rzeczy proste (np. $y_t = \beta_0 + \beta_1 x_{1t} + \beta_2 x_{2t} + \dots + \beta_k x_{kt} + \varepsilon_t$); dla tej klasy modeli istnieje szeroka gama metod estymacji i weryfikacji hipotez,
- modele **nielinowe** – ma tu miejsce różnorodność form wynikająca ze skomplikowanego charakteru związków pomiędzy zjawiskami ekonomicznymi.

3. Ze względu na rolę odgrywaną przez czas modele ekonometryczne dzielimy na:

- modele **statyczne**: $y_t = f(x_{1t}, x_{2t}, \dots, x_{kt}, \varepsilon_t)$ – w nich związek ma charakter jednoczesny, więc wszystkie subskrypty mają wartość t , a w zbiorze zmiennych objaśniających nie występuje wyraźnie czas¹,
- modele **dynamiczne** – w nich jedną ze zmiennych objaśniających jest czas (t) lub występują zmienne opóźnione bądź przyspieszone (do klasy tej należą m.in. modele **trendu**, modele **autoregresyjne**).

4. Ze względu na poznawcze cechy modele ekonometryczne dzielimy na:

- modele **przyczynowo-opisowe** – ich zadaniem jest poznanie struktury badanych zjawisk ekonomicznych; występujące w nich zmienne objaśniające mają charakter obiektywnych przyczyn w stosunku do wyróżnionych zmiennych endogenicznych,
- modele **symptomatyczne** – przynajmniej niektóre spośród uwzględnionych w nich zmiennych objaśniających nie stanowią bezpośredniej przyczyny dla zmiennej endogenicznej danego równania, ale są z nią silnie skorelowane,

¹ W wypadku danych przekrojowych dla wyróżnienia obiektów, do których odnosi się analizowany związek, stosuje się zamiast t subskrypty i .



- c. modele **tendencji rozwojowej** – opisują one za pomocą funkcji zmiennej czasowej t wahania zmiennych endogenicznych w czasie (tu: modele **trendu**, modele **wahań sezonowych**),
- d. modele **autoregresyjne** – tutaj zmienna endogeniczna wyjaśniana jest za pomocą opóźnionych swych własnych wartości (np. $y_t = f(y_{t-1}, x_{1t}, x_{2t}, \dots, x_{kt}, \varepsilon_t)$),
- e. modele **adaptacyjne** – dopuszczają możliwość zmian struktury ekonomicznej; budowane są w sposób sekwencyjny, w miarę napływu nowych informacji statystycznych.

W modelowaniu ekonometrycznym zaznaczają się dwie tendencje:

- a. modelowanie **strukturalne** – będące rozwinięciem modelowania matematyczno-ekonomicznego, nastawione na poznanie mechanizmów rozwojowych, przyczyniające się do weryfikacji i rozwoju teorii ekonomicznej,
- b. modelowanie **predyktywne** – nie prowadzące do rozwoju teorii ekonomicznej, ale tego typu modele odgrywają istotną rolę praktyczną.

5. ETAPY MODELOWANIA EKONOMETRYCZNEGO

1. Specyfikacja równania modelowego:

- a. dobór zmiennych objaśniających (specyfikacja statyczna),
- b. określenie opóźnień bądź przyspieszeń (specyfikacja dynamiczna),
- c. wybór formy związku (liniowa, nieliniowa) i ewentualne nałożenie warunków na parametry równania,
- d. określenie własności składnika losowego (specyfikacja stochastyczna).

Punktem wyjścia przy budowie modelu ekonometrycznego są hipotezy dotyczące danego związku, sformułowane na gruncie teorii ekonomii, wsparte dotychczasowymi doświadczeniami badawczymi. Hipotezy te podlegają weryfikacji ekonometrycznej.

Kroki od a. do c. określają tzw. strukturę deterministyczną równania, zaś krok d. strukturę stochastyczną. Struktura może być stała w czasie lub zmienna.

Najczęstszymi błędami specyfikacji są: pominięcie ważnej zmiennej objaśniającej lub wprowadzenie zmiennej bez znaczenia.

2. Zgromadzenie danych empirycznych

Dane mogą mieć formę:

- szeregów czasowych (np. dane roczne, kwartalne, miesięczne, dzienne, tikowe),
- danych przekrojowych odnoszących się do poszczególnych obiektów rozpatrywanych w tej samej jednostce czasu (np. dotyczące budżetów gospodarstw domowych),
- danych przekrojowo-czasowych.

3. Estymacja parametrów modelu

Wybór metody estymacji: MNK, MNW, 2MNK...

Wybór odpowiedniego oprogramowania: Statistica, PcGive, SPSS, SAS, gretl...

4. Weryfikacja merytoryczna i statystyczna modelu

- a. merytoryczna analiza uzyskanych ocen parametrów,
- b. badanie dopasowania modelu do danych empirycznych (S_e , S_e^2 , R^2 , R , W_e ...),
- c. wyznaczenie precyzji oszacowań parametrów strukturalnych (statystyki t-Studenta, przedziały ufności dla parametrów),
- d. badanie rozkładu odchyłeń losowych (losowość, normalność, brak autokorelacji, homoskedastyczność).

5. Ekonomiczna interpretacja uzyskanych wyników

II. DOBÓR ZMIENNYCH OBJAŚNIAJĄCYCH DO MODELU EKONOMETRYCZNEGO

1. Zasady doboru zmiennych
2. Metoda eliminacji zmiennych quasi-stałych
3. Metoda analizy macierzy współczynników korelacji
4. Metoda Hellwiga doboru zmiennych do modelu

1. ZASADY DOBORU ZMIENNYCH

Zazwyczaj zestaw zmiennych, które mają wystąpić w modelu w charakterze zmiennych objaśniających nie jest określony jednoznacznie. Wynika to przykładowo z:

- mało precyzyjnej teorii ekonomicznej,
- większej liczebności zbioru potencjalnych zmiennych objaśniających niż liczba dostępnych obserwacji.

Metody doboru zmiennych objaśniających do modelu składają się z dwóch elementów:

1. kryterium, na podstawie którego można porównywać różne zestawy zmiennych,
2. sposobu „przeglądania” zbioru potencjalnych zmiennych objaśniających.

Ad 1.

Tu stosuje się mierniki jakości modelu (np. współczynnik determinacji) lub ich zestawy (np. współczynnik determinacji + wymóg statystycznej istotności ocen parametrów).

Ad 2.

- metoda eliminacji a priori (zaczynam od jednej zmiennej i ewentualnie dokładam do modelu kolejne),
- metoda selekcji a posteriori (zaczynam od modelu ze wszystkimi zmiennymi, potem usuwam kolejno najgorsze z nich),
- metoda przeglądu bezpośredniego (badanie własności wszystkich kombinacji zbioru potencjalnych zmiennych objaśniających i wybór najlepszej).

2. METODA ELIMINACJI ZMIENNYCH QUASI-STAŁYCH

Miarą poziomu zmienności zmiennej jest współczynnik zmienności:

$$v_i = \frac{S_i}{\bar{x}_i},$$

gdzie $\bar{x}_i = \frac{1}{m} \sum_{t=1}^m x_{it}$ jest średnią arytmetyczną dla zmiennej X_i , a $S_i = \left(\frac{1}{m} \sum_{t=1}^m (x_{it} - \bar{x}_i)^2 \right)^{\frac{1}{2}}$

odchyleniem standardowym zmiennej X_i .

Po ustaleniu przez badacza wartości krytycznej v^* przyrównuje się ją do wartości współczynnika zmienności dla danej zmiennej. Jeśli $v_i \leq v^*$, to zmienną uznaje się za quasi-stałą i eliminuje ze zbioru potencjalnych zmiennych objaśniających.

3. METODA ANALIZY MACIERZY WSPÓŁCZYNNIKÓW KORELACJI

Niech Y – zmienna objaśniana, X_1, \dots, X_m – potencjalne zmienne objaśniające („kandydatki”), n - ilość obserwacji na zmiennych.

Metoda analizy macierzy współczynników korelacji polega na wyborze takich zmiennych objaśniających do modelu, które są silnie skorelowane ze zmienną objaśnianą i jednocześnie słabo skorelowane pomiędzy sobą.



Przypomnienie:

Do badania zależności dwóch cech mierzalnych służy **współczynnik korelacji liniowej Pearsona**. Jest to stosunek iloczynów odchyłeń wartości zmiennych od średnich arytmetycznych do iloczynu liczebności zbiorowości i odchyłeń standardowych tych zmiennych:

$$r_{xy} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{n \sigma_x \sigma_y} = \frac{\text{cov}_{xy}}{\sigma_x \sigma_y} = \frac{\sum (x_i - \bar{x}) \cdot (y_i - \bar{y})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

gdzie:

r_{xy} – współczynnik korelacji liniowej Pearsona,

n – liczebność próby,

x_i, y_i – wartości zmiennej X i zmiennej Y ,

\bar{x}, \bar{y} – średnie arytmetyczne zmiennej X i zmiennej Y ,

σ_x, σ_y – odchylenia standardowe zmiennej X i zmiennej Y ,

$$\sigma_x = \sqrt{\frac{\sum (x_i - \bar{x})^2}{n}}, \quad \sigma_y = \sqrt{\frac{\sum (y_i - \bar{y})^2}{n}}$$

Współczynnik korelacji liniowej Pearsona przyjmuje wartości z przedziału $\langle -1, 1 \rangle$.

W przypadku braku związku między zmiennymi $r_{xy} = 0$. W przypadku związku funkcyjnego dodatniego bądź ujemnego (korelacja doskonała) współczynnik przyjmuje wartość 1 lub -1 . Im siła zależności między X i Y większa, tym wartość bezwzględna współczynnika jest bliższa jedności.

Punktem wyjścia w metodzie analizy macierzy współczynników korelacji jest znajomość wektora \mathbf{R}_0 i macierzy \mathbf{R} :

$$\mathbf{R}_0 = \begin{bmatrix} r_1 \\ r_2 \\ \vdots \\ r_m \end{bmatrix}, \text{ gdzie } r_i \text{ jest współczynnikiem korelacji pomiędzy zmienną objaśnianą } Y \text{ oraz}$$

potencjalną zmienną objaśniającą X_i .

$$\mathbf{R} = \begin{bmatrix} 1 & r_{12} & \cdots & r_{1m} \\ r_{21} & 1 & \cdots & r_{2m} \\ \cdots & \cdots & \cdots & \cdots \\ r_{m1} & r_{m2} & \cdots & r_{mm} \end{bmatrix}, \text{ gdzie } r_{ij} \text{ jest współczynnikiem korelacji pomiędzy potencjalnymi}$$

zmiennymi objaśniającymi X_i oraz X_j .

Macierz \mathbf{R} jest symetryczna (tzn. $r_{ij} = r_{ji}$), a na jej diagonalnej znajdują się jedynki (ponieważ $r_{ii} = 1$).

Przyjmując poziom istotności α oraz dla $n-2$ stopni swobody, wyznacza się wartość krytyczną współczynnika korelacji:

$$r^* = \left(\frac{(t^*)^2}{(t^*)^2 + n - 2} \right)^{1/2},$$

gdzie t^* jest wartością statystyki odczytanej z tablic dla rozkładu t-Studenta dla α i $n-2$ stopni swobody.

Kolejno wykonywane są kroki następującej procedury:

1. Ze zbioru potencjalnych zmiennych objaśniających eliminuje się te zmienne, dla których zachodzi:

$$|r_i| \leq r^*,$$

są to bowiem zmienne nieistotnie skorelowane ze zmienną objaśnianą.

2. Spośród pozostałych zmiennych jako zmienną objaśniającą w modelu powołuje się taką zmienną X_h , dla której:

$$|r_h| = \max_i \{|r_i|\};$$

zmienną X_h zwie się nośnikiem największego zasobu informacji o zmiennej objaśnianej.

3. Ze zbioru potencjalnych zmiennych objaśniających eliminuje się te zmienne, dla których zachodzi:

$$|r_{hi}| > r^*,$$

są to bowiem zmienne zbyt silnie skorelowane ze zmienną objaśniającą X_h , czyli powielające informacje wnoszone do modelu przez tą zmienną.

Powyższą procedurę powtarzamy, aż do momentu wyczerpania zbioru potencjalnych zmiennych objaśniających.

4. METODA HELLWIGA DOBORU ZMIENNYCH DO MODELU

Idea tej metody polega na wyborze optymalnej kombinacji zmiennych objaśniających, tzn. takiej, której elementy będą w największym stopniu skorelowane z zmienną endogeniczną i równocześnie słabo skorelowane między sobą w porównaniu z innymi możliwymi kombinacjami zmiennych.

Nośnikiem informacji o zmiennej endogenicznej jest potencjalna zmienna objaśniająca. Liczba wszystkich możliwych kombinacji potencjalnych zmiennych objaśniających spośród m zmiennych wynosi: $C = 2^m - 1$.

Pojemnością indywidualną nośnika informacji, będącą miernikiem wielkości informacji o zmiennej Y wnoszonej przez zmienną X_j w l -tej kombinacji, jest

$$h_{lj} = \frac{r_j^2}{\sum_{i \in I_l} |r_{ij}|} \quad (l = 1, 2, \dots, C; j \in I_l)$$

r_j - współczynnik korelacji liniowej między zmienną endogeniczną a j -tą zmienną egzogeniczną,

r_{ij} - współczynnik korelacji liniowej między i -tą i j -tą zmienną egzogeniczną występującą w danej kombinacji,

I_l - zbiór numerów zmiennych tworzących l -tą kombinację.

Pojemnością integralną nośników informacji, łączną pojemnością wszystkich zmiennych objaśniających występujących w l -tej kombinacji, nazywamy wyrażenie:

$$H_l = \sum_{j \in I_l} h_{lj} \quad (l = 1, 2, \dots, C)$$

Pojemność integralna stanowi kryterium doboru kombinacji zmiennych objaśniających do modelu ekonometrycznego. Spośród wszystkich możliwych kombinacji zmiennych wybiera się tą, dla której pojemność integralna jest największa.



III. ESTYMACJA MODELU REGRESJI LINIOWEJ Z JEDNĄ ZMIENNĄ OBJAŚNIAJĄCĄ

1. Pojęcie estymacji i estymatora
2. Specyfikacja zależności
3. Metoda najmniejszych kwadratów
4. Metoda momentów
5. Interpretacja współczynników równania

1. POJĘCIE ESTYMACJI I ESTYMATORA

Celem niniejszego rozdziału jest wyznaczenie oszacowań nieznanymi parametrów β_0 i β_1 w liniowym modelu ekonometrycznym z jedną zmienną objaśniającą opisanym równaniem

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i .$$

Dokonywać tego na drodze tzw. **estymacji statystycznej**, poprzez którą najogólniej rozumiemy wnioskowanie o wartościach parametrów rozkładu zmiennej losowej w populacji generalnej, na podstawie zaobserwowanych wartości tej zmiennej w próbie losowej. W statystyce każdą jednoznacznie określoną funkcję wyników próby, za pomocą której wnioskujemy o wartości nieznanego a priori parametru nazywamy **estymatorem parametru**. Estymator jest zmienną losową, natomiast przybrana przez estymator konkretna wartość liczbową to jego **ocena (szacunek parametru)**.

Teoria estymacji statystycznej dzieli się na dwa działy:

1. estymacja **punktowa** – w jej ramach znajduje się pewną funkcję wyników obserwacji i wartość tej funkcji przyjmuje się za przybliżenie prawdziwej wartości parametru populacji,
2. estymacja **przedziałowa** – polega na konstrukcji (na podstawie wyników z próby) pewnego przedziału liczbowego, o którym z ustalonym z góry prawdopodobieństwem słuszności sądu można zakładać, że zawiera w sobie wartość nieznanego parametru populacji.

W dalszej części rozdziału poświęcimy uwagę pierwszemu sposobowi wnioskowania.

2. SPECYFIKACJA ZALEŻNOŚCI

Uwagę koncentrujemy na zależności pomiędzy zmiennymi Y oraz X postaci:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n, n > 2$$

gdzie

- y_i - i -ta obserwacja zmiennej objaśnianej,
- x_i - i -ta obserwacja zmiennej objaśniającej,
- ε_i - składnik losowy,
- β_0, β_1 - nieznanne parametry strukturalne modelu,
- n - liczebność próby.

Zależność ta, ze względu na obecność składnika losowego w równaniu, nie jest zależnością deterministyczną, lecz **stochastyczną**.

Można wskazać trzy podstawowe przyczyny występowania składnika losowego:

1. nieprzewidywalny element losowości związany z ludzkimi zachowaniami,
2. wpływ wielu pominiętych zmiennych lub nieadekwatnej postaci analitycznej modelu,
3. błąd pomiaru zmiennej Y .

Aby estymatory parametrów modelu $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ miały pewne wymagane własności statystyczne, należy przyjąć następujące założenia na temat składnika losowego ε_i :

Założenie 1. Zmienna objaśniająca X jest nielosowa, jej wartości są ustalonymi liczbami rzeczywistymi.

Założenie 2. Wartości oczekiwane składników losowych są równe zero: $E(\varepsilon_i) = 0$ dla wszystkich i . Czyli występujące zakłócenia, które prezentuje składnik losowy, mają tendencje do wzajemnej redukcji.

Założenie 3. Składnik losowy jest sferyczny, tzn.:

- a. wariancje składników losowych są stałe: $D^2(\varepsilon_i) = \sigma^2$ dla wszystkich i , czyli składnik losowy jest homoskedastyczny,
- b. składniki losowe ε_i i ε_j są od siebie niezależne dla wszystkich $i \neq j$, czyli nie występuje autokorelacja składnika losowego.

Założenie 4. Składniki losowe mają rozkład normalny, co zapisujemy $\varepsilon_i : N(0, \sigma^2)$.

Założenie 5. Liczebność próby jest większa niż liczba szacowanych parametrów, czyli $n > 2$.

Dla rozważanego modelu $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ wartość oczekiwana zmiennej objaśnianej równa jest:

$$E(y_i) = \beta_0 + \beta_1 x_i.$$

To ostatnie równanie nosi nazwę **linii regresji w populacji**.

Kiedy parametry β_0 i β_1 zastąpimy ich oszacowaniami, otrzymamy **linię regresji próby**:

$$\hat{y}_i = b_0 + b_1 x_i$$

gdzie:

\hat{y}_i - wartość teoretyczna zmiennej Y odpowiadająca i -tej obserwacji,

b_0, b_1 - estymatory parametrów β_0, β_1 .

3. METODA NAJMNIEJSZYCH KWADRATÓW

Nieznane parametry β_0, β_1 mogą być oszacowane na podstawie n -elementowej próby statystycznej (x_i, y_i) , $i = 1, 2, \dots, n$. Estymatorami parametrów strukturalnych β_0, β_1 są pewne funkcje obserwacji dokonanych na zmiennych objaśnianej i objaśniającej.

Różnice pomiędzy wartościami empirycznymi a teoretycznymi dla zmiennej Y nazywane są **resztami** i oznaczane symbolem e_i :

$$e_i = y_i - \hat{y}_i$$

Estymacja parametrów modelu za pomocą metody najmniejszych kwadratów (MNK) sprowadza się do znalezienia linii, która będzie najlepsza z punktu widzenia następującego kryterium:

$$\sum_{i=1}^n e_i^2 = \sum_{i=1}^n (y_i - \hat{y}_i)^2 = \sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \rightarrow \min \quad (\text{minimum sumy kwadratów reszt})$$



Powyższa funkcja osiąga minimum w punkcie, gdzie zerują się wartości pierwszych pochodnych względem b_0, b_1 .

$$\frac{\partial \left(\sum_{i=1}^n e_i^2 \right)}{\partial b_0} = \frac{\partial \left(\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \right)}{\partial b_0} = -2 \sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0$$

$$\frac{\partial \left(\sum_{i=1}^n e_i^2 \right)}{\partial b_1} = \frac{\partial \left(\sum_{i=1}^n (y_i - b_0 - b_1 x_i)^2 \right)}{\partial b_1} = -2 \sum_{i=1}^n x_i (y_i - b_0 - b_1 x_i) = 0$$

Po uporządkowaniu otrzymujemy tzw. układ równań normalnych:

$$\sum_{i=1}^n y_i = n b_0 + b_1 \sum_{i=1}^n x_i \quad (*)$$

$$\sum_{i=1}^n x_i y_i = b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2 \quad (**)$$

Ponieważ $\sum_{i=1}^n x_i = n \bar{x}$ oraz $\sum_{i=1}^n y_i = n \bar{y}$ z pierwszego równania (*) otrzymamy: $\bar{y} = b_0 + b_1 \bar{x}$,

$$\text{a w konsekwencji: } b_0 = \bar{y} - b_1 \bar{x} \quad (***)$$

Drugie równanie (**) przekształcamy następująco: $\sum_{i=1}^n x_i y_i = b_0 n \bar{x} + b_1 \sum_{i=1}^n x_i^2$

i podstawiamy do niego wyrażenie (***): $\sum_{i=1}^n x_i y_i = n \bar{x} \bar{y} - b_1 n \bar{x}^2 + b_1 \sum_{i=1}^n x_i^2$

Po dalszych przekształceniach otrzymujemy: $\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y} = b_1 \left(\sum_{i=1}^n x_i^2 - n \bar{x}^2 \right)$

Ostatecznie:

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{\sum_{i=1}^n (x_i - \bar{x})(y_i - \bar{y})}{\sum_{i=1}^n (x_i - \bar{x})^2}$$

$$b_0 = \bar{y} - b_1 \bar{x}$$

Można dowieść, że przy wcześniej poczynionych założeniach estymatory MNK mają własności:

1. liniowości, ponieważ są liniową funkcją zmiennych losowych y_i ,
2. nieobciążoności, ponieważ ich wartość oczekiwana jest równa wartości szacowanych parametrów, co zapisujemy jako $E(b_j) = \beta_j$,
3. efektywności, czyli mają najmniejszą wariancję w klasie nieobciążonych estymatorów liniowych,
4. zgodności, ponieważ są stochastycznie zbieżne do szacowanych parametrów, co zapisujemy jako $\lim_{n \rightarrow \infty} P\{|b_j - \beta_j| < \delta\} = 1$ dla dowolnego $\delta > 0$.

Fakt, iż MNK-estymatory są najlepszymi nieobciążonymi estymatorami w klasie estymatorów liniowych określa się mianem BLUE (ang. *best linear unbiased estimators*).

Przykład 2.

Mamy do dyspozycji dane na temat dostaw serów (Y) i masła (X) w tys. ton w pewnym państwie w latach 2001-2006. Aby wyznaczyć oceny parametrów modelu $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$ przeprowadzamy niezbędne obliczenia:

lata	sery Y	masło X	$x_i y_i$	x_i^2
2001	8	43	344	1849
2002	8	43	344	1849
2003	7	41	287	1681
2004	9	54	486	2916
2005	10	61	610	3721
2006	12	64	768	4096
Suma	54	306	2839	16112

$$\bar{x} = \frac{306}{6} = 51, \quad \bar{y} = \frac{54}{6} = 9$$

$$b_1 = \frac{\sum_{i=1}^n x_i y_i - n \bar{x} \bar{y}}{\sum_{i=1}^n x_i^2 - n \bar{x}^2} = \frac{2839 - 6 \cdot 51 \cdot 9}{16112 - 6 \cdot 51 \cdot 51} = \frac{85}{506} = 0,16798$$

$$b_0 = \bar{y} - b_1 \bar{x} = 9 - 0,16798 \cdot 51 = 0,43302$$

W rezultacie otrzymujemy: $\hat{y}_i = 0,43302 + 0,16798 \cdot x_i$

Dodatkowo wyznaczamy wartości teoretyczne \hat{y}_i oraz reszty modelowe e_i i ich kwadraty:

y_i	\hat{y}_i	$e_i = y_i - \hat{y}_i$	e_i^2
8	7,656126	0,343874	0,118249
8	7,656126	0,343874	0,118249
7	7,320158	-0,32016	0,102502
9	9,503953	-0,50395	0,253966
10	10,67984	-0,67984	0,462182
12	11,18379	0,816206	0,666192
Suma	X	0,000004	1,72134

4. METODA MOMENTÓW

Niech b_0 i b_1 będą oszacowaniami parametrów β_0 i β_1 równania regresji $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i$. Odpowiednikiem składnika losowego ε_i w próbie jest jego oszacowanie $e_i = y_i - \hat{y}_i$, zwane resztą modelową.

Równania, które pozwalają wyznaczyć b_0 i b_1 metodą momentów, otrzymujemy zastępując założenia na temat populacji ich odpowiednikami w próbie.

Założenie $E(\varepsilon_i) = 0$, mówiące o tym, że wartości oczekiwane składników losowych są równe

zeru, ma swój odpowiednik w próbie następujący: $\frac{1}{n} \sum_{i=1}^n e_i = 0$.



Założenie dotyczące nielosowości zmiennej objaśniającej X oznacza, że e_i oraz x_i są niezależne, co można zapisać symbolicznie $\text{cov}_{xe} = 0$; na podstawie próby można to wyrazić

$$\text{jako } \frac{1}{n} \sum_{i=1}^n x_i e_i = 0.$$

W konsekwencji otrzymujemy dwa równania:

$$\sum_{i=1}^n (y_i - b_0 - b_1 x_i) = 0$$

$$\sum_{i=1}^n x_i (y_i - b_0 - b_1 x_i) = 0$$

które po przekształceniach dają układ równań normalnych:

$$\sum_{i=1}^n y_i = n b_0 + b_1 \sum_{i=1}^n x_i$$

$$\sum_{i=1}^n x_i y_i = b_0 \sum_{i=1}^n x_i + b_1 \sum_{i=1}^n x_i^2$$

Jego rozwiązaniem są poszukiwane szacunki b_0 i b_1 . Za pomocą metody momentów otrzymujemy takie same oszacowania parametrów modelu regresji liniowej, jak za pomocą metody najmniejszych kwadratów, gdyż stosujemy te same równania normalne.

5. INTERPRETACJA WSPÓŁCZYNNIKÓW RÓWNANIA

W oszacowanym modelu $\hat{y}_i = b_0 + b_1 x_i$ ocena b_1 informuje, o ile wzrost (gdy $b_1 > 0$) lub zmalał (gdy $b_1 < 0$) średni poziom zmiennej objaśnianej Y pod wpływem zwiększenia się zmiennej niezależnej X o jednostkę.

Wyraz wolny b_0 jest oceną wielkości zmiennej Y , gdy zmienna X jest równa zero. Często pozostawiamy go bez interpretacji. Czasami wyraz wolny uznać można za oszacowanie stałego – niezależnego od zmiennej X – poziomu zmiennej Y , np. za oszacowanie kosztu stałego w modelach kosztów lub minimalnego spożycia w modelach popytu.

Przykład 3.

W modelu $\hat{y}_i = 0,43302 + 0,16798 \cdot x_i$ z przykładu 2. ocena parametru b_1 ma następującą interpretację:

„Wzrost dostaw masła o 1 tysięcy ton (czyli o 1 jednostkę) wywoła średni przyrost dostaw serów o 167,98 ton.”

IV. WERYFIKACJA MODELU

1. Ocena jakości modelu
2. Badanie istotności zmiennych objaśniających modelu
3. Weryfikacja reszt

1. OCENA JAKOŚCI MODELU

Po oszacowaniu parametrów modelu ekonometrycznego należy przeprowadzić analizę dopasowania modelu do danych empirycznych. Miary określające stopień zgodności modelu z danymi empirycznymi obliczane są na podstawie reszt $e_i = y_i - \hat{y}_i$. Model tym lepiej pasuje do danych empirycznych, im reszty co do wartości bezwzględnej są mniejsze.

O zgodności modelu z danymi empirycznymi mówi wariancja składnika losowego σ^2 . Estymatorem wariancji σ^2 składnika losowego w liniowym modelu z k zmiennymi objaśniającymi szacowanym KMNK jest **wariancja reszt**:

$$S_e^2 = \frac{1}{n-k-1} \cdot \sum_{i=1}^n e_i^2$$

Odchylenie standardowe reszt wyznaczamy jako:

$$S_e = \sqrt{\frac{1}{n-k-1} \cdot \sum_{i=1}^n e_i^2}$$

Odchylenie standardowe reszt (inaczej: błąd estymacji, *standard error of estimation*) mówi, o ile przeciętnie zaobserwowane wartości zmiennej objaśnianej różnią się od wartości teoretycznych tej zmiennej wyznaczonych z modelu.

Przykład 4.

Dla danych z przykładu 2:

$$S_e = \sqrt{\frac{1}{n-k-1} \sum_{i=1}^n e_i^2} = \sqrt{\frac{1}{6-1-1} \cdot 1,721} = \sqrt{0,430} = 0,656$$

Wartości empiryczne dostaw serów różnią się od wartości teoretycznych przeciętnie o 656 ton.

Wadą powyższych miar jest fakt, że przeskalowanie zmiennej objaśnianej powoduje zmianę ich wartości.

Dokonyamy teraz **dekompozycji wariancji zmiennej objaśnianej**. Całkowitą zmienność zmiennej objaśnianej można rozbić na dwie części:

$$\sum_{i=1}^n (y_i - \bar{y})^2 = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2 + \sum_{i=1}^n e_i^2$$

Inaczej zapisując:

$$SST = SSR + SSE$$

$SST = \sum_{i=1}^n (y_i - \bar{y})^2$ to suma kwadratów odchylenia wartości empirycznych y -ka od wartości średniej, tzw. **całkowita zmienność zmiennej objaśnianej** (*sum of squares-total*),

$SSR = \sum_{i=1}^n (\hat{y}_i - \bar{y})^2$ to suma kwadratów odchylenia wartości teoretycznych y -ka od wartości średniej, tzw. **zmienność zmiennej objaśnianej wyjaśniona przez model** (*sum of squares-*



regression), a $SSE = \sum_{i=1}^n e_i^2$ - suma kwadratów reszt, czyli **zmienność zmiennej objaśnianej nie wyjaśniona przez model** (*sum of squares-error*).

Dzieląc stronami wzór na całkowitą zmienność y -ka przez SST otrzymujemy: $1 = \frac{SSR}{SST} + \frac{SSE}{SST}$.
Stąd wynikają kolejne miary dopasowania modelu do danych empirycznych:

Współczynnik determinacji R^2 (*R-squared*):

$$R^2 = \frac{SSR}{SST} = \frac{\sum_{i=1}^n (\hat{y}_i - \bar{y})^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Informuje on, jaka część całkowitej zmienności zmiennej objaśnianej została wyjaśniona przez model.

Przykład 5.

Wyznamy wartość współczynnika determinacji dla modelu $\hat{y}_i = 0,43302 + 0,16798 \cdot x_i$ z przykładu 2.

y_i	\hat{y}_i	$(\hat{y}_i - \bar{y})^2$	$(y_i - \bar{y})^2$
8	7,656126	1,805997	1
8	7,656126	1,805997	1
7	7,320158	2,821869	4
9	9,503953	0,253969	0
10	10,67984	2,821862	1
12	11,18379	4,768939	9
Suma	X	14,27863	16

$$R^2 = \frac{14,27863}{16} = 0,892415$$

Otrzymany wynik oznacza, że 89,24% całkowitej zmienności zmiennej objaśnianej zostało wyjaśnione przez model.

Współczynnik zbieżności (indeterminacji) ϕ^2 :

$$\phi^2 = \frac{SSE}{SST} = \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Informuje on, jaka część całkowitej zmienności zmiennej objaśnianej nie została wyjaśniona przez model.

Współczynnik determinacji można również wyznaczyć ze wzoru:

$$R^2 = 1 - \phi^2 = 1 - \frac{\sum_{i=1}^n e_i^2}{\sum_{i=1}^n (y_i - \bar{y})^2}$$

Współczynnik R^2 przyjmuje wartości z przedziału $\langle 0,1 \rangle$. $R^2=1$ dotyczy doskonałego dopasowania modelu do danych empirycznych. Wtedy wszystkie reszty przyjmują wartość 0. Wyznaczona MNK hiperpłaszczyzna przechodzi wówczas przez wszystkie punkty empiryczne. $R^2=0$ odpowiada modelowi $\hat{y}_i = b_0$, w którym żadne zmienne objaśniające nie wyjaśniają zmienności y -ka. Im większa jest wyjaśniona modelem zmienność zmiennej objaśnianej, tym wartość R^2 bliższa jest jedynce. W przypadku modelu z jedną zmienną objaśniającą R^2 równy jest kwadratowi współczynnika korelacji z próby między zmienną objaśnianą a objaśniającą.

Wśród innych miar dopasowania modelu do danych empirycznych występuje **średni błąd bezwzględny** (*mean absolute error*):

$$MAE = \frac{1}{n} \sum_{i=1}^n |e_i|$$

Im mniejsze są wartości MAE , tym dopasowanie modelu lepsze.

Współczynnik zmienności losowej $W_e = \frac{S_e}{\bar{y}} \cdot 100\%$ określa, jaki procent średniej wartości

zmiennej objaśnianej stanowi odchylenie standardowe reszt. Im mniejsza jest wartość W_e , tym dopasowanie modelu lepsze.

Współczynnik korelacji wielorakiej R określa siłę związku liniowego zmiennej objaśnianej ze wszystkimi zmiennymi objaśniającymi modelu.

$$R = \sqrt{R^2}$$

Ponieważ $R = r_{yy}$, stąd współczynnik korelacji wielorakiej informuje w jakim stopniu są ze sobą skorelowane empiryczne i teoretyczne wartości zmiennej objaśnianej.

2. BADANIE ISTOTNOŚCI ZMIENNYCH OBJAŚNIAJĄCYCH MODELU

W rozdziale niniejszym zajmiemy się wnioskowaniem o parametrach liniowego modelu ekonometrycznego:

$$y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$$

Jego oszacowana postać jest następująca:

$$\hat{y}_i = b_0 + b_1 x_i.$$

b_0, b_1 są estymatorami parametrów β_0, β_1 mającymi rozkłady normalne.

Wartości oczekiwane tych estymatorów równe są odpowiednio β_0, β_1 .

Standardowy błąd szacunku parametru S_{b_j} informuje, o ile średnio ocena b_j różni się od parametru β_j , jeśli parametr β_j został oszacowany na podstawie różnych prób składających się z tej samej liczby obserwacji.

Dla modelu z jedną zmienną objaśniającą $y_i = \beta_0 + \beta_1 x_i + \varepsilon_i, \quad i = 1, 2, \dots, n$, celem wyznaczenia standardowych błędów szacunków parametrów β_0, β_1 zastosowanie znajdują wzory:

$$S_{b_0} = S_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} \qquad S_{b_1} = \frac{S_e}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}}$$



Standardowy błąd szacunku powinien być jak najmniejszy w stosunku do oceny parametru. W praktyce przyjmuje się, że nie powinien przekraczać 50% jej wartości bezwzględnej, jeśli liczba stopni swobody jest większa niż 20.

Standardowy błąd szacunku zapisujemy w nawiasie kwadratowym pod oceną parametru strukturalnego modelu.

Przykład 6.

Wyznamy standardowe błędy szacunku parametrów dla modelu z przykładu 2:

$$\hat{y}_t = 0,432 + 0,168x_t$$

$$S_{b_0} = S_e \sqrt{\frac{1}{n} + \frac{\bar{x}^2}{\sum_{i=1}^n (x_i - \bar{x})^2}} = 0,656 \sqrt{\frac{1}{6} + \frac{2601}{506}} = 0,656 \cdot 2,304 = 1,511$$

$$S_{b_1} = \frac{S_e}{\sqrt{\sum_{i=1}^n (x_i - \bar{x})^2}} = \frac{0,656}{\sqrt{506}} = 0,029$$

$$\hat{y}_t = 0,432 + 0,168x_t, \quad t = 1, \dots, 6$$

[1,511] [0,029]

Przedział ufności dla parametru strukturalnego

Estymacja przedziałowa polega na wyznaczeniu, na podstawie wyników obserwacji, takiego przedziału liczbowego, o którym, z ustalonym z góry, bliskim jedności, prawdopodobieństwem słuszności sądu, można twierdzić, że zawiera w sobie prawdziwą wartość szacowanego parametru. Przedział taki nosi nazwę **przedziału ufności**, zaś jego końce, odpowiednio, nazwę **dolnej i górnej granicy ufności**. Prawdopodobieństwo tego, że przedział ufności pokryje nieznaną wartość szacowanego parametru, nazywa się **współczynnikiem ufności** (ozn. $1-\alpha$). Wartość α nazywamy **poziomem istotności**. W celu skonstruowania przedziału ufności poziomu istotności przyjmujemy z góry.

Wiemy, że estymator b_j ma rozkład normalny ze średnią β_j i odchyleniem standardowym σ_{b_j} ,

(co zapisujemy jako $b_j \sim N(\beta_j, \sigma_{b_j})$), zatem $\frac{b_j - \beta_j}{\sigma_{b_j}} \sim N(0,1)$.

W praktyce zamiast nieznanego σ_{b_j} stosuje się S_{b_j} . W związku z tym następująca zmienna ma rozkład t-Studenta o $(n-k-1)$ stopniach swobody:

$$\frac{b_j - \beta_j}{S_{b_j}} \sim t_{n-k-1}$$

Aby wyznaczyć przedział ufności dla parametru β_j , $j = 0, 1, \dots, k$, należy dobrać z tablic rozkładu t-Studenta taką wartość $t_{\alpha, n-k-1}$, aby spełniona była relacja:

$$P\left\{\left|\frac{b_j - \beta_j}{S_{b_j}}\right| \leq t_{\alpha, n-k-1}\right\} = 1 - \alpha$$

Inaczej: $P\{b_j - t_{\alpha, n-k-1} \cdot S_{b_j} \leq \beta_j \leq b_j + t_{\alpha, n-k-1} \cdot S_{b_j}\} = 1 - \alpha$

$(1-\alpha)$ -100-procentowy przedział ufności dla parametru β_j jest więc postaci:

$$(b_j - t_{\alpha, n-k-1} \cdot S_{b_j}, b_j + t_{\alpha, n-k-1} \cdot S_{b_j}), \quad j = 0, 1, \dots, k.$$

W skrócie wartość krytyczną $t_{\alpha, n-k-1}$ będziemy oznaczać t^* .

Długość przedziału ufności zależy od poziomu istotności α , liczby stopni swobody oraz wielkości standardowych błędów szacunku parametrów. Przedział ufności jest tym węższy im wyższa jest wartość poziomu istotności α , większa liczba stopni swobody (a więc bardziej liczna próba) oraz niższa wartość standardowego błędu szacunku parametru. Z praktycznego punktu widzenia będzie nas interesowało, czy skonstruowany przedział ufności zawiera liczbę zero. Będziemy twierdzić, że jeśli do przedziału ufności należy 0, to dany parametr strukturalny modelu jest statystycznie nieistotny.

Przykład 7.

$$\hat{y}_t = 0,432 + 0,168x_t, \quad t = 1, \dots, 6$$

[1,511] [0,029]

Wartość krytyczna statystyki t-Studenta dla poziomu istotności $\alpha=0,05$ i $n-k-1=6-1-1=4$ stopni swobody wynosi $t^*=2,777$.

Przedział ufności dla parametru β_1 :

$$(b_1 - t_{0,05;5} \cdot S_{b_1}; b_1 + t_{0,05;5} \cdot S_{b_1}) = (0,168 - 2,777 \cdot 0,029; 0,168 + 2,777 \cdot 0,029) = (0,087; 0,249)$$

Można sądzić na 95%, że przedział ten obejmuje nieznaną wartość parametru β_1 .

Do przedziału tego nie należy liczba 0, więc można twierdzić, że parametr β_1 jest statystycznie istotny.

Zmienna X wywiera tym samym statystycznie istotny wpływ na zmienną Y .

Test t-Studenta na istotność parametru strukturalnego

Zajmiemy się teraz weryfikowaniem hipotezy dotyczącej braku statystycznej istotności parametru β_j w modelu ekonometrycznym. Hipotezy badawcze formułujemy następująco:

$$H_0 : \beta_j = 0 \text{ (parametr jest statystycznie nieistotny)}$$

$$H_1 : \beta_j \neq 0 \text{ (parametr jest statystycznie istotny)}$$

Jeśli hipoteza zerowa jest prawdziwa, to przy założeniu, że składnik losowy modelu ma

rozkład normalny, statystyka $t_j = \frac{b_j}{S_{b_j}}$, ma rozkład t-Studenta z $n-k-1$ stopniami swobody.

Procedura postępowania:

1. Dla parametru β_j wyznaczamy wartość empiryczną statystyki t_j .
2. Z tablic dla rozkładu t-Studenta dla zadanego poziomu istotności α i $n-k-1$ stopni swobody odczytujemy krytyczną wartość t^* .
3. Jeżeli $|t_j| > t^*$, to hipotezę zerową H_0 odrzucamy na rzecz hipotezy alternatywnej H_1 . Mówi się wtedy o **statystycznej istotności parametru** β_j . Uznajemy, że zmienna X_j wywiera statystycznie istotny wpływ na zmienną Y .
Jeżeli $|t_j| \leq t^*$, to nie ma podstaw do odrzucenia hipotezy H_0 . Mówi się wtedy o **statystycznej nieistotności parametru** β_j . Uznajemy, że zmienna X_j nie wywiera statystycznie istotnego wpływu na zmienną Y .



Jeśli liczba stopni swobody jest większa niż 20, to wtedy dla $\alpha = 0,05$ wartość krytyczna $t^* \approx 2$. W konsekwencji standardowy błąd szacunku parametru nie powinien przekraczać 50% wartości bezwzględnej oceny parametru.

Przykład 8.

$$\hat{y}_t = 0,432 + 0,168x_t, \quad t = 1, \dots, 6$$

[1,511] [0,029]

Zweryfikujemy hipotezę $H_0: \beta_1 = 0$ wobec hipotezy alternatywnej $H_1: \beta_1 \neq 0$.

Wartość krytyczna statystyki t-Studenta dla $\alpha = 0,05$ i 4 stopni swobody wynosi $t^* = 2,777$.

Wartość empiryczna statystyki t-Studenta dla β_1 wynosi: $t_1 = \frac{0,168}{0,029} = 5,76$.

Ponieważ $|t_1| > t^*$, zatem hipotezę zerową odrzucamy na rzecz hipotezy alternatywnej.

Parametr β_1 jest statystycznie istotny.

Zmienna X wywiera statystycznie istotny wpływ na zmienną Y .

Wartość p jest poziomem prawdopodobieństwa, przy którym nie ma podstaw do odrzucenia hipotezy zerowej przy obliczonej na podstawie próby wartości empirycznej statystyki t_j . Przyjmując zazwyczaj poziom $\alpha = 0,05$, H_0 odrzuca się na korzyść H_1 , gdy $p \leq 0,05$.

Przyczyny braku statystycznej istotności parametrów

Brak statystycznej istotności parametru strukturalnego może wynikać z faktycznego braku związku pomiędzy zmienną objaśniającą a zmienną objaśnianą, ale może też być spowodowana innymi przyczynami:

- niską jakością danych statystycznych,
- małą liczebnością próby,
- niewłaściwie dobranym zestawem zmiennych objaśniających,
- niewłaściwą postacią analityczną modelu.

3. WERYFIKACJA RESZT

Sprawdzimy teraz, czy w szacowanym modelu ekonometrycznym są spełnione założenia KMNK dotyczące własności składnika losowego. Weryfikację taką przeprowadzimy na podstawie reszt będących oszacowaniami składników losowych.

Badanie losowości

Test liczby serii

Za pomocą testu liczby serii weryfikujemy hipotezę o trafności doboru postaci analitycznej modelu. Stawiamy hipotezy:

H_0 : przyjęta postać analityczna modelu jest poprawna

H_1 : przyjęta postać analityczna modelu nie jest poprawna

Procedura postępowania:

1. Porządkujemy niemalejąco reszty w próbie według zmiennej porządkującej. Zmienną porządkującą jest zmienna czasowa, gdy dane pochodzą z szeregów czasowych, natomiast dla danych przekrojowych – jedna ze zmiennych objaśniających.
2. Dla uporządkowanego ciągu reszt obliczamy liczbę serii reszt modelu (ozn. S). Seria jest sekwencją reszt o takim samym znaku. Jeśli reszty mają charakter losowy, to liczba serii jest zmienną losową.

3. Z tablic testu liczby serii dla liczby reszt dodatnich n_1 , liczby reszt ujemnych n_2 oraz przyjętego poziomu istotności $\alpha/2$ i $1-\alpha/2$ odczytujemy krytyczne liczby serii S_1^* i S_2^* .
 4. Jeśli $S_1^* < S < S_2^*$, to nie ma podstaw do odrzucenia hipotezy H_0 , zatem reszty mają charakter losowy. Jeśli $S \leq S_1^*$ lub $S \geq S_2^*$ hipotezę H_0 należy odrzucić, rozkład odchyleń uznajemy za nielosowy i twierdzimy, że postać modelu została źle dobrana.
- Popularne tablice testu serii skonstruowano dla $n_1 < 20$ oraz $n_2 < 20$. Ze wzrostem liczb n_1 i n_2 rozkład liczby serii dąży do rozkładu normalnego. Wtedy do oceny losowości ciągu reszt wykorzystujemy się statystykę:

$$Z = \frac{S - E(S)}{\sigma_S}, \quad \text{gdzie } E(S) = \frac{2n_1n_2}{n_1 + n_2} + 1, \quad \sigma_S = \sqrt{\frac{2n_1n_2(2n_1n_2 - n_1 - n_2)}{(n_1 + n_2)^2(n_1 + n_2 - 1)}}.$$

Statystyka Z ma asymptotyczny rozkład normalny $N(0, 1)$.

Badanie normalności

a) Test Shapiro-Wilka

H_0 : składnik losowy modelu ma rozkład normalny

H_1 : składnik losowy modelu nie ma rozkładu normalnego

Procedura postępowania:

1. Porządkujemy reszty według wartości niemalejących, otrzymując ciąg $e_{(1)}, e_{(2)}, \dots, e_{(n)}$.

$$2. \text{ Obliczamy wartość statystyki } W = \frac{[\sum_{i=1}^{\lfloor n/2 \rfloor} a_{n,i} \cdot (e_{(n-i+1)} - e_{(i)})]^2}{\sum_{i=1}^n e_i^2},$$

gdzie $\lfloor n/2 \rfloor$ – część całkowita liczby $n/2$,

$a_{n,i}$ – współczynnik Shapiro-Wilka odczytany z tablic.

3. Z tablic testu Shapiro-Wilka dla poziomu istotności α i wielkości próby n odczytujemy wartość krytyczną W^* .
4. Jeżeli $W \geq W^*$, to nie ma podstaw do odrzucenia hipotezy H_0 .
Jeżeli $W < W^*$, hipotezę H_0 odrzucamy na korzyść hipotezy H_1 .

Test ten może być stosowany dla małych prób (np. gdy $n \leq 30$).

b) Test Jarque'a-Bery

Test ten polega na porównaniu, jak wartości miar asymetrii i kurtozy, obliczonych na podstawie reszt, różnią się od wartości tych miar dla rozkładu normalnego.

H_0 : składnik losowy modelu ma rozkład normalny

H_1 : składnik losowy modelu nie ma rozkładu normalnego

Procedura postępowania:

1. Na podstawie reszt modelowych obliczamy:

współczynnik asymetrii:

$$A = \frac{M_3}{S^3} = \frac{\frac{1}{n} \sum_{i=1}^n e_i^3}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \right)^3}$$

$A > 0$ w przypadku asymetrii prawostronnej.

$A < 0$ w przypadku asymetrii lewostronnej.

$A = 0$ dla rozkładów symetrycznych (m.in. dla rozkładu normalnego).



współczynnik skupienia (kurtozę):

$$K = \frac{M_4}{S^4} = \frac{\frac{1}{n} \sum_{i=1}^n e_i^4}{\left(\sqrt{\frac{1}{n} \sum_{i=1}^n e_i^2} \right)^4}$$

$K < 3$, jeśli rozkład jest bardziej spłaszczony niż rozkład normalny.

$K > 3$, jeśli rozkład jest bardziej wysmukły niż rozkład normalny.

Dla rozkładu normalnego $K=3$.

- Wyznaczamy wartość statystyki $JB = n \left(\frac{1}{6} A^2 + \frac{1}{24} (K-3)^2 \right)$. Ma ona asymptotyczny rozkład χ^2 z 2 stopniami swobody.
- Z tablic dla rozkładu χ^2 dla poziomu istotności α i 2 stopni swobody odczytujemy wartość krytyczną χ_*^2 .
- Jeśli $JB > \chi_*^2$, to hipotezę H_0 o normalności składnika losowego odrzucamy.
Jeśli $JB \leq \chi_*^2$, nie ma podstaw do odrzucenia H_0 .

Test Jarque'a-Bery stosujemy jedynie dla dużych prób.

Odrzucenie hipotezy o normalności składnika losowego oznacza, że estymatory parametrów modelu nie mają rozkładu normalnego. Należy wtedy ostrożnie wnioskować na podstawie testów, wykorzystujących założenie normalności składnika losowego (np. testu t-Studenta).

Badanie autokorelacji

Negatywne zjawisko autokorelacji składników losowych może się pojawić na skutek błędnej specyfikacji modelu (brak ważnej zmiennej objaśniającej, niewłaściwa postać analityczna) lub z powodu powolnego wygasania efektów czynników przypadkowych zaburzających normalny przebieg prawidłowości ekonomicznych.

Mówimy, że **nie występuje autokorelacja składników losowych** (są one nieskorelowane), jeśli $\text{cov}(\varepsilon_i, \varepsilon_j) = 0$ dla $i \neq j, i, j = 1, 2, \dots, n$.

Jeśli $\text{cov}(\varepsilon_i, \varepsilon_j) \neq 0$ dla $j=i-p, i=p+1, \dots, n$, to mówimy, że **wystąpiła autokorelacja rzędu p** .

Jeśli $\text{cov}(\varepsilon_i, \varepsilon_{i-1}) \neq 0$ dla $i=2, \dots, n$, to **wystąpiła autokorelacja I-ego rzędu**.

Miernikiem autokorelacji składników losowych rzędu p są współczynniki autokorelacji z próby, będące współczynnikami korelacji liniowej pomiędzy resztami odległymi od siebie o p okresów.

Do badania autokorelacji I-ego rzędu wykorzystuje się następujący współczynnik $\hat{\rho}$:

$$\hat{\rho} = \frac{\sum_{i=2}^n e_i e_{i-1}}{\sqrt{\sum_{i=1}^n e_i^2 \cdot \sum_{i=2}^n e_{i-1}^2}}$$

Współczynnik $\hat{\rho}$ przyjmuje wartości z przedziału $\langle -1, 1 \rangle$, jego znak określa kierunek zależności, a bezwzględna wartość wskazuje na siłę zależności.

a) Test Durбина-Watsona weryfikuje istnienie autokorelacji pierwszego rzędu.

W teście tym stawiamy hipotezy następująco:

$H_0: \rho = 0$ (nie występuje autokorelacja I-ego rzędu składników losowych)

$H_1: \rho > 0$ (ma miejsce dodatnia autokorelacja I-ego rzędu składników losowych)

albo

$H_0: \rho = 0$ (nie występuje autokorelacja I-ego rzędu składników losowych)

$H_1: \rho < 0$ (ma miejsce ujemna autokorelacja I-ego rzędu składników losowych)
Sprawdzianem jest statystyka Durбина-Watsona postaci:

$$d = \frac{\sum_{i=2}^n (e_i - e_{i-1})^2}{\sum_{i=1}^n e_i^2} \approx 2(1 - \hat{\rho})$$

Statystyka d przyjmuje wartości z przedziału $\langle 0, 4 \rangle$.

Wartość $d = 2$ świadczy o braku autokorelacji (wtedy $\hat{\rho} = 0$).

Gdy $d \approx 0$, to ma miejsce silna autokorelacja dodatnia (wtedy $\hat{\rho} \approx 1$).

Gdy $d \approx 4$, to ma miejsce silna autokorelacja ujemna (wtedy $\hat{\rho} \approx -1$).

Rozkład statystyki d zależy od liczby obserwacji i liczby szacowanych parametrów. Tablice zawierają wartości krytyczne: górną d_U i dolną d_L .

Wnioskowanie na podstawie d :

1. Jeśli $\hat{\rho} > 0$ (czyli $d < 2$), to weryfikuje się hipotezy:

$$H_0: \rho = 0 \quad \text{wobec} \quad H_1: \rho > 0$$

Gdy $d > d_U$, to brak podstaw do odrzucenia hipotezy H_0 .

Gdy $d < d_L$, to odrzucamy hipotezę H_0 i ma miejsce autokorelacja dodatnia.

Gdy $d_L \leq d \leq d_U$ brak rozstrzygnięcia testu.

2. Jeśli $\hat{\rho} < 0$ (czyli $d > 2$), to weryfikuje się hipotezy:

$$H_0: \rho = 0 \quad \text{wobec} \quad H_1: \rho < 0$$

Obliczamy $d' = 4 - d$.

Gdy $d' > d_U$, to brak podstaw do odrzucenia hipotezy H_0 .

Gdy $d' < d_L$, to odrzucamy hipotezę H_0 i ma miejsce autokorelacja ujemna.

Gdy $d_L \leq d' \leq d_U$ brak rozstrzygnięcia testu.

Testu Durбина-Watsona nie stosuje się, jeśli w modelu jako zmienna objaśniająca występuje opóźniona zmienna objaśniana.

Gdy mamy do czynienia z autokorelacją składnika losowego, estymatory wariancji parametrów strukturalnych szacowanych MNK są obciążone. W rezultacie fałszywych informacji dostarczają wartości statystyk t-Studenta oraz przedziały ufności.

W takiej sytuacji parametry modelu należy szacować uogólnioną metodą najmniejszych kwadratów lub metodą różniczki zupełnej.

Badanie homoskedastyczności

W KMNK zakłada się, że wariancje składników losowych są stałe: $D^2(\varepsilon_i) = \sigma^2$ dla $i = 1, 2, \dots, n$ (składnik losowy jest **homoskedastyczny**). W praktyce, zwłaszcza w modelach dla danych przekrojowych, założenie to często nie jest spełnione - składnik losowy jest **heteroskedastyczny** (np. model wydatków gospodarstw w zależności od ich dochodów).

Test Goldfelda-Quandt

Dla dwóch części populacji, co do których przypuszczamy, że cechują je różne wariancje składnika losowego, formułujemy następujące hipotezy:

$$H_0: \sigma_1^2 = \sigma_2^2 \quad (\text{składnik losowy jest homoskedastyczny})$$

$$H_1: \sigma_1^2 \neq \sigma_2^2 \quad (\text{składnik losowy jest heteroskedastyczny})$$

gdzie σ_1^2, σ_2^2 - wariancje składników losowych w pierwszej i drugiej części.

Procedura postępowania:

1. Porządkujemy niemalejąco obserwacje w próbie według zmiennej porządkującej. Zmienną porządkującą może być zmienna czasowa lub – dla danych przekrojowych – zmienna objaśniająca podejrzana o wywoływanie heteroskedastyczności.



2. Wybieramy dwie skrajne podpróby. Pominięta liczba obserwacji nie powinna przekraczać $1/3$ liczebności całej próby. Niech n_1 - liczba obserwacji w pierwszej podpróbie, a n_2 - liczba obserwacji w drugiej. Szacujemy model indywidualnie w każdej podpróbie i wyznaczamy odpowiednie wariancje resztowe S_1^2 i S_2^2 .
3. Obliczamy $F = S_2^2 / S_1^2$ (w liczniku większa z wariancji). Jeśli hipoteza H_0 jest prawdziwa, to F ma rozkład F-Snedecora.
4. W tablicach statystycznych dla poziomu istotności α oraz $m_1 = n_2 - k - 1$, $m_2 = n_1 - k - 1$ stopni swobody odczytujemy wartość krytyczną F^* .
5. Jeśli $F \leq F^*$, to brak podstaw do odrzucenia hipotezy H_0 .
Jeśli $F > F^*$, to hipotezę H_0 odrzucamy na rzecz hipotezy H_1 .

Jeśli heteroskedastyczność została spowodowana łącznie przez kilka zmiennych objaśniających, stosujemy inne testy, np. **test White'a** lub **Harveya-Godfrey'a**.

W wypadku heteroskedastyczności składnika losowego estymatory parametrów strukturalnych uzyskane MNK mają obciążoną wariancję, co uniemożliwia poprawną weryfikację hipotez dotyczących wartości parametrów strukturalnych. W takiej sytuacji do estymacji parametrów modelu należy zastosować ważoną MNK.

V. ANALIZA ROZWOJU ZJAWISK W CZASIE

1. Szeregi czasowe
2. Modele trendu
3. Wygładzanie szeregu czasowego metodą średniej ruchomej

1. SZEREGI CZASOWE

Szeregiem czasowym nazywa się uporządkowany według czasu zbiór obserwacji statystycznych. W wypadku szeregu czasowego przyjmuje się, że wartości zmiennej charakteryzującej określone zjawisko (zmiennej zależnej) zależą od czasu (zmiennej niezależnej).

Rozróżniamy dwa typy szeregów czasowych:

- **szeregi okresów** - zawierają dane dotyczące kształtowania się poziomu zjawiska w całym okresie przyjętym za jednostkę czasu. Istnieje możliwość wydłużania okresów, w jakich dokonuje się pomiaru, np. przechodzenie z danych miesięcznych na kwartalne i roczne, za pomocą agregacji danych.
- **szeregi momentów** - zawierają informacje jedynie o poziomie zjawiska w wyróżnionych momentach czasu (tzn. w chwili dokonywania pomiaru) i nie wiemy, jak się ono kształtowało między kolejnymi momentami obserwacji. Można jednak wyznaczyć przeciętny poziom zjawiska na podstawie pomiarów przeprowadzonych w kilku momentach należących do okresu przyjętego za nową jednostkę czasu (np. mając dane o liczbie zatrudnionych w kolejnych miesiącach można podać średni poziom zatrudnienia w ciągu roku).

Celem analizy szeregów czasowych jest:

- oszacowanie parametrów wybranego modelu kształtowania się zmiennej i ocena dokładności dopasowania modelu do danych empirycznych,
- wykorzystanie oszacowanego modelu dla prognozy kształtowania się zjawiska w przyszłych okresach.

Z punktu widzenia działania różnego rodzaju przyczyn, które wywołują określone zmiany w rozwoju zjawiska, wyróżniamy:

- **przyczyny działające w sposób trwały**, powodując wystąpienie określonej tendencji rozwojowej (trendu),
- **przyczyny zakłócające** tą tendencję w sposób **regularny lub nieregularny**.

Dekompozycja szeregu czasowego polega na stwierdzeniu występowania w szeregu czasowym takich składników, jak:

- stały (przeciętny) poziom,
- tendencja rozwojowa (trend),
- wahania okresowe,
- wahania cykliczne,
- wahania przypadkowe.

Tendencja rozwojowa to systematyczne, jednokierunkowe zmiany (wzrost lub spadek) poziomu badanego zjawiska zachodzące w długim okresie. Wyrównaną linię wyznaczającą zasadniczy kierunek rozwojowy badanego procesu nazywamy **trendem**.

Wahania okresowe (periodyczne) są to rytmiczne wahania o określonym cyklu (okresie przebiegu). Najczęściej obserwuje się wahania o cyklu rocznym, przy czym podokresami cyklu w takim przypadku mogą być półrocza, kwartały, miesiące, a nawet dni.



Wahania sezonowe – wahania o okresie rocznym. U podstaw występowania wahań sezonowych znajdują się takie czynniki egzogeniczne, jak klimatyczno-przyrodnicze lub kalendarzowe (np. czas trwania dnia lub nocy, temperatura, opady).

Wahania cykliczne to długookresowe, rytmiczne wahania wartości zmiennej wokół stałego (przeciętnego) poziomu lub wokół trendu badanej zmiennej.

Wahania koniunkturalne to systemowe, falowe wahania rozwoju gospodarki obserwowane w dłuższych od roku okresach. Analiza tego rodzaju wahań wymaga wieloletnich obserwacji.

Wahania przypadkowe – powodują pewne odchylenia od zmian regularnych. Są wynikiem działania przyczyn ubocznych, ujawniających się nieregularnie.

W celu identyfikacji poszczególnych składowych szeregu czasowego dla konkretnej zmiennej wykorzystuje się **wykresy** zaobserwowanych danych. Analiza wykresu szeregu czasowego pozwala również na wykrycie **obserwacji nietypowych** oraz **punktów zwrotnych**.

Prawidłowości rozwoju zjawiska wykryte w szeregach czasowych można wyrazić modelem opisującym zależność poziomu zjawiska od czasu, czyli w postaci funkcji:

$$y_t = f(t) + g_l(t) + \varepsilon_t \quad \text{lub} \quad y_t = f(t) \cdot g_l(t) \cdot 10^{\varepsilon_t},$$

gdzie:

y_t - poziom badanego zjawiska zaobserwowany w chwili t ,

$f(t)$ - funkcja trendu,

$g_l(t)$ - funkcja wahań okresowych dla $l=1,2,\dots,L$ gdzie L oznacza liczbę podokresów składowej periodycznej,

ε_t - zmienna losowa, charakteryzująca efekty oddziaływania wahań przypadkowych.

Modele powyższe nazywa się modelami wahań w czasie odpowiednio typu **addytywnego** i **multiplikatywnego**.

Dekompozycję szeregu czasowego na poszczególne składowe przeprowadza się stosując odpowiednie **metody**. Wyodrębnianie tendencji rozwojowej można przeprowadzić za pomocą metody:

1. analitycznej, polegającej na oszacowaniu funkcji trendu,
2. mechanicznej, opartej na średnich.

2. MODELE TRENDU

Modele trendu, zwane również **modelami tendencji rozwojowej**, zawierają tylko jedną zmienną objaśniającą, którą jest czas t , czyli:

$$y_t = f(t) + \varepsilon_t,$$

gdzie:

f - symbol dowolnej funkcji,

t - zmienna czasowa przyjmująca najczęściej wartości $t=1,2,\dots,T$,

y_t - zmienna objaśniana,

ε_t - składnik losowy.

Zadanie wyznaczenia funkcji $f(t)$ jest nazywane **wygładzaniem** (wyrównywaniem) szeregu czasowego.

Analityczna metoda wyrównywania szeregów czasowych polega na takim dopasowaniu funkcji $f(t)$ do danych empirycznych, aby jak najlepiej obrazowała ona ogólną tendencję rozwojową, eliminując wahania okresowe i przypadkowe.

W zależności od postaci analitycznej funkcji f wyróżnia się różne rodzaje trendu. Najczęściej w praktyce stosuje się:

- **funkcję liniową** $f(t) = \beta_0 + \beta_1 t$
Postać trendu liniowego stosowana jest w przypadku, gdy można przyjąć założenie o stałych przyrostach wartości zmiennej y w jednostce czasu. Wtedy parametr β_1 wyraża stały przyrost z okresu na okres wartości zmiennej objaśnianej. Funkcję powyższą można oszacować za pomocą MNK.
- **funkcję wielomianową** $f(t) = \beta_0 + \beta_1 t + \dots + \beta_n t^n$
Każdą funkcję ciągłą i ograniczoną można w skończonym przedziale aproksymować wielomianem odpowiedniego stopnia. Korzystając więc z funkcji wielomianowej trendu możemy uzyskać wysoką zgodność modelu z obserwacjami empirycznymi. Nie ma jednak gwarancji, że otrzymana funkcja trendu równie dobrze opisze zmiany zmiennej y w przyszłości. W wypadku wielomianowej funkcji trendu stosując odpowiednie podstawienia sprowadzamy model do postaci liniowej.
- **funkcję potęgową** $f(t) = \beta_0 t^{\beta_1}$
Tę postać funkcji trendu stosuje się najczęściej w przypadku, gdy wykładnik β_1 jest większy od zera, lecz mniejszy od jedności, wtedy funkcja powyższa charakteryzuje się malejącym tempem wzrostu. Jest to przykład funkcji nieliniowej względem parametrów. Aby sprowadzić ją do postaci liniowej, logarytmujemy obie strony wyrażenia, a następnie szacujemy parametry MNK.
- **funkcję wykładniczą** $f(t) = e^{\beta_0 + \beta_1 t}$
Własnością charakterystyczną tej funkcji trendu są stałe w czasie przyrosty względne. Przed estymacją funkcję sprowadzamy ją do postaci liniowej za pomocą logarytmowania obu stron.
- **funkcję logarytmiczną** $f(t) = \beta_0 + \beta_1 \ln t$
Trend logarytmiczny wybieramy, gdy wzrost badanej zmiennej jest coraz wolniejszy. Po wprowadzeniu podstawienia $t' = \ln t$ otrzymujemy równanie liniowe, którego parametry szacuje się MNK. Jest to więc przykład trendu nieliniowego, ale liniowego względem parametrów.
- **funkcję logistyczną** $f(t) = \frac{\beta_1}{1 + e^{\beta_2 - \beta_3 t}}$

3. WYGŁADZANIE SZEREGU CZASOWEGO METODĄ ŚREDNIEJ RUCHOMEJ

Idea wyrównywania szeregu czasowego za pomocą średnich ruchomych polega na zastąpieniu wyjściowych wartości zmiennej średnimi arytmetycznymi, obliczonymi sekwencyjnie dla wybranej liczby obserwacji.

Średnią ruchomą prostą k -okresową wyznaczamy z następującego wzoru:

$$\bar{y}_{t-\frac{k-1}{2}} = \frac{1}{k} \sum_{i=t-k+1}^t y_i \quad \text{dla} \quad t \geq k.$$

Wyznaczone wartości średnie przyporządkowuje się środkowym obserwacjom, na których podstawie obliczono średnie. Średnie te bywają nazywane średnimi łańcuchowymi ze względu na wzajemne zazębianie się okresów biorących udział w liczeniu.



Przykład 9.

Przeanalizujemy ilość zachorowań na gruźlicę w Polsce w latach 1999-2008, wyznaczając dla tego szeregu średnie ruchome 3-letnią i 5-letnią. Celem określenia kolejnych wartości średnich wykorzystamy wzory:

- 3-okresowa średnia ruchoma dla T -elementowego szeregu czasowego:

$$\bar{y}_2 = \frac{y_1 + y_2 + y_3}{3}, \bar{y}_3 = \frac{y_2 + y_3 + y_4}{3}, \dots, \bar{y}_{T-1} = \frac{y_{T-2} + y_{T-1} + y_T}{3}$$

- 5-okresowa średnia ruchoma dla T -elementowego szeregu czasowego:

$$\bar{y}_3 = \frac{y_1 + y_2 + y_3 + y_4 + y_5}{5}, \dots, \bar{y}_{T-2} = \frac{y_{T-4} + y_{T-3} + y_{T-2} + y_{T-1} + y_T}{5}$$

lata	liczba osób	średnia ruchoma 3-letnia	średnia ruchoma 5-letnia
1999	12 179		
2000	11 477	11442,7	
2001	10 672	10874,7	10985,4
2002	10 475	10423,7	10448,2
2003	10 124	10030,7	10006,6
2004	9 493	9628,7	9589,6
2005	9 269	9116,3	9217,4
2006	8 587	8823,3	8808,8
2007	8 614	8427,3	
2008	8 081		

W sytuacji, kiedy średnie ruchome mają być wykorzystane do eliminacji wahań okresowych, muszą obejmować ściśle określoną liczbę jednostek czasu, odpowiadającą zaobserwowanemu cyklowi wahań. Na przykład pojawienie się wahań przy danych kwartalnych oznacza potrzebę zastosowania 4-okresowej średniej ruchomej, a wahań dla danych miesięcznych - średniej 12-okresowej.

Pojawia się wtedy problem liczenia średnich ruchomych z parzystej liczby okresów. Są to tzw. **średnie scentrowane**, które wyznacza się zgodnie z relacją:

$$\bar{y}_{t-\frac{k}{2}} = \frac{1}{k} \left(\frac{1}{2} \cdot y_{t-k} + \sum_{i=t-k+1}^{t-1} y_i + \frac{1}{2} y_t \right) \quad \text{dla } t \geq k+1$$

gdzie:

- \bar{y}_i - wartości średnich ruchomych obliczone dla kolejnych podokresów,
- k - krok uśredniania.

Przykład 10.

Dla danych kwartalnych dotyczących dni nieprzepracowanych z powodu choroby pracujących na rachunek własny w latach 1988-1990 wyznaczmy średnie scentrowane 4-okresowe. Korzystamy ze wzorów:

$$\bar{y}_3 = \frac{0,5 \cdot y_1 + y_2 + y_3 + y_4 + 0,5 \cdot y_5}{4}, \dots, \bar{y}_{T-2} = \frac{0,5 \cdot y_{T-4} + y_{T-3} + y_{T-2} + y_{T-1} + 0,5 \cdot y_T}{4}$$

lata	kwartały	dni w tys.	średnie ruchome scentrowane
1988	I	20	
	II	18	
	III	16	19,4
	IV	23	19,6
1989	I	21	19,9
	II	19	20,5
	III	17	20,6
	IV	25	20,8
1990	I	22	21,0
	II	19	21,4
	III	19	
	IV	26	

Wady mechanicznej metody wyrównywania szeregów czasowych:

- skracanie szeregów czasowych,
- brak możliwości przedstawienia trendu w formie matematycznej funkcji trendu.



VI. PROGNOZOWANIE EKONOMETRYCZNE

1. Prognoza
2. Błędy prognozy
3. Prognozowanie na podstawie modeli tendencji rozwojowej

1. PROGNOZA

Finalnym etapem analizy ekonometrycznej jest wykorzystanie oszacowanego i zweryfikowanego modelu do prognozy zmiennej objaśnianej.

Prognozowanie ekonometryczne oznacza wnioskowanie o przyszłości za pomocą modelu, który został poddany wszechstronnej weryfikacji w poprzednich etapach. **Prognoza** to konkretny wynik tego wnioskowania.

Z modelu prognozuje się przyszłe wartości zmiennej objaśnianej (zmiennej prognozowanej) na podstawie danych przyszłych wartości zmiennych objaśniających.

Termin „prognozowanie” jest najczęściej używany z myślą o przewidywaniu przyszłości, jednak zasady predykcji ekonometrycznej mogą być stosowane i dla danych przekrojowych.

Przy prognozowaniu na podstawie modelu ekonometrycznego powinny być spełnione następujące **założenia**:

1. Ma miejsce stabilność relacji ujętych w modelu, tzn. postać analityczna, parametry i zbiór zmiennych objaśniających nie zmieniają się.
2. Ma miejsce stabilność rozkładu odchyłeń losowych modelu.
3. Znane są wartości zmiennych objaśniających, na podstawie których buduje się prognozę.
4. Dopuszczalna jest ekstrapolacja modelu poza próbę statystyczną.

Prognozę konkretnej zmiennej stanowi liczba (**prognoza punktowa**) lub przedział liczbowy, który z określonym prawdopodobieństwem zawiera wartość zmiennej prognozowanej (**prognoza przedziałowa**).

Niech oszacowany model ma postać $\hat{y}_i = b_0 + b_1 x_i$, $i=1, \dots, n$.

Oznaczmy przez y_m^P wartość **prognozy punktowej** zmiennej Y na moment m . Przyjmując za wartość zmiennej objaśniającej x_m^* , prognozę definiujemy następująco:

$$y_m^P = b_0 + b_1 x_m^*.$$

Sposoby określenia przyszłych wartości zmiennej objaśniającej mogą być następujące:

- analiza wartości przeszłych i przyjęcie dla okresu prognozowanego np. średniej wartości lub różnych wariantów wartości zmiennej (symulacja ekonometryczna),
- zbudowanie odrębnego modelu objaśniającego przebieg interesującej zmiennej,
- wykorzystanie metod ekstrapolacji szeregów czasowych.

Przykład 11.

W przykładzie 2 na podstawie 6 obserwacji oszacowano następujący model:

$$\hat{y}_i = 0,43302 + 0,16798 \cdot x_i,$$

gdzie y_i – dostawy serów w tys. ton, x_i – dostawy masła w tys. ton.

Jeśli chcemy prognozować, jakie byłyby dostawy serów w roku 2007 w przypadku dostaw masła w tym roku na poziomie 70 tys. ton, wyznaczmy prognozę punktową:

$$y_{2007}^P = 0,43302 + 0,16798 \cdot x_{2007}^* = 0,43302 + 0,16798 \cdot 70 = 12,19162$$

2. BŁĘDY PROGNOZY

Przy ocenie jakości wyznaczonej prognozy punktowej wykorzystuje się dwa typy błędów:

- oczekiwany błąd prognozy (**błąd ex ante**), nazywany błędem predykcji lub predyktora,
- zrealizowany błąd prognozy (**błąd ex post**), nazywany błędem prognozy.

Niech y_m^* oznacza rzeczywistą wartość zmiennej prognozowanej, natomiast y_m^P prognozę punktową dla tej zmiennej. Wtedy błąd prognozy jest równy $y_m^* - y_m^P$.

Średnim błędem predykcji ex ante jest estymator odchylenia standardowego błędu prognozy. Wyznaczany jest w momencie budowy prognozy i informuje, o ile oszacowana wartość zmiennej prognozowanej y_m^P odchyła się średnio od rzeczywistej wartości zmiennej prognozowanej y_m^* .

Gdy w modelu ekonometrycznym uwzględniona jest tylko jedna zmienna objaśniająca, to średni błąd predykcji ex ante wyraża się jako:

$$v_m = S_e \sqrt{1 + \frac{1}{n} + \frac{(x_m^* - \bar{x})^2}{\sum_{i=1}^n (x_i - \bar{x})^2}}$$

Do porównywania dokładności prognoz wyznaczonych dla różnych okresów czasu lub dla różnych zmiennych przydatny jest błąd względny predykcji postaci: $wv_m = \frac{v_m}{y_m^P}$.

Średni błąd predykcji ex ante wykorzystywany jest do wyznaczenia **prognozy przedziałowej** dla pojedynczej wartości y_m^* .

$(1-\alpha) \cdot 100$ -procentowy **przedział ufności dla prognozy** (inaczej **przedział predykcji**), jest postaci:

$$(y_m^P - t^* \cdot v_m; y_m^P + t^* \cdot v_m),$$

gdzie y_m^P – prognoza punktowa,

t^* – wartość krytyczna z tablic dla rozkładu t-Studenta dla $n-k-1$ stopni swobody i poziomu istotności α ,

v_m – średni błąd predykcji ex ante.

Z prawdopodobieństwem $1-\alpha$ można twierdzić, że przedział powyższy pokryje rzeczywistą wartość zmiennej prognozowanej y_m^* .

Błąd prognozy ex post wyznacza się dopiero po dokonaniu pomiaru rzeczywistej wartości zmiennej prognozowanej. Jest on różnicą między zaobserwowaną wartością zmiennej prognozowanej y_m^* i wartością prognozy y_m^P :

$$e_m^* = y_m^* - y_m^P$$

Błąd względny prognozy ex post określony jest jako: $we_m^* = \frac{y_m^* - y_m^P}{y_m^*}$

Do najczęściej stosowanych w praktyce innych miar dokładności prognoz ex post należą:

- średni błąd (*mean error*):

$$ME = \frac{1}{T} \sum_{m=1}^T (y_m^* - y_m^P)$$



- średni błąd bezwzględny (*mean absolute error*):

$$MAE = \frac{1}{T} \sum_{m=1}^T |y_m^* - y_m^P|$$

- średni absolutny błąd procentowy (*mean absolute percentage error*):

$$MAPE = \frac{1}{T} \sum_{m=1}^T \left| \frac{y_m^* - y_m^P}{y_m^*} \right| \cdot 100$$

- błąd średniokwadratowy (*mean squared error*):

$$MSE = \frac{1}{T} \sum_{m=1}^T (y_m^* - y_m^P)^2$$

- pierwiastek błędu średniokwadratowego (*root mean squared error*):

$$RMSE = \sqrt{\frac{1}{T} \sum_{m=1}^T (y_m^* - y_m^P)^2}$$

gdzie:

y_m^* - wartość rzeczywista zmiennej prognozowanej,

y_m^P - wyznaczona prognoza,

T - horyzont prognozy.

3. PROGNOZOWANIE NA PODSTAWIE MODELI TENDENCJI ROZWOJOWEJ

Prognozowanie na podstawie modelu tendencji rozwojowej $\hat{y}_t = f(t) + g_t(t)$ polega na prostej **ekstrapolacji funkcji trendu** (i ewentualnie funkcji $g_t(t)$) przez podstawienie do modelu w miejsce zmiennej czasowej numeru momentu m , na który wyznacza się prognozę:

$$y_m^P = f(m) + g_t(m) \quad \text{dla } m > n.$$

Przykładowo założmy, że mamy oszacowany model trendu liniowego postaci: $\hat{y}_t = b_0 + b_1 t$.

Wówczas prognozę punktową na moment m -ty wyznacza się jako: $y_m^P = b_0 + b_1 m$.

Przykład 12.

Zmieniającą się w ciągu 12 miesięcy cenę pewnego dobra opisuje model trendu:

$$\hat{y}_t = 30,08 + 0,55t.$$

Podstawiając $t = 13$ wyznaczmy prognozę ceny na kolejny miesiąc:

$$y_{13}^P = 30,08 + 0,55 \cdot 13 = 37,23.$$

Wyodrębnianie wahań sezonowych przeprowadza się za pomocą tzw. **wskaźnika sezonowości**, który można wyznaczyć

$$\text{analitycznie: } S_l = \frac{\sum_{i=0}^{T-1} y_{l+iL}}{\sum_{i=0}^{T-1} \hat{y}_{l+iL}} \quad \text{lub mechanicznie: } S_l = \frac{\sum_{i=0}^{T-1} y_{l+iL}}{\sum_{i=0}^{T-1} \bar{y}_{l+iL}}$$

gdzie:

\hat{y}_t - wartości teoretyczne wyznaczone na podstawie oszacowanej funkcji trendu,

\bar{y}_t - wartości średnich ruchomych obliczone dla kolejnych podokresów,

S_l - surowe wskaźniki sezonowości,

$t = 1, 2, \dots, T, l = 1, 2, \dots, L$ (dla danych kwartalnych $L=4$, miesięcznych $L=12$).

Suma surowych wskaźników sezonowości powinna być równa L , czyli: $\sum_{l=1}^L S_l = L$.

Jeśli ta relacja nie zachodzi, to należy wyznaczyć **współczynniki skorygowane** postaci:

$$S_l^v = v \cdot S_l$$

gdzie: v - współczynnik korygujący, który wyznacza się jako: $v = \frac{L}{\sum_{l=1}^L S_l}$.

Absolutną wielkość odchyłeń sezonowych wyznacza się z relacji:

$$g_l(t) = S_l^v \cdot \bar{y} - \bar{y},$$

gdzie: \bar{y} - średni poziom zjawiska y_t , wyznaczony dla wszystkich obserwacji $t = 1, 2, \dots, T$.



VII. PROGRAMOWANIE LINIOWE

1. Model procesu decyzyjnego
2. Programowanie liniowe
3. Dualność w programowaniu liniowym

1. MODEL PROCESU DECYZYJNEGO

Zespół metod matematycznych służących do rozwiązywania problemów z zakresu podejmowania decyzji obejmuje dziedzinę o nazwie **badania operacyjne**.

Zastosowania badań operacyjnych dotyczą przykładowo:

- planowania produkcji i gospodarki zapasami,
- problemów optymalnych przydziałów (tu: zagadnienie transportowe), optymalnego rozmieszczenia pracowników czy przydziału maszyn do realizacji różnych zadań,
- planowania remontów i wymiany urządzeń,
- optymalizacji mieszanek, a w tym zagadnienia diety,
- optymalizacji procesów technologicznych.

Matematyczny model decyzyjny to opis sytuacji decyzyjnej w języku matematyki. W opisie tym występują: zmienne decyzyjne, parametry, funkcja celu, warunki ograniczające oraz warunki brzegowe.

Klasyfikacja modeli decyzyjnych:

1. z uwagi na zmienne decyzyjne:

- a. liniowe,
- b. nieliniowe,

2. ze względu na parametry:

- a. deterministyczne – gdy wszystkie parametry są stałe i znane,
- b. stochastyczne – gdy choć jeden parametr jest niestały,

3. z uwagi na czas:

- a. statyczne – gdy dotyczą jednego okresu,
- b. dynamiczne – gdy dotyczą wielu okresów.

W trakcie rozwiązywania modelu, czyli wyznaczania optymalnej decyzji decydenta, w zależności od rodzaju modelu stosuje się różne metody: programowanie liniowe, programowanie nieliniowe, programowanie całkowitoliczbowe, programowanie dynamiczne, teorię gier, programowanie zapasów, programowanie wielokryterialne i inne. Rozwiązanie polega na wyznaczeniu rozwiązań dopuszczalnych (rozwiązań spełniających wszystkie warunki zadania) oraz rozwiązania optymalnego (najlepszego spośród rozwiązań dopuszczalnych).

2. PROGRAMOWANIE LINIOWE

Model programowania liniowego (PL) w postaci standardowej można zapisać następująco:

$$Z(x_1, x_2, \dots, x_n) = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \max$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n \leq b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n \leq b_2$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n \leq b_m$$

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$$

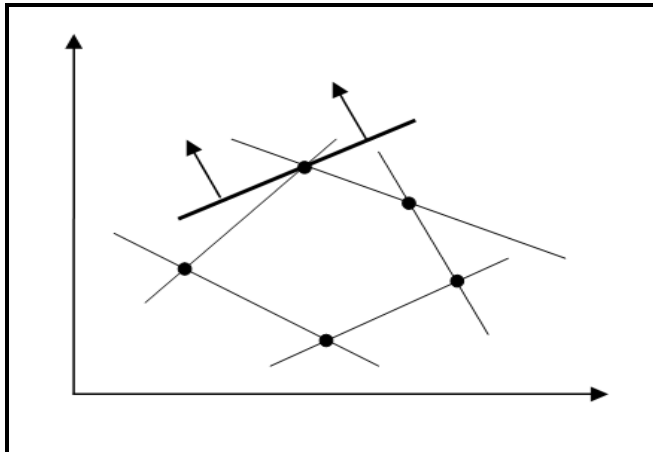
lub



Niepusty, wypukły zbiór rozwiązań dopuszczalnych może być zbiorem ograniczonym lub nieograniczonym.

Jeśli funkcja celu $Z(x)$ zagadnienia programowania liniowego jest ciągła, to na mocy twierdzenia Weierstrassa osiąga ona na zbiorze ograniczonym ekstremum. Ekstremum to znajduje się przynajmniej w jednym punkcie wierzchołkowym.

Przykładowo:



Można sformułować następujące twierdzenia:

Twierdzenie. Jeśli istnieje optymalne rozwiązanie zadania programowania liniowego, to istnieje optymalne rozwiązanie wierzchołkowe (bazowe).

Twierdzenie. Jeśli istnieją dwa optymalne rozwiązania zadania programowania liniowego, to dowolna kombinacja liniowa tych rozwiązań jest również rozwiązaniem optymalnym tego zadania.

W sytuacji, w której zbiór rozwiązań dopuszczalnych zagadnienia programowania liniowego jest nieograniczony, funkcja celu osiąga maksimum (minimum) w co najmniej jednym punkcie wierzchołkowym, jeśli jest ograniczona z góry (z dołu). Jeśli funkcja celu jest nieograniczona, ekstremum znajduje się w nieskończoności.

Postać kanoniczna zadania programowania liniowego:

$$Z(x_1, x_2, \dots, x_n) = c_1x_1 + c_2x_2 + \dots + c_nx_n \rightarrow \max(\min)$$

$$a_{11}x_1 + a_{12}x_2 + \dots + a_{1n}x_n = b_1$$

$$a_{21}x_1 + a_{22}x_2 + \dots + a_{2n}x_n = b_2$$

$$\vdots \qquad \qquad \qquad \vdots$$

$$a_{m1}x_1 + a_{m2}x_2 + \dots + a_{mn}x_n = b_m$$

$$x_1 \geq 0, x_2 \geq 0, \dots, x_n \geq 0$$

Inaczej:

$$Z(\mathbf{x}) = \mathbf{c}^T \mathbf{x} \rightarrow \max$$

$$\mathbf{A} \mathbf{x} = \mathbf{b}$$

$$\mathbf{x} \geq \mathbf{0}$$

gdzie:

$$\mathbf{A} = \begin{bmatrix} a_{11} & a_{12} & \dots & a_{1n} \\ a_{21} & a_{22} & \dots & a_{2n} \\ \dots & & & \\ a_{m1} & a_{m2} & \dots & a_{mn} \end{bmatrix}; \quad \mathbf{b} = \begin{bmatrix} b_1 \\ b_2 \\ \vdots \\ b_m \end{bmatrix}; \quad \mathbf{c} = \begin{bmatrix} c_1 \\ c_2 \\ \vdots \\ c_n \end{bmatrix}; \quad \mathbf{x} = \begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_n \end{bmatrix}$$

$$n > m$$

$$\text{rz}(\mathbf{A}) = m$$

Rozwiązanie bazowe to rozwiązanie, które powstaje przez przyjęcie, że $n-m$ zmiennych jest równe 0 i rozwiązanie układu równań $\mathbf{Ax}=\mathbf{b}$. Maksymalna liczba tych rozwiązań to:

$$\binom{n}{m} = \frac{n!}{(n-m)!m!}$$

Dopuszczalne rozwiązania bazowe nazywamy **rozwiązaniami podstawowymi**.

Rozwiązanie podstawowe jest niezdegenerowane, jeśli ma dokładnie m dodatnich wartości x_j . Rozwiązanie podstawowe jest zdegenerowane, jeśli ma mniej niż m dodatnich wartości x_j .

Metoda simpleks rozwiązywania zadań programowania liniowego to ukierunkowany przegląd zbioru rozwiązań bazowych. W przeglądzie tym pomija się rozwiązania niedopuszczalne oraz gorsze od już znalezionych.

Przykład 13.

Przedsiębiorstwo wytwarza dwa wyroby (A i B). Spośród surowców zużywanych w procesie produkcji trzy są limitowane. Limity dziennego zużycia wynoszą odpowiednio: dla surowca S_1 120 kg, dla surowca S_2 420 kg, dla surowca S_3 110 kg.

Tablica zawiera jednostkowe zużycia tych surowców:

Maszyna	Wyrób	
	A	B
S_1	1	1
S_2	2	4
S_3	1	0

Zyski jednostkowe ze sprzedaży obu wyrobów wynoszą odpowiednio 3 i 4 złote. Wielkości produkcji x_1 i x_2 maksymalizujące zysk przedsiębiorstwa są rozwiązaniami następującego problemu decyzyjnego:

$$Z(\mathbf{x}) = 3x_1 + 4x_2 \rightarrow \max$$

$$x_1 + x_2 \leq 120$$

$$2x_1 + 4x_2 \leq 420$$

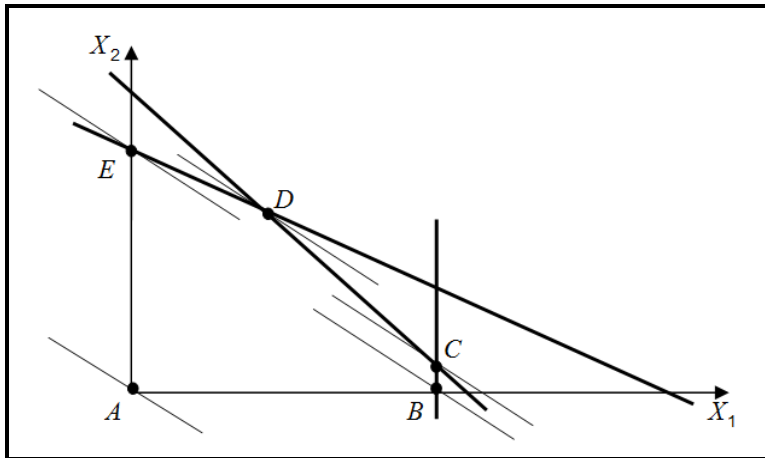
$$x_1 \leq 110$$

$$x_1, x_2 \geq 0$$

W problemie tym $3x_1 + 4x_2$ jest funkcją celu, $x_1 + x_2 \leq 120$, $2x_1 + 4x_2 \leq 420$, $x_1 \leq 110$ – ograniczeniami nałożonymi na dzienne zużycia surowców, natomiast $x_1, x_2 \geq 0$ – warunkami nieujemności rozwiązania.



Rozwiązanie graficzne powyższego liniowego problemu decyzyjnego przedstawia się następująco:



Punkty $A-E$ są wierzchołkami zbioru, do którego należą kombinacje produkcyjne $x^T = [x_1, x_2]$ spełniające układ ograniczeń oraz warunki nieujemności rozwiązania. Każdą taką kombinację nazywamy rozwiązaniem dopuszczalnym. Współrzędne punktów $A-E$ znajdujemy rozwiązując 5 układów par równań powstałych z warunków ograniczających i warunków nieujemności:

Wierzchołek	x_1	x_2	Z
A	0	0	0
B	110	0	330
C	110	10	370
D	30	90	450
E	0	105	420

Przez punkty te, dla $Z(\mathbf{x}) = \bar{Z}$ równego odpowiednio 0, 330, 370, 420 oraz 450, można poprowadzić tzw. „linie jednakowego zysku” o postaci $x_2 = \frac{\bar{Z}}{4} - \frac{3}{4}x_1$. Kombinacja $x_0^T = [30, 90]$ jest najlepsza, ponieważ dla niej osiągnany jest najwyższy poziom zysku $Z = 450$. Nazwiemy ją w związku z tym rozwiązaniem optymalnym.

Powyższy problem decyzyjny ma również swoją postać kanoniczną:

$$\begin{aligned} Z(x_1, x_2, x_3, x_4, x_5) &= 3x_1 + 4x_2 + 0x_3 + 0x_4 + 0x_5 \rightarrow \max \\ x_1 + x_2 + x_3 &= 120 \\ 2x_1 + 4x_2 + x_4 &= 420 \\ x_1 + x_5 &= 110 \\ x_1, x_2, x_3, x_4, x_5 &\geq 0 \end{aligned}$$

x_1, x_2 – zmienne decyzyjne, x_3, x_4, x_5 – zmienne swobodne (uzupełniające).
Postać kanoniczna modelu stanowi punkt wyjścia w metodzie simpleks.



VIII. ANALIZA PRZEPŁYWÓW MIĘDZYGAŁĘZIOWYCH

1. Tabele przepływów międzygałęziowych
2. Model Leontiefa

1. TABELE PRZEPŁYWÓW MIĘDZYGAŁĘZIOWYCH

Analiza przepływów międzygałęziowych (PM; inne nazwy: analiza nakładów i wyników, analiza input-output) jest rodzajem rachunku makroekonomicznego, związanego z badaniem złożonych struktur ekonomicznych. Punktem wyjścia jest bilans gospodarczy w postaci umożliwiającej ilościowe ujęcie zależności między elementami tych struktur. Twórcą metody jest Wasily Leontief (lata 30-te XX. wieku).

Tabela przepływów międzygałęziowych (TPM) dla systemu zamkniętego:

i	X_i	x_{ij}				Y_i
		1	2	...	n	
1	X_1	x_{11}	x_{12}	...	x_{1n}	Y_1
2	X_2	x_{21}	x_{22}	...	x_{2n}	Y_2
...
n	X_n	x_{n1}	x_{n2}	...	x_{nn}	Y_n
	x_{0j}	x_{01}	x_{02}	...	x_{0n}	
	Z_j	Z_1	Z_2	...	Z_n	
	X_j	X_1	X_2	...	X_n	

Zakładamy, że mamy do czynienia z n gałęziami gospodarki; $i, j = 1, 2, 3, \dots, n$. Symbole posiadają następujące oznaczenia:

- X_i – wartość produktu globalnego i -tej gałęzi gospodarki,
- x_{ij} – przepływ z gałęzi i do gałęzi j ,
- Y_i – wartość produktu finalnego (końcowego) gałęzi i -tej,
- x_{0i} – płace gałęzi i -tej,
- Z_i – zysk gałęzi i -tej,
- D_i – wartość dodana gałęzi i -tej.

Przyjmujemy, że i -ty wiersz odpowiada działalności i -tej gałęzi jako producenta. Suma $x_{i1} + x_{i2} + \dots + x_{in}$ oznacza zużycie produkcyjne wyrobów i -tej gałęzi. Produkt końcowy Y_i odpowiada wartości popytu końcowego. Zużycie produkcyjne odpowiada wartości popytu pośredniego. Bilans działalności i -tej gałęzi zapisujemy w postaci **równania podziału**:

$$X_i = x_{i1} + x_{i2} + \dots + x_{in} + Y_i$$

Każda gałąź w procesie produkcyjnym zużywa pewne produkty. Wartości ich zużycia umieszczone są w kolejnych kolumnach.

Suma przepływów do j -tej gałęzi stanowi tzw. **koszty materiałowe** KM_j :

$$KM_j = x_{1j} + x_{2j} + \dots + x_{nj}$$

Łączny **koszt produkcji** K_j stanowią koszty materiałowe KM_j oraz płace x_{0j} :

$$K_j = x_{0j} + KM_j.$$

Zysk Z_j gałęzi jest różnicą między wartością wielkości produkcji globalnej, a poniesionymi kosztami: $Z_j = X_j - K_j$.

Wartością dodaną D_j w gałęzi jest różnica między wartością produkcji globalnej, a kosztami materiałowymi: $D_j = X_j - (x_{1j} + x_{2j} + \dots + x_{nj}) = x_{0j} + Z_j$.

Suma składników w j -tej kolumnie daje wartość produktu globalnego j -tej gałęzi. Bilans kosztów i zysków gałęzi ujmuje **równanie kosztów**: $X_j = x_{0j} + x_{1j} + x_{2j} + \dots + x_{nj} + Z_j$.

Łączną produkcję globalną całego systemu można wyrazić jako:

$$\sum_{i=1}^n X_i = \sum_{i=1}^n \sum_{j=1}^n x_{ij} + \sum_{i=1}^n Y_i \quad \text{lub} \quad \sum_{j=1}^n X_j = \sum_{j=1}^n \sum_{i=1}^n x_{ij} + \sum_{j=1}^n (x_{0j} + Z_j).$$

Skąd wynika równanie wiążące produkt finalny z płacami i zyskami: $\sum_{i=1}^n Y_i = \sum_{j=1}^n (x_{0j} + Z_j)$.

Gdy tabela przepływów międzygałęziowych opisuje całą gospodarkę, ostatnie równanie jest tzw. **warunkiem równowagi ogólnej**. Wtedy strony równania opisują produkt krajowy (lub dochód narodowy). Lewa strona to opis podażowy (rzeczowy). Prawa strona to opis popytowy (pieniężny).

Strona podażowa dotyczy wyprodukowanych towarów, które są tu wyrażone w ujęciu wartościowym. Zmiana cen produktów ma wpływ na wartość lewej strony równania. Strona popytowa dotyczy środków pieniężnych, np. wynagrodzeń, podatków. Zmiana poziomu płac lub podatków wpływa na prawą stronę równania. Zmiana jednej ze stron równania narusza równowagę i w rezultacie powoduje zmiany po drugiej stronie równania (np. zmiany inflacyjne).

Przykład 14.

Zakładamy trójgałęziowy system gospodarczy.

i	X_i	x_{ij}			Y_i
		1	2	3	
1	400	50	195	0	155
2	300	100	0	90	110
3	150	80	45	15	10
	x_{0j}	100	30	15	
	Z_j	70	30	30	
	X_j	400	300	150	

Bilans podziału produkcji gałęzi 1: $400 = 50 + 195 + 0 + 155$

Bilans kosztów w gałęzi 1: $400 = 50 + 100 + 80 + 100 + 70$

Warunek równowagi ogólnej: $155 + 110 + 10 = 100 + 70 + 30 + 30 + 15 + 30$

Tabela przepływów międzygałęziowych (TPM) dla systemu otwartego:

W tabeli uwzględnimy import, export i amortyzację oraz podział produktu finalnego zgodnie z jego przeznaczeniem: spożycie, inwestycje, przyrost środków obrotowych i rezerw, eksport.

i	X _i	x _{ij}				Y _i			
		1	2	...	n	spożycie (1)	inwestycje (2)	śr.obrot rez. (2)	eksport (4)
1	X ₁	x ₁₁	x ₁₂	...	x _{1n}	Y ₁ ⁽¹⁾	Y ₁ ⁽²⁾	Y ₁ ⁽³⁾	Y ₁ ⁽⁴⁾
2	X ₂	x ₂₁	x ₂₂	...	x _{2n}	Y ₂ ⁽¹⁾	Y ₂ ⁽²⁾	Y ₂ ⁽³⁾	Y ₂ ⁽⁴⁾
...
n	X _n	x _{n1}	x _{n2}	...	x _{nn}	Y _n ⁽¹⁾	Y _n ⁽²⁾	Y _n ⁽³⁾	Y _n ⁽⁴⁾
n+1	import	x _{n+1,1}	x _{n+1,2}	...	x _{n+1,n}	x _{n+1} ⁽¹⁾	x _{n+1} ⁽²⁾	x _{n+1} ⁽³⁾	x _{n+1} ⁽⁴⁾
n+2	amortyzacja	x _{n+2,1}	x _{n+2,2}	...	x _{n+2,n}				
		x ₀₁	x ₀₂	...	x _{0n}				
		Z ₁	Z ₂	...	Z _n				
		X ₁	X ₂	...	X _n				

Zysk j-tej gałęzi formułujemy teraz jako:

$$Z_j = X_j - K_j = X_j - (x_{0j} + x_{1j} + x_{2j} + \dots + x_{nj} + x_{n+1,j} + x_{n+2,j}),$$

natomiast **wartość dodana (produkcja czysta)**, czyli różnica między wartością **produkcji globalnej** a **kosztami materialnymi**, to:

$$D_j = X_j - (x_{1j} + x_{2j} + \dots + x_{nj} + x_{n+1,j} + x_{n+2,j}) = x_{0j} + Z_j$$

Równanie kosztów dla j-tej gałęzi ma postać:

$$X_j = x_{0j} + x_{1j} + x_{2j} + \dots + x_{nj} + x_{n+1,j} + x_{n+2,j} + Z_j$$

Równanie podziału produkcji dla i-tej gałęzi ma postać:

$$X_i = x_{i1} + x_{i2} + \dots + x_{in} + Y_i^{(1)} + Y_i^{(2)} + Y_i^{(3)} + Y_i^{(4)}$$

Warunek równowagi systemu:

$$\sum_{i=1}^n \sum_{k=1}^4 Y_i^{(k)} - \sum_{j=1}^n x_{n+1,j} = \sum_{j=1}^n x_{n+2,j} + \sum_{j=1}^n (x_{0j} + Z_j)$$

Gdy analiza przepływów międzygałęziowych dotyczy całej gospodarki to ostatnie równanie ukazuje wartość dochodu narodowego wytworzonego brutto: lewa strona równania wskazuje kierunki jego zużycia, zaś prawa źródła powstania. Zgodnie z równaniem dochód narodowy wytworzony brutto jest sumą wartości dodanej (produkcji czystej) w gałęziach produkcji materialnej i amortyzacji.

Na podstawie tabeli przepływów międzygałęziowych możliwe jest policzenie pewnych współczynników charakteryzujących proces produkcji:

Współczynnik materiałochłonności produkcji globalnej w j-tej gałęzi:
$$m_j = \frac{KM_j}{X_j} = \frac{\sum_{i=1}^{n+1} x_{ij}}{X_j}$$



Po uporządkowaniu otrzymujemy model Leontiefa:

$$\begin{aligned}(1-a_{11})X_1 - a_{12}X_2 - \dots - a_{1n}X_n &= Y_1 \\ -a_{21}X_1 + (1-a_{22})X_2 - \dots - a_{2n}X_n &= Y_2 \\ \dots & \\ -a_{n1}X_1 - a_{n2}X_2 - \dots + (1-a_{nn})X_n &= Y_n\end{aligned}$$

Model ten ma następującą postać macierzową: $(\mathbf{I} - \mathbf{A})\mathbf{X} = \mathbf{Y}$,

gdzie \mathbf{I} oznacza macierz jednostkową wymiaru $n \times n$, \mathbf{X} wektor produkcji globalnej, a \mathbf{Y} wektor produkcji finalnej. Macierz $(\mathbf{I} - \mathbf{A})$ nazywamy macierzą Leontiefa.

Model Leontiefa służy do określania wielkości produkcji finalnej, dla ustalonej produkcji globalnej lub określania produkcji globalnej koniecznej do osiągnięcia wymaganej produkcji finalnej.

Przykład 16.

Dane jak w przykładzie 15. Macierz Leontiefa:

$$\mathbf{I} - \mathbf{A} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} - \begin{bmatrix} 0,125 & 0,65 & 0 \\ 0,25 & 0 & 0,6 \\ 0,2 & 0,15 & 0,1 \end{bmatrix} = \begin{bmatrix} 0,875 & -0,65 & 0 \\ -0,25 & 1 & -0,6 \\ -0,2 & -0,15 & 0,9 \end{bmatrix}$$

Wyznaczenie produkcji finalnej dla danej produkcji globalnej $\mathbf{X}^T = [400 \ 300 \ 150]$:

$$\mathbf{Y} = (\mathbf{I} - \mathbf{A})\mathbf{X} = \begin{bmatrix} 0,875 & -0,65 & 0 \\ -0,25 & 1 & -0,6 \\ -0,2 & -0,15 & 0,9 \end{bmatrix} \begin{bmatrix} 400 \\ 300 \\ 150 \end{bmatrix} = \begin{bmatrix} 155 \\ 110 \\ 10 \end{bmatrix}$$

O ile wzrośnie produkt finalny jeśli, produkt globalny \mathbf{X} zmieni się o $\Delta\mathbf{X}$?

$$\Delta\mathbf{X}^T = [40 \ 30 \ -5]$$

$$\Delta\mathbf{Y} = (\mathbf{I} - \mathbf{A})\Delta\mathbf{X} = \begin{bmatrix} 0,875 & -0,65 & 0 \\ -0,25 & 1 & -0,6 \\ -0,2 & -0,15 & 0,9 \end{bmatrix} \begin{bmatrix} 40 \\ 30 \\ -5 \end{bmatrix} = \begin{bmatrix} 15,5 \\ 23 \\ -17 \end{bmatrix}$$

Przy założeniu odwracalności macierzy Leontiefa otrzymujemy: $(\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y} = \mathbf{X}$,

czyli postać modelu Leontiefa pozwalającą określić produkcję globalną na podstawie produkcji finalnej.

Przykład 17.

Ciąg dalszy przykładu 16. Policzmy macierz odwrotną do macierzy Leontiefa:

$$(\mathbf{I} - \mathbf{A})^{-1} = \begin{bmatrix} 1,672 & 1,207 & 0,805 \\ 0,712 & 1,625 & 1,084 \\ 0,49 & 0,539 & 1,471 \end{bmatrix}$$

Wyznaczenie produkcji globalnej \mathbf{X} , przy której produkcja finalna $\mathbf{Y}^T = [155 \ 110 \ 10]$:

$$\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1} \mathbf{Y} = \begin{bmatrix} 1,672 & 1,207 & 0,805 \\ 0,712 & 1,625 & 1,084 \\ 0,49 & 0,539 & 1,471 \end{bmatrix} \cdot \begin{bmatrix} 155 \\ 110 \\ 10 \end{bmatrix} = \begin{bmatrix} 400 \\ 300 \\ 150 \end{bmatrix}$$

Prognozowanie na podstawie modelu Leontiefa

Model Leontiefa służy do krótkookresowego prognozowania przyszłej wartości produkcji finalnej lub globalnej pod warunkiem, że założono niezmienną technologię produkcji (elementy macierzy \mathbf{A} są stałe w czasie).

Prognozy I rodzaju dokonujemy, gdy w oparciu o model $\mathbf{Y} = (\mathbf{I} - \mathbf{A})\mathbf{X}$ wyznaczamy wektor produkcji finalnej dla zadanego przyszłego wektora produkcji globalnej.

Prognozy II rodzaju dokonujemy, gdy ustalony jest pożądany przyszły wektor produkcji finalnej i na podstawie modelu $\mathbf{X} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}$ wyznaczmy wektor produkcji globalnej, który umożliwi osiągnięcie produktu końcowego na oczekiwanym poziomie.

Prognoza mieszana polega na prognozowaniu wybranych elementów wektora produkcji globalnej i finalnej, jeśli ustalone są pozostałe elementy obu wektorów.

Przykład 18.

Na podstawie danych ekonomicznych oceniamy, że w kolejnym okresie ($t+1$) produkcja globalna w gałęziach 1 i 2 wzrośnie o 10%, zaś w gałęzi 3 nie zmieni się. Jaka będzie produkcja finalna w okresie $t+1$?

Produkt globalny w okresie $t+1$: $\mathbf{X}_{t+1}^T = [440 \ 330 \ 150]$

Produkt finalny w okresie $t+1$ (prognoza I rodzaju):

$$\mathbf{Y}_{t+1} = (\mathbf{I} - \mathbf{A})\mathbf{X}_{t+1} = \begin{bmatrix} 0,875 & -0,65 & 0 \\ -0,25 & 1 & -0,6 \\ -0,2 & -0,15 & 0,9 \end{bmatrix} \begin{bmatrix} 440 \\ 330 \\ 150 \end{bmatrix} = \begin{bmatrix} 170,5 \\ 130 \\ -2,5 \end{bmatrix}$$

Do zrealizowania zakładanej produkcji globalnej potrzebny jest import produktów gałęzi 3 o wartości 2,5 jednostki.

Jaki powinien być wektor produkcji globalnej, aby uniknąć konieczności importu towarów z gałęzi 3 (produkcja finalna ma wynieść: $\mathbf{Y}_{t+1}^T = [170,5 \ 130 \ 0]$)?

Produkt globalny w okresie $t+1$ (prognoza II rodzaju) obliczamy następująco:

$$\mathbf{X}_{t+1} = (\mathbf{I} - \mathbf{A})^{-1}\mathbf{Y}_{t+1} = \begin{bmatrix} 1,672 & 1,207 & 0,805 \\ 0,712 & 1,625 & 1,084 \\ 0,49 & 0,539 & 1,471 \end{bmatrix} \cdot \begin{bmatrix} 170,5 \\ 130 \\ 0 \end{bmatrix} = \begin{bmatrix} 442,012 \\ 332,709 \\ 153,677 \end{bmatrix}$$