

## **PROGNOZOWANIE**

**Prognozowanie** to finalny etap analizy ekonometrycznej polegający na wykorzystaniu oszacowanego i zweryfikowanego modelu do prognozy zmiennej endogenicznej.

**Prognozowanie** przeważnie dotyczy szeregów czasowych.

## **ZADANIE PROGNOZOWANIA**

- ✓ Wnioskowanie o przyszłości na podstawie modelu uprzednio zweryfikowanego,

**Prognoza** jest wynikiem wnioskowania o przyszłości.

## MODEL LINIOWY

Najprostsza postać modelu ekonometrycznego to jednorównaniowy model liniowy zwany **modelem regresji prostej**:

$$Y = b_0 + b_1 X + u ,$$

gdzie:

Y to zmienna **zależna** lub inaczej zmienna **objaśniana**

X jest zmienną **niezależną** lub inaczej zmienną **objaśniającą** w równaniu modelu.

$b_0$  i  $b_1$  to **parametry strukturalne modelu**.

$u$  jest **składnikiem losowym**.

**Składnik losowy** to zmienna losowa o określonych własnościach, która reprezentuje tzw. **błąd w równaniu**.

## POSTAĆ OGÓLNA MODELU

Postać ogólną jednorównaniowego modelu ekonometrycznego formułujemy dla większej liczby zmiennych objaśniających:  $X_1, X_2, \dots, X_k$ .

Jest to model tzw. **regresji wielorakiej**.

Model jednorównaniowy z  $k$  zmiennymi objaśniającymi zapisujemy jako:

$$Y = b_0 + b_1 X_1 + b_2 X_2 + \dots + b_k X_k + u$$

Model posiada  **$k + 1$  parametrów strukturalnych**.

# LINIOWOŚĆ I NIELINIOWOŚĆ

W ekonometrii rozróżniamy następujące, podstawowe typy modeli:

- ściśle liniowe — oznacza to że jest on liniowy względem zmiennych objaśniających jak i parametrów strukturalnych, model,

$$Y = b_0 + b_1X_1 + b_2X_2 + \dots + b_kX_k + u$$

- liniowe względem parametrów , model,

$$Y = a + bX + cX^2 + u_1$$

- liniowy względem zmiennych , model,

$$Y = f + gX + g^2Z + u_2$$

- ściśle nieliniowy — nie ma możliwości przekształcenia na model liniowy

$$Y = h(X^i + Z^j)u_3.$$

## ZASADA MNK

**Znaleźć takie wartości parametrów strukturalnych modelu (oceny), które minimalizują następujące wyrażenie:**

$$S = \sum_{t=1}^n \hat{u}_t^2 = \sum_{t=1}^n \left( Y_t - \hat{Y}_t \right)^2$$

**gdzie „daszki” oznaczają oszacowania.**

**Funkcja  $S$  minimalizuje sumę kwadratów wyrażen, które nazywamy składnikami resztowymi (resztami) modelu.**

**Zauważmy, że  $S$  jest funkcja parametrów modelu, bowiem:**

$$\hat{Y}_t = \hat{b}_0 + \hat{b}_1 X_{1t} + \dots + \hat{b}_k X_{kt}$$

## MNK - graficznie

W przypadku  $k=1$  rozwiązaniem zadania minimalizacji są dwie liczby:

- wyraz wolny,
- parametr nachylenia prostej regresji.

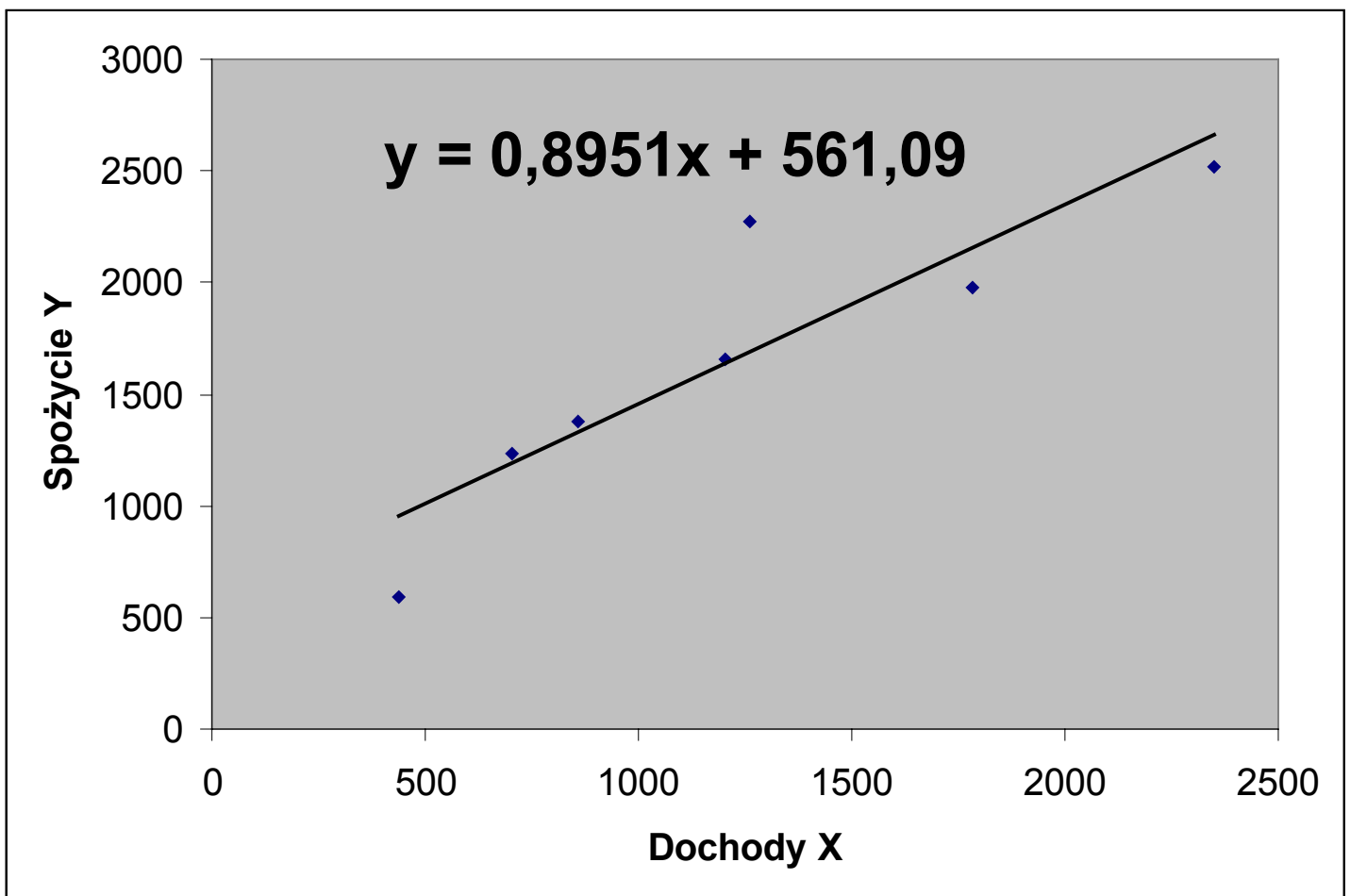
*Przykład - wydatki na spożycie (Y) w wybranych 7 rodzinach w Warszawie w lipcu 1996 roku (w zł)*

$Y_t$	$X_t$
1784	1978
439	590
858	1381
2351	2518
1262	2272
702	1231
1205	1656

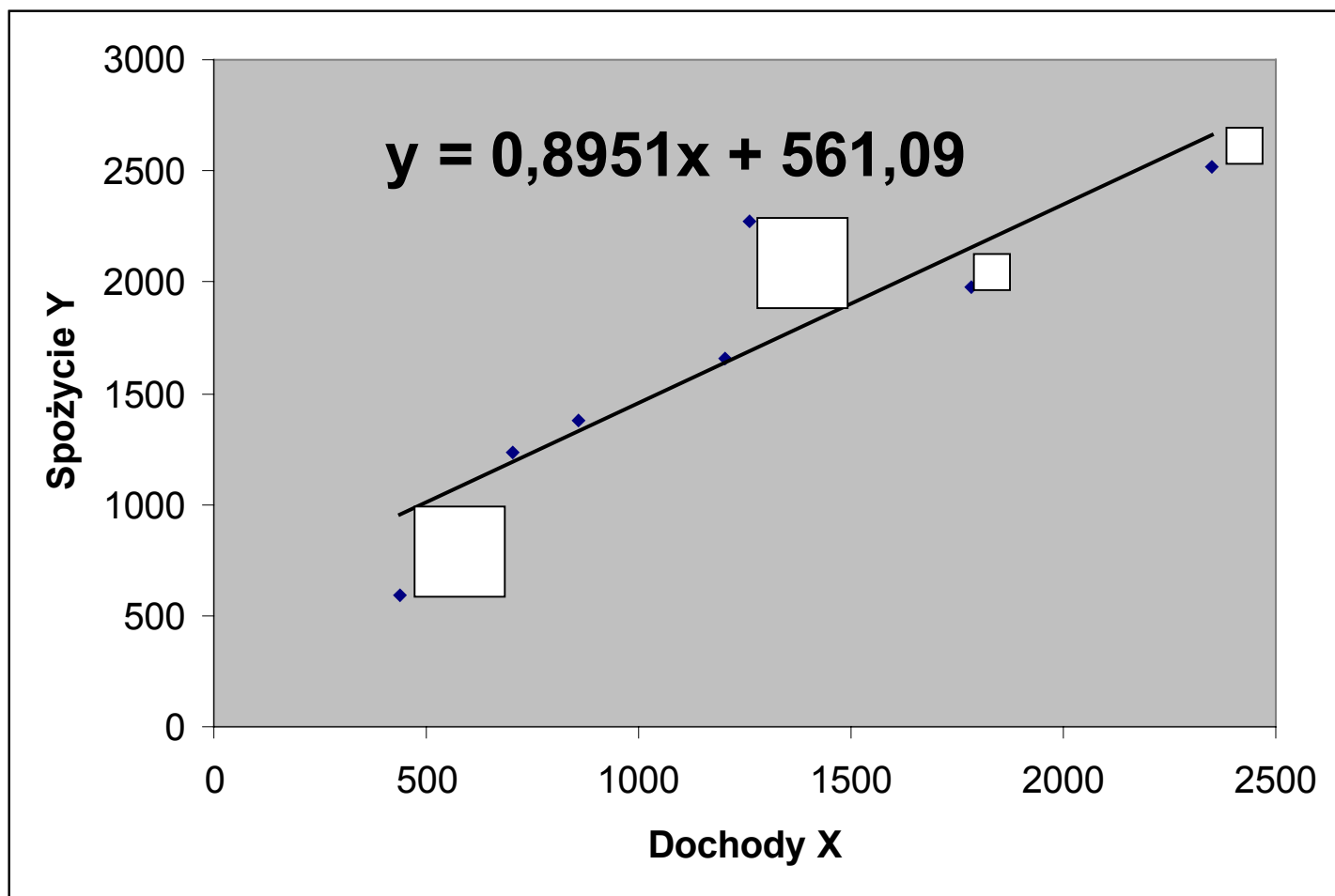
Oszacowanie MNK przedstawia prosta:

$$\hat{Y}_t = -561,01 + 0,89 X_t$$

**Wnioskujemy, że krańcowa skłonność do konsumpcji (KSK) w tej grupie rodzin wynosi 0,89.**







## KLASYCZNA METODA NAJMNIEJSZYCH KWADRATÓW (KMNK)

Metoda najmniejszych kwadratów została opracowana przez C.F. Gaussa na początku XIX wieku. Podstawą do uzyskiwania estymatorów metodą MNK jest twierdzenie Gaussa-Markowa, które głosi, że jeżeli wektory błędów  $\varepsilon_i$  mają taką samą wariancję i są nieskorelowane to estymator otrzymany metodą MNK jest najlepszy tzn. ma najmniejszą wariancję. W klasycznym modelu regresji liniowej najlepszymi nieobciążonymi estymatorami liniowymi współczynników funkcji regresji są estymatory uzyskane metodą MNK. Oznacza to, że estymatory współczynników pozwalają oceniać parametry klasycznego modelu regresji liniowej bez błędu systematycznego i z najmniejszym przeciętnym błędem losowym. Estymatory MNK są nieobciążone, zgodne i efektywne.

- Estymator parametru jest nieobciążony jeżeli wartość oczekiwana tego estymatora równa się szacowanemu parametrowi. Przy jednej próbie ocena różni się od szacowanego parametru, ale stosując estymator nieobciążony wiemy, że odchylenia wartości estymatora od wartości parametru nie mają tendencyjnego charakteru, są wśród nich zarówno odchylenia „na plus” jak i „na minus” o różnych wartościach (ujemnych i dodatnich). Przy wyliczaniu średniej wartości niwelują się wszystkie różnice, a więc nieobciążony estymator gwarantuje, że oszacowany na jego podstawie parametr nie jest obciążony systematycznym błędem.
- Estymator jest zgodny jeżeli jest stochastycznie zbieżny do szacowanego parametru, im większa liczebność próby to prawdopodobieństwo zgodności wartości estymatora z szacowanym parametrem zmierza do 1, a więc przy dużej próbie prawdopodobieństwo popełnienia dużego błędu przy szacowaniu parametru populacji jest niewielkie.
- Najefektywniejszy z pośród wszystkich nieobciążonych estymatorów danego parametru jest taki estymator, który ma najmniejszą wariancję.

Wartości zmiennej objaśnianej nazwiemy wartościami teoretycznymi zmiennej objaśnianej i oznaczmy je przez  $\hat{y}_t$  :

$$\hat{y}_t = a_0 + a_1x_{1t} + a_2x_{2t} + \dots + a_kx_{kt}$$

gdzie  $t = 1, 2, \dots, n$

Różnicę między wartością empiryczną a teoretyczną zmiennej objaśnianej dla okresu  $t$  można nazwać resztą -  $e_t$

$$e_t = y_t - \hat{y}_t$$

Wektorowy zapis wartości teoretycznej zmiennej objaśnianej oraz reszty wyglądają następująco:

$$\hat{y} = \begin{bmatrix} \hat{y}_1 \\ \vdots \\ \hat{y}_n \end{bmatrix}, \quad e = \begin{bmatrix} e_1 \\ \vdots \\ e_n \end{bmatrix}$$

W ten oto sposób otrzymujemy macierzowy zapis równań

$$\hat{y} = Xa$$

oraz równań (3.13):

$$e = y - \hat{y} = y - Xa$$

Idea metody najmniejszych kwadratów jest wyznaczenie takiego wektora oceny  $a$  wektora parametrów  $\alpha$ , dla którego funkcja  $S(a) = e^T e$ , przyjmuje wartości najmniejsze.

Funkcja  $S(a)$  przedstawia sumę kwadratów odchyleń teoretycznych wartości zmiennej objaśnianej od empirycznych wartości tej zmiennej i można ją zapisać w następującej postaci:

$$S(a) = e^T e = (y - Xa)^T (y - Xa) = y^T y - 2a^T X^T + a^T X^T Xa$$

Poszukiwanie punktu stacjonarnego funkcji  $S$  z warunku koniecznego istnienia ekstremum funkcji,

$$\partial S(a) / \partial a = 0$$

prowadzi do równania macierzowego

$$-2X^T y + 2X^T Xa = 0$$

a w końcu pojawia się układ równań normalnych względem  $a$  o postaci

$$X^T X a = X^T y$$

Macierz  $X^T X$  jest macierzą kwadratową, symetryczną stopnia  $k+1$ , która jest warunkiem otrzymania jedyne rozwiązanie układu (3.20) które przyjmuje postać:

$$a = (X^T X)^{-1} X^T y$$

Poszukiwanie punktu stacjonarnego dla przedstawionego wyżej problemu pokażemy wykonując proste przekształcenia algebraiczne. Dla uproszczenia zapisu pominiemy wskaźnik  $t$  przy zmiennej  $x$  oraz teoretyczne wartości zmiennej objaśnianej oznaczmy przez  $y$ . Uprościmy też model liniowy do postaci:  $y_i = ax_i + b$ . Poszukiwać będziemy wartości współczynników  $a$  i  $b$  danych następującą zależnością funkcyjną:

$$f(a, b) = \sum_{i=1}^n (y_i - ax_i - b)^2,$$

co możemy zapisać:

$$f(a, b) = \sum_{i=1}^n (y_i^2 + a^2 x_i^2 + b^2 + 2abx_i - 2ax_i y_i - 2by_i).$$

Budujemy układ równań:

$$\begin{cases} \frac{\partial f(a, b)}{\partial a} = 0 \\ \frac{\partial f(a, b)}{\partial b} = 0 \end{cases},$$

Przekształcając dalej, otrzymujemy:

$$\begin{cases} \frac{\partial f}{\partial a} = \sum_{i=1}^n 2ax_i^2 + \sum_{i=1}^n 2bx_i - \sum_{i=1}^n 2x_i y_i = 0 \\ \frac{\partial f}{\partial b} = \sum_{i=1}^n 2b + \sum_{i=1}^n 2ax_i - \sum_{i=1}^n 2y_i = 0 \end{cases},$$

$$\begin{cases} a \sum_{i=1}^n x_i^2 + b \sum_{i=1}^n x_i - \sum_{i=1}^n x_i y_i = 0 \\ \sum_{i=1}^n b + a \sum_{i=1}^n x_i - \sum_{i=1}^n y_i = 0 \end{cases}.$$

Obliczając b z drugiego z powyższych równań otrzymujemy:

$$b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n}.$$

Wstawiając do równania pierwszego w miejsce b mamy:

$$a \sum_{i=1}^n x_i^2 + \frac{\sum_{i=1}^n x_i \sum_{i=1}^n y_i - a \left( \sum_{i=1}^n x_i \right)^2}{n} - \sum_{i=1}^n x_i y_i = 0$$

$$an \sum_{i=1}^n x_i^2 + \sum_{i=1}^n x_i \sum_{i=1}^n y_i - a \left( \sum_{i=1}^n x_i \right)^2 - n \sum_{i=1}^n x_i y_i = 0$$

$$a \left[ n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2 \right] = n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i$$

$$a = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}$$

Teraz możemy obliczyć współczynnik  $b$ .

$$b = \frac{\sum_{i=1}^n y_i - a \sum_{i=1}^n x_i}{n} = \bar{y} - a\bar{x}.$$

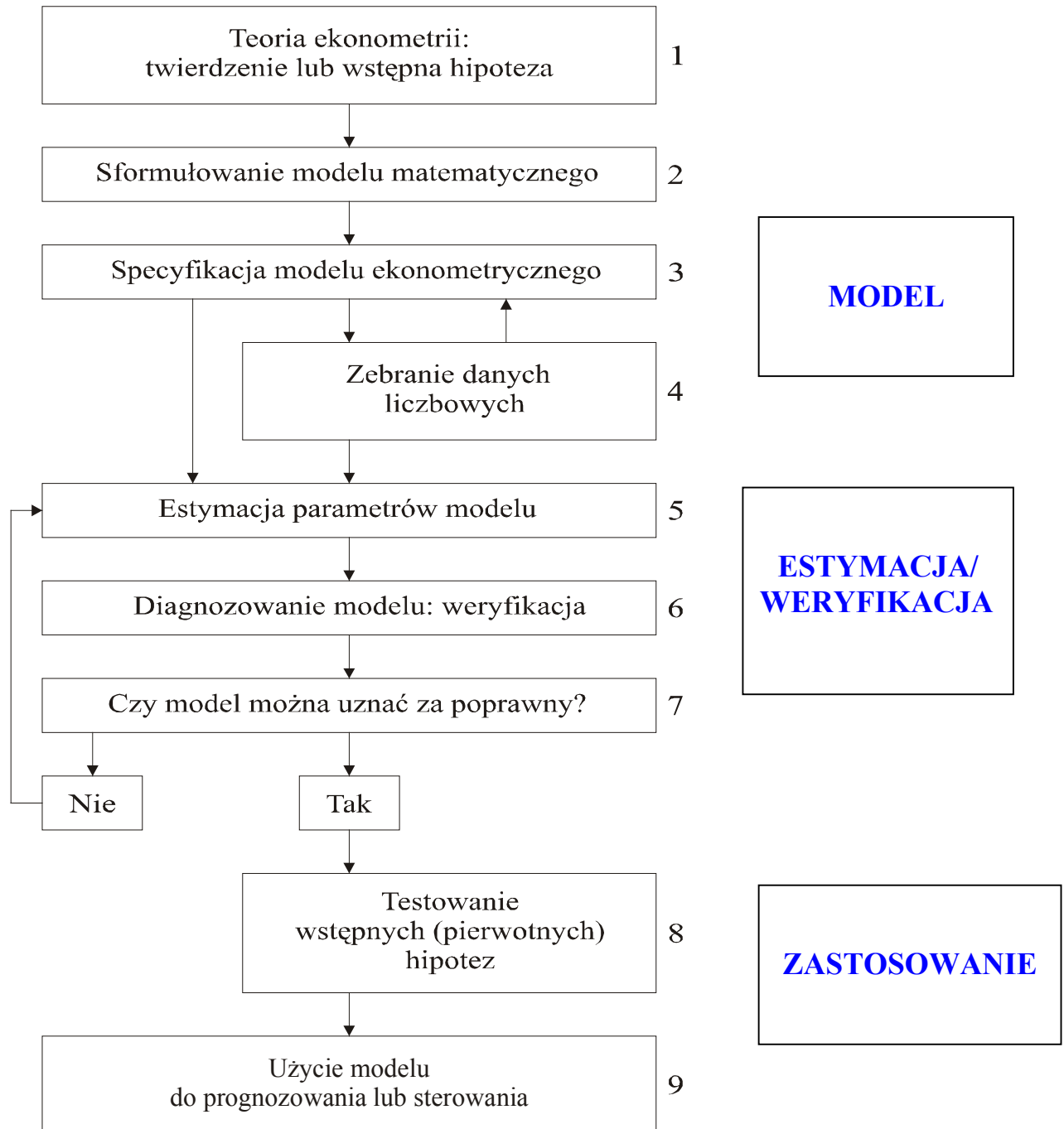
Stąd, otrzymujemy:

$$b = \frac{n \sum_{i=1}^n x_i y_i - \sum_{i=1}^n x_i \sum_{i=1}^n y_i}{n \sum_{i=1}^n x_i^2 - \left( \sum_{i=1}^n x_i \right)^2}.$$

Procedura obliczenia wektora ocen kończy etap estymacji parametrów strukturalnych modelu ekonometrycznego. Analogiczne wyniki do powyższych otrzymamy posługując się metodą macierzową.

Po wyznaczeniu parametrów strukturalnych metodą MNK należy dokonać **weryfikacji modelu** ekonometrycznego pod względem merytorycznym i statystycznym.

# BUDOWA MODELU EKONOMETRYCZNEGO



## **Etap 1. Teoria ekonomii: twierdzenie lub wstępna hipoteza**

Punktem wyjścia do budowy modelu jest istniejąca **teoria** lub **hipoteza** ekonomiczna.

### *Przykłady*

1. Teoria mówi, że popyt na dane dobro jest malejącą funkcją ceny tego dobra.
2. Istnieje związek pomiędzy poziomem inflacji płac oraz poziomem bezrobocia: im niższa inflacja tym wyższe bezrobocie (hipoteza Phillipsa).

Na tym etapie **dokonujemy wyboru istniejącej teorii lub formułujemy własną hipotezę**, którą chcielibyśmy weryfikować w oparciu o informacje statystyczne.



## **Etap 2. Ustalenie modelu matematycznego**

Po określeniu teorii lub wstępnej hipotezy badawczej następuje jej uściślenie matematyczne.

*Przykłady*

**1. funkcja popytu w postaci liniowej:**

$$\text{POPYT} = a + b \text{ CENA} \quad \text{gdzie } b < 0$$

**2. krzywa Phillipsa w postaci hiperboli**

$$INF = \alpha + \beta \frac{1}{BEZ}$$

$$\beta > 0$$

INF – stopa wzrostu płac

**BEZ – stopa bezrobocia**

### **Etap 3. Specyfikacja modelu ekonometrycznego**

Etap specyfikacji to sformułowanie modelu matematycznego w formie nadającej się do weryfikacji empirycznej. Funkcje ustalone w poprzednim etapie uzupełniamy o wielkość reprezentująca zakłócenia losowe:

#### *Przykłady*

1. funkcja popytu w postaci liniowej:

$$\text{POPYT} = a + b \text{ CENA} + u \quad \text{gdzie} \quad b < 0$$

2. krzywa Phillipsa w postaci hiperboli

$$\text{INF} = \alpha + \beta \frac{1}{\text{BEZ}} + u$$

gdzie  $\beta > 0$

## **Etap 4. Zebranie danych liczbowych**

Do oszacowania parametrów modelu potrzebne są dane liczbowe o wartościach zmiennych. Zwykle dane te zostały zebrane przez kogoś innego np. przez urzędy statystyczne.

## **Etap 5. Estymacja parametrów modelu**

Etap estymacji polega na zastosowaniu odpowiednich metod statystycznych w celu otrzymania ocen parametrów modelu na podstawie danych liczbowych o zmiennych.

W wyniku tego etapu następuje **kwantyfikacja**, czyli nadanie wartości liczbowej parametrom postulowanej relacji ekonomicznej.

## **Etap 6 - 7. Diagnozowanie modelu: weryfikacja**

Weryfikacja modelu polega na stwierdzeniu:

- 1.czy otrzymane oszacowanie modelu wytrzymuje konfrontację z oczekiwaniami i teorią (**weryfikacja merytoryczna**)
- 2.czy oszacowanie to statystycznie „mocne”, tzn. czy dane potwierdzają poprawność postulowanego modelu (weryfikacja statystyczna)

## **Etap 8. Testowanie wstępnych (pierwotnych) hipotez**

Jeżeli uzyskany model stanowi akceptowalne oszacowanie rzeczywistości, wówczas wykorzystuje się go do weryfikacji zasadniczych, sformułowanych na wstępie **hipotez badawczych**.

### **Przykład**

W modelu popytu można testować, czy parametr stojący przy zmiennej **CENA** jest mniejszy od zera. Potwierdziłoby to wstępną hipotezę o tym, że w miarę wzrostu ceny popyt maleje.

## **Etap 9. Użycie modelu do prognozowania i sterowania**

Na tym etapie, oszacowany i zweryfikowany model ekonometryczny jest wykorzystywany do:

- 1.formułowania **prognoz** wartości zmiennych objaśnianych
- 2.wyznaczania wartości „instrumentów” (zmiennych objaśniających”) w celu uzyskania żadanego poziomu „celów” (zmiennych objaśnianych); jest to zagadnienie **sterowania**.

## ESTYMACJA

Estymacja polega na szacowaniu parametrów strukturalnych  $\alpha$  występujących przy zmiennej objaśniającej  $X$ . Polega to na sprawdzeniu przeciętnego wzrostu ( $\alpha > 0$ ) lub spadku ( $\alpha < 0$ ) wartości zmiennej objaśnianej  $Y$  jeżeli wartość zmiennej objaśniającej  $X$  zostanie zwiększona o 1 a pozostałe zmienne objaśniające nie ulegną zmianie.

Rozpatrywana jest liniowa zależność zmiennej objaśnianej od zmiennych objaśniających i - tego składnika losowego.

$$Y = \alpha_0 + \alpha_1 X_1 + \alpha_2 X_2 + \dots + \alpha_k X_k + \varepsilon$$

gdzie:  $Y$  - zmienna objaśniana,

$X_j$  - zmienne objaśniające,  $j = 1, 2, \dots, k$ ,

$\alpha_j$  - nieznane parametry strukturalne modelu,  $j = 0, 1, \dots, k$ ,

$\varepsilon$  - składnik losowy.



Celem oszacowania modelu korzysta się z narzędzi statystyki matematycznej. Najważniejsze z nich to:

**Estymator i test statystyczny.**

**Estymatorem** parametru strukturalnego modelu jest pewna funkcja obserwacji dokonanych na zmiennych: objaśnianej i objaśniających. Konkretna wartość tej funkcji nazywa się **oceną** parametru.

Z reguły estymatory otrzymuje się w wyniku zastosowania procedury numerycznej zwanej metodą najmniejszych kwadratów. Mają one wówczas pożądane własności – o ile spełnione są pewne istotne założenia. Główne z nich dotyczą specyfikacji modelu i własności składnika losowego.

## WAŻNE ZAŁOŻENIA

### I. ZAŁOŻENIA DOTYCZĄCE MODELU I DANYCH

#### **Założenie 1**

Model jest liniowy względem parametrów.

$$Y_t = b_0 + b_1 x_{1t} + b_2 x_{2t} + \dots + b_k x_{kt} + u_t$$

#### **Założenie 2**

Zmienne objaśniające  $\mathbf{x}$  są nielosowe.

#### **Założenie 3**

Liczba obserwacji  $\mathbf{n}$  jest większa od liczby parametrów do oszacowania.

W praktyce żądamy by  $\mathbf{n}$  było liczbą kilkakrotnie większą od  $\mathbf{k}+1$  parametrów do oszacowania, np. **3 krotnie**.

#### **Założenie 4**

Żadna ze zmiennych  $\mathbf{x}$  nie jest kombinacją liniową innych zmiennych objaśniających.

## II. ZAŁOŻENIA DOTYCZĄCE SKŁADNIKA LOSOWEGO

Mamy do czynienia z **n** składnikami losowymi, każdy dla innego okresu obserwacji.

### Założenie 1

Składnik losowy ma **wartość oczekiwaną** równą zero (dla wszystkich  $t=1,...,n$ ):

$$E(u_t) = 0$$

Oznacza to, że czynniki nie uwzględnione w modelu nie oddziałują w systematyczny sposób na średnią wartość zmiennej Y: wpływy „dodatnie” i „ujemne” „znoszą się” i średni efekt jest zerowy.

## **Założenie 2**

**Wariancja zmiennej losowej  $u_t$  jest taka sama dla wszystkich obserwacji.**

$$Var(u_t) = \sigma^2$$

dla  $t = 1, \dots, n$ .

Tak więc przyjmujemy, że zmienne losowe **u** mają jednakową dyspersję. Oznacza to, że wpływy czynników nie ujętych w modelu na zmienną **Y** mają takie same rozproszenie, niezależnie od numeru obserwacji.

### Założenie 3

Zmienne losowe  $u_t$  są **nieskorelowane** czyli nie ma tzw. **autokorelacji składników losowych**.

Oznacza to, że wpływy czynników nie ujętych w modelu są nieskorelowane pomiędzy różnymi obserwacjami. Założenie to jest mocne i często nie spełnione w modelach dla szeregów czasowych.

$$\text{cov}(u_t, u_s) = 0$$

$$t \neq s, \quad t = 1, \dots, n; \quad s = 1, \dots, n.$$

**COV** oznacza **kowariancję** pary zmiennych losowych.

## **ZAŁOŻENIA A RZECZYWISTOŚĆ**

Założenia klasyczne okazują się często mało realistyczne dla konkretnego modelu i konkretnego zbioru danych.

Przyjęcie modelu klasycznego jest zwykle **punktem startowym** w analizie ekonometrycznej.

Na późniejszym etapie dokonuje się estymacji parametrów klasycznego modelu.

Ewentualna zmiana założeń następuje na etapie **weryfikacji** modelu.