Ekonometria

Wykład 7

dr hab. Małgorzata Radziukiewicz, prof. PSW Biała Podlaska

Metoda najmniejszych kwadratów w ogólnym przypadku modelu z wieloma zmiennymi objaśniającymi.

Jeśli oszacowania parametrów $\alpha_0, \alpha_1, \ldots, \alpha_k$ oznaczymy odpowiednio przez $a_0, a_1, \ldots, a_k$, wówczas funkcja regresji z próby będzie miała postać:

$$\hat{Y} = a_0 + a_1 X_1 + \ldots + a_k X_k \qquad (3)$$

a podstawiając pod X_j odpowiednią wartość z próby X_{ji} ($j=1,2,\ldots,k$, $i=1,2,\ldots,n$) otrzymujemy:

$$\hat{Y}_i = a_0 + a_1 X_{1i} + \ldots + a_k X_{ki} \qquad i = 1, 2, \ldots, n \qquad (4)$$

Tak jak poprzednio, przez resztę e_i będziemy rozumieć różnicę:

$$e_i = Y_i - \hat{Y}_i \qquad i = 1, 2, \ldots, n \qquad (5)$$

Uwzględniając (4), równość (5) może być zapisana w postaci:

$$Y_i = a_0 + a_1 X_{1i} + \ldots + a_k X_{ki} + e_i \qquad i = 1, 2, \ldots, n \qquad (6)$$

Jeśli estymatory a_j parametrów α_j mają być estymatorami MNK, to muszą być one tak zbudowane, aby suma kwadratów reszt e_i była możliwie najmniejsza, tj.:

$$\sum_{i=1}^{n} e_i^2 = \sum_{i=1}^{n} (Y_i - a_0 - a_1 X_{1i} - \ldots - a_k X_{ki})^2 \rightarrow min \qquad (7)$$

- funkcję (7), traktowaną jako funkcję $k+1$ zmiennych $a_0, a_1, \ldots, a_k$, należy więc zminimalizować poprzez dobór odpowiednich wartości tych zmiennych.
- trzeba zatem wyznaczyć $k+1$ pochodnych funkcji (7), każdą względem kolejnej zmiennej a_j, oraz przyrównać je do zera.
- rozwiązanie otrzymanego układu $k+1$ równań, pozwala na wyznaczenie wartości $a_0, a_1, \ldots, a_k$, które - uzyskane w ten sposób - są estymatorami MNK.
- założenia 1 i 2 gwarantują, że układ $k+1$ równań normalnych, ma jednoznaczne rozwiązanie.

Wprowadzamy następujące oznaczenia:

$$\boldsymbol{Y} = \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix}_{n\times 1} \qquad \boldsymbol{X} = \begin{bmatrix} 1 & X_{11} & X_{21} & \dots & X_{k1} \\ 1 & X_{12} & X_{22} & \dots & X_{k2} \\ & . & . & . & . \dots . \\ & . & . & . & . \dots . \\ 1 & X_{1n} & X_{2n} & \dots & X_{kn} \end{bmatrix}_{n\times(k+1)} \qquad \boldsymbol{\alpha} = \begin{bmatrix} \alpha_0 \\ \alpha_1 \\ . \\ . \\ \alpha_k \end{bmatrix}_{(k+1)\times 1} \tag{8}$$

oraz:

$$\boldsymbol{a} = \begin{bmatrix} a_0 \\ a_1 \\ . \\ . \\ a_n \end{bmatrix}_{(k+1)\times 1} \qquad \boldsymbol{e} = \begin{bmatrix} e_1 \\ e_2 \\ . \\ . \\ e_n \end{bmatrix}_{n\times 1} \qquad \boldsymbol{\varepsilon} = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ . \\ . \\ \varepsilon_k \end{bmatrix}_{n\times 1} \tag{9}$$

Przy przyjętych oznaczeniach, model (2) można zwięźle zapisać jako:

$$\boldsymbol{Y} = \boldsymbol{X\alpha} + \boldsymbol{\varepsilon} \tag{10}$$

Zapis ten przedstawia po prostu n równań (2).

Suma kwadratów reszt (7), którą oznaczymy przez $\boldsymbol{Q}$, w zapisie macierzowym przyjmuje postać:

$$\boldsymbol{Q} = \boldsymbol{e}^T\boldsymbol{e} \tag{11}$$

gdzie symbol T oznacza transpozycję macierzy.

Ponieważ równania (4), w zapisie macierzowym, mają postać:

$$\hat{\boldsymbol{Y}} = \boldsymbol{Xa} \tag{12}$$

wektor $\boldsymbol{e}$ przedstawia się jako różnica dwóch wektorów $\boldsymbol{Y}$ i $\hat{\boldsymbol{Y}}$, tj:

$$\boldsymbol{e} = \boldsymbol{Y} - \hat{\boldsymbol{Y}} = \boldsymbol{Y} - \boldsymbol{Xa} \tag{13}$$

Korzystając z (13), sumę kwadratów reszt (11), można zapisać następująco:

$$Q = (\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{a})^T(\boldsymbol{Y} - \boldsymbol{X}\boldsymbol{a}) = \boldsymbol{Y}^T\boldsymbol{Y} - 2\,\boldsymbol{Y}^T\boldsymbol{X}\boldsymbol{a} + \boldsymbol{a}^T\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} \qquad (14)$$

Powyższą sumę kwadratów należy zminimalizować względem wektora $\boldsymbol{a}$. W tym celu należałoby obliczyć pochodne cząstkowe sumy Q względem kolejnych składowych wektora $\boldsymbol{a}$, a następnie pochodne te przyrównać do zera, otrzymując układ $k+1$ równań z $k+1$ niewiadomymi.

Korzystamy ze wzorów na różniczkowanie w notacji macierzowej.

- formy liniowej $\boldsymbol{c}^T\boldsymbol{x}$ względem wektora $\boldsymbol{x}$:

$$\frac{\partial(\boldsymbol{c}^T\boldsymbol{x})}{\partial\boldsymbol{x}} = \boldsymbol{c}$$

- formy kwadratowej $\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x}$ względem wektora $\boldsymbol{x}$:

$$\frac{\partial(\boldsymbol{x}^T\boldsymbol{A}\boldsymbol{x})}{\partial\boldsymbol{x}} = 2\boldsymbol{A}\boldsymbol{x}$$

Wykorzystując te wzory otrzymujemy wektor pochodnych sumy kwadratów Q, który przyrównujemy do zera:

$$\frac{\partial Q}{\partial \boldsymbol{a}} = -2\boldsymbol{X}^T\boldsymbol{Y} + 2\boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} = \boldsymbol{0} \qquad (15)$$

gdzie $\boldsymbol{0}$ oznacza wektor zerowy.

W rezultacie układ równań normalnych może być zapisany jako:

$$\boldsymbol{X}^T\boldsymbol{Y} = \boldsymbol{X}^T\boldsymbol{X}\boldsymbol{a} \qquad (16)$$

Rozwiązaniem tego układu równań jest następujący wektor $\boldsymbol{a}$:

$$\boldsymbol{a} = (\boldsymbol{X}^T\boldsymbol{X})^{-1}\boldsymbol{X}^T\boldsymbol{Y} \qquad (17)$$

gdzie symbol (-1) oznacza działanie odwracania macierzy.

Dla modelu z jedną zmienną X, ze wzoru (17) wynikają wzory:

$$\boldsymbol{X}^T\boldsymbol{X} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} 1 & X_1 \\ 1 & X_2 \\ . & . \\ . & . \\ 1 & X_n \end{bmatrix} = \begin{bmatrix} n & \sum X_i \\ \sum X_i & \sum X_i^2 \end{bmatrix} \tag{17a}$$

$$\boldsymbol{X}^T\boldsymbol{Y} = \begin{bmatrix} 1 & 1 & \dots & 1 \\ X_1 & X_2 & \dots & X_n \end{bmatrix} \begin{bmatrix} Y_1 \\ Y_2 \\ . \\ . \\ Y_n \end{bmatrix} = \begin{bmatrix} \sum Y_i \\ \sum X_i Y_i \end{bmatrix} \tag{17b}$$

$$(\boldsymbol{X}^T\boldsymbol{X})^{-1} = \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 & -\sum X_i \\ -\sum X_i & n \end{bmatrix} \tag{17c}$$

a stąd:

$$\boldsymbol{a} = \frac{1}{n\sum X_i^2 - (\sum X_i)^2} \begin{bmatrix} \sum X_i^2 \sum Y_i - \sum X_i \sum X_i Y_i \\ -\sum X_i \sum Y_i + n\sum X_i Y_i \end{bmatrix} \tag{17d}$$

- wektor $\boldsymbol{a}$ określony przez wzór (17) jest wektorem estymatorów MNK.
- aby istniało rozwiązanie układu równań (16), macierz $\boldsymbol{X}^T\boldsymbol{X}$, która jest macierzą kwadratową o wymiarach $(k+1)\times(k+1)$, musi być nieosobliwa, gdyż tylko wtedy działanie odwracania macierzy jest wykonalne.
- nieosobliwość macierzy $\boldsymbol{X}^T\boldsymbol{X}$ oznacza, że jej wyznacznik jest różny od zera, tj.:

$$|\boldsymbol{X}^T\boldsymbol{X}| \neq 0 \tag{18}$$

Twierdzenie 1. (Gaussa-Markowa)

W klasycznym modelu regresji liniowej najlepszym nieobciążonym estymatorem wektora α jest wektor otrzymany MNK:

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

o macierzy kowariancji:

$$D^2(a) = \sigma^2(\xi)(X^T X)^{-1} \qquad (2)$$

Twierdzenie 2.

W klasycznym modelu regresji liniowej nieobciążonmym estymatorem wariancji składnika losowego $\sigma^2(\xi)$ jest:

$$S^2(e) = \frac{e^T e}{n-(k+1)}$$

gdzie n - (k+1) jest liczbą stopni swobody (liczba obserwacji pomniejszona o liczbę szacowanych parametrów).

Twierdzenie 3.

W klasycznym modelu regresji liniowej nieobciążonym estymatorem macierzy $D^2(a)$ jest:

$$S^2(a) = S^2(\xi)(X^T X)^{-1}$$

Te błędy średnie informują o ile, średnio rzecz biorąc, pomylilibyśmy się in plus lub in minus szacując ten sam parametr modelu na podstawie różnych zbiorów obserwacji statystycznych, ale składających się zawsze z tej samej liczby obserwacji.

W ten sposób **średni błąd szacunku staje się ważnym elementem informacji w zakresie dokładności oszacowania modelu i umożliwia wydanie sądu, czy otrzymaną ocenę parametru można uważać za wystarczająco dokładną dla celów praktycznych, czy nie**. Dlatego też zwykle obok obliczonych wartości ocen parametru modelu podaje się wartości średnich błędów szacunku.

Zazwyczaj podczas estymacji parametrów modelu ekonometrycznego oblicza się także wartości syntetycznych miar dopasowania modelu do danych empirycznych:

- współczynnika zbieżności,
- współczynnika korelacji wielorakiej i jego kwadratu,
- współczynnika zbieżności resztowej.

Współczynniki te związane są z takimi ocenami parametrów składnika losowego jak wariancja resztowa i odchylenie standardowe reszt.

Weryfikacja modelu ekonometrycznego

Etap VI budowy modelu następuje z chwilą, gdy otrzymano numeryczne oceny parametrów strukturalnych i parametrów stochastycznej struktury modelu.

Etap ten polega na weryfikacji modelu, tj. analizie, czy otrzymane wartości ocen parametrów strukturalnych są rozsądne i czy model z dostateczną dokładnością opisuje wahania zmiennych endogenicznych.

1. Może się zdażyć, że otrzymane oceny parametrów są tego rodzaju, że kłócą się z teorią ekonomiczną lub zdrowym rozsądkiem.
2. Niekiedy zdarza się, że wprawdzie oceny poszczególnych parametrów wyglądają rozsądnie, ale wariancja składnika losowego $D^2(\epsilon)$ jest bardzo duża, tak że odchylenia zmiennej endogenicznej od jej wartości "teoretycznej" wyznaczonej przez model przekreślają możliwość praktycznego wykorzystania modelu.

W rezultacie etap VI jest etapem poprawiania modelu, modyfikacji jego równań, ewentualnej zmiany analitycznej niektórych równań modelu.
Tak poprawiony model poddawany jest ponownemu procesowi estymacji parametrów, poczym następuje ponowna analiza, czy "nowy" model jest poprawny, czy dalej należy go "poprawiać".

Narzędzia weryfikacji.

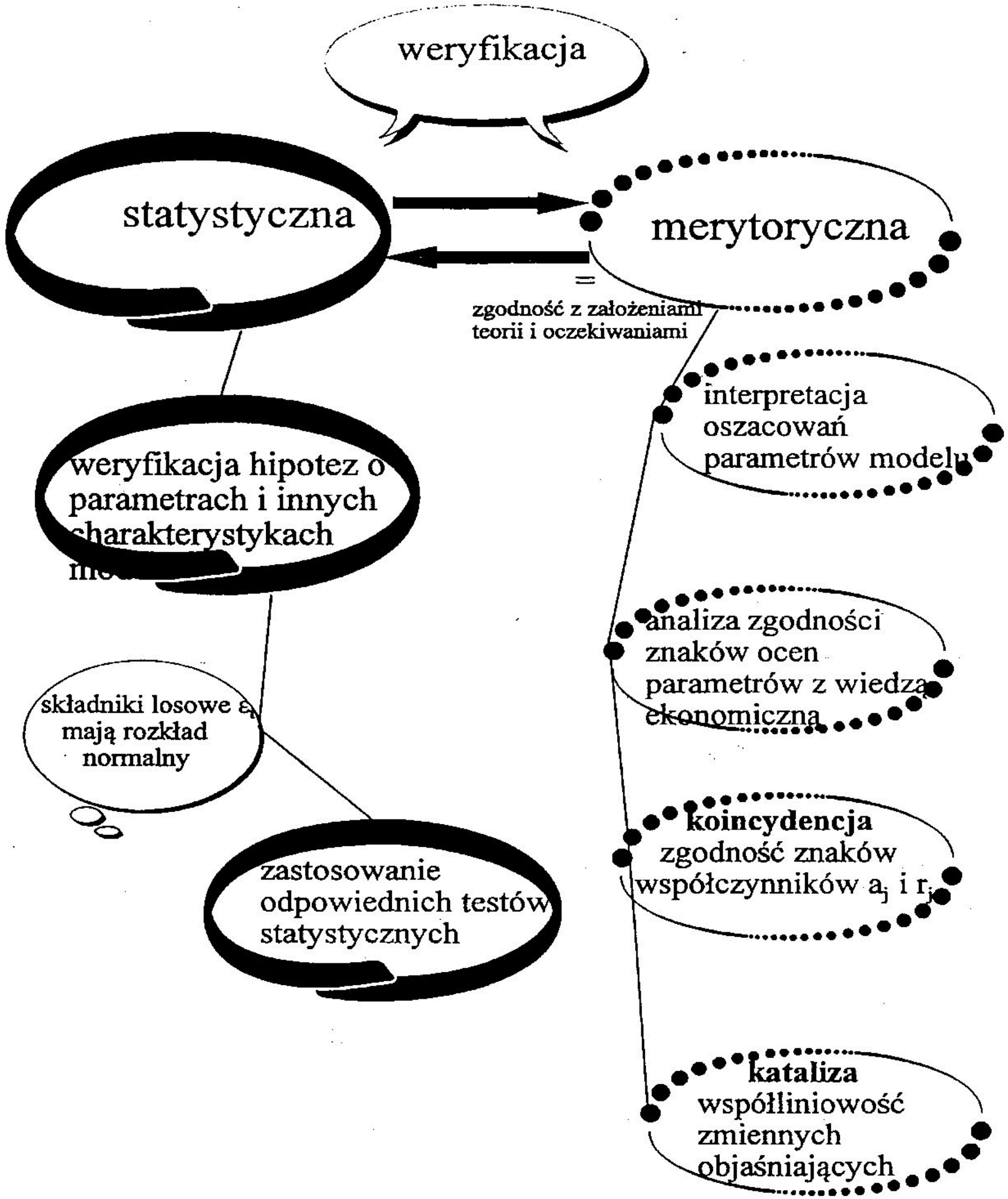
Każdy model powinien być poddany weryfikacji.

W zestawie podstawowych narzędzi weryfikacji są:

(1) **konfrontacja z teorią i oczekiwaniami** - to domena tzw. weryfikacji merytorycznej;

(2) **mierniki dopasowania** - są to **MNK-reszty**, $\mathbf{R^2}$ (i jego odmiany) oraz $\mathbf{S^2}$;

(3) **testy statystyczne** - stosowane w trakcie tzw. weryfikacji statystycznej modelu.


weryfikacja
statystyczna
merytoryczna
=
zgodność z założeniami teorii i oczekiwaniami
weryfikacja hipotez o
parametrach i innych
charakterystykach
składniki losowe ε
mają rozkład
normalny
zastosowanie
odpowiednich testów
statystycznych
interpretacja
oszacowań
parametrów modelu
analiza zgodności
znaków ocen
parametrów z wiedzą
ekonomiczną
koincydencja
zgodność znaków
współczynników a_j i r_j
kataliza
współliniowość
zmiennych
objaśniających

Weryfikacja oszacowanego modelu ekonometrycznego

Weryfikację modelu ekonometrycznego przeprowadza się w dwóch ujeciach:

1. **weryfikacja merytoryczna modelu**;
2. **weryfikacja statystyczna modelu.**

Weryfikacja merytoryczna oszacowanego modelu obejmuje:

- **interpretację oszacowań** parametrów modelu;
- **i ich analizę** (w szczególności bada się **zgodność znaków ocen parametrów** z wiedzą ekonomiczną o modelowanym zjawisku).

Ocena a_j parametru strukturalnego α_j, dla j=1,2,...k, występującego przy zmiennej objaśniającej X_j oznacza, o ile przeciętnie wzrasta (gdy $a_j > 0$) albo zmniejsza się (gdy $a_j < 0$) wartość zmiennej objaśnianej Y, gdy przy niezmienionych wartościach zmiennych objaśniających wartość zmiennej X_j wzrasta o 1 jednostkę.

Wyraz wolny na ogół nie ma znaczącej interpretacji.

Przykład.

- Do modelu wybrano zmienne objaśniające X_1 oraz X_2.
- Macierz obserwacji na zmiennych objaśniających modelu jest postaci:

$$X = \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 2 \\ 1 & 4 & 3 \end{bmatrix}$$

- Wektor wartości zmiennej objaśnianej Y:

$$Y = [10, \quad 12, \quad 13, \quad 15, \quad 20]$$

Twierdzenie 1 (Gaussa-Markowa)

- Wektor ocen parametrów strukturalnych jest postaci:

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

$$X^T X = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 2 & 1 & 2 & 1 & 4 \\ 1 & 1 & 3 & 2 & 3 \end{bmatrix} \cdot \begin{bmatrix} 1 & 2 & 1 \\ 1 & 1 & 1 \\ 1 & 2 & 3 \\ 1 & 1 & 2 \\ 1 & 4 & 3 \end{bmatrix} = \begin{bmatrix} 5 & 10 & 10 \\ 10 & 26 & 23 \\ 10 & 23 & 24 \end{bmatrix}$$

$$a=(X^TX)^{-1}X^TY$$

$$X^TY=\begin{bmatrix}1&1&1&1&1\\2&1&2&1&4\\1&1&3&2&3\end{bmatrix}\cdot\begin{bmatrix}10\\12\\13\\15\\20\end{bmatrix}=\begin{bmatrix}70\\153\\151\end{bmatrix}$$

- Macierz odwrotna do macierzy X^TX

$$(X^TX)^{-1}=\begin{bmatrix}1{,}267&-0{,}133&-0{,}400\\-0{,}133&0{,}267&-0{,}200\\-0{,}400&-0{,}200&0{,}400\end{bmatrix}$$

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

- Obliczamy wartości ocen parametrów strukturalnych modelu ekonometrycznego:

$$a = \begin{bmatrix} 1{,}267 & -0{,}133 & -0{,}400 \\ -0{,}133 & 0{,}267 & -0{,}200 \\ -0{,}400 & -0{,}200 & 0{,}400 \end{bmatrix} \cdot \begin{bmatrix} 70 \\ 153 \\ 151 \end{bmatrix} = \begin{bmatrix} 7{,}941 \\ 1{,}341 \\ 1{,}800 \end{bmatrix}$$

- Model ekonometryczny jest postaci:

$$\hat{Y} = 7{,}941 + 1{,}341 X_1 + 1{,}800 X_2$$

$$\hat{Y} = 7{,}941 + 1{,}341X_1 + 1{,}800X_2$$

- **Interpretacja:**
- a_0 = 7,941 to średnia wartość Y w przypadku, gdy zmienne objaśniające X_1 i X_2 są równe 0;
- a_1 = 1,341 oznacza o ile przeciętnie wzrośnie Y, jeżeli zmienna objaśniająca X_1 wzrośnie o jednostkę, podczas gdy zmienna objaśniająca X_2 pozostanie bez zmian;
- a_2 = 1,800 oznacza, o ile przeciętnie wzrośnie Y, jeżeli zmienna objaśniająca X_2 wzrośnie o jednostkę, podczas gdy zmienna objaśniająca X_1 pozostanie bez zmian.

Weryfikacja = sprawdzenie własności modelu

stopnia "przylegania" modelu

- **precyzja estymatora** (dokładność szacunku parametrów $\mathbf{a_j}$)
  - średni błąd szacunku $S(\mathbf{a_j})$
  - względny średni błąd szacunku $|S(\mathbf{a_j})/\mathbf{a_j}|\cdot 100\%$
- **miary dobroci** dopasowania modelu do danych empirycznych:
  $\mathbf{R^2}$ i jego odmiany, ϕ^2, $\mathbf{S^2(e)}$, $\mathbf{S(e)}$, **V**.

zestawu zmiennych objaśniających

- istotność ocen parametrów $\mathbf{a_j}$
- istotność pojedyńczej zmiennej objaśniającej
  test t-Studenta
- jednoczesna istotność wszystkich zmiennych objaśniających
  test Walda
- model jest koincydentny? ⇒ **nie**
- występuje efekt katalizy? ⇒ **tak**

zmiana zmiennych

rozkładu elementu losowego celem zweryfikowania słuszności założeń

- losowość reszt
  test serii
- symetria składnika losowego
- autokorelacja składnika losowego
  test Durbina-Watsona
  test mnożnika Lagrange'a
- heteroscedastyczność składnika losowego
  test Harrisona-McCabe'a

☻
"Trzy złote zasady ekonometrii to: testować, testować, testować."
D.F.Heangry

Miary dobroci dopasowania modelu do danych empirycznych.

Mają one na celu sprawdzenie, w jakim stopniu oszacowany model wyjaśnia kształtowanie się zmiennej objaśnianej.

- **współczynnik determinacji R^2:**

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

jest miarą określającą, jaka część zmienności zmiennej objaśnianej Y jest wyjaśniana przez model.

Wartość współczynnika R^2 jest liczbą z przedziału $\langle 0, 1 \rangle$.

$R^2=1$ - dopasowanie modelu do danych empirycznych jest doskonałe; $R^2=0$ - bardzo złe.

- **współczynnik zbieżności**

$$\phi^2 = 1 - R^2 \tag{2}$$

wskazujący, jaka część zmienności zmiennej Y nie jest objaśniana przez model (jaka część zmienności zmiennej Y stanowi zmienność odchyleń losowych).

- **odchylenie standardowe reszt modelu**

$$S(e) = \sqrt{S^2(e)} = \sqrt{\frac{\sum_{i=1}^{n} e_i^2}{n-(k+1)}} \tag{3}$$

wkskazuje na przeciętną różnicę między zaobserwowanymi wartościami zmiennej objaśnianej i wartościami teoretycznymi.

- **współczynnik zmienności** V:

$$V = \frac{S(e)}{\bar{y}} \tag{4}$$

informuje, jaka część średniej wartości zmiennej objaśnianej Y stanowi jej odchylenie standardowe.

(model dopuszczalny, gdy wartość V jest bliska zeru).

Twierdzenie 2 (Gaussa-Markowa)

- Wariancja składnika resztowego (estymator wariancji składnika losowego) według wzoru:

$$S^2(e) = \frac{e^T e}{n-(k+1)}$$

- Do obliczenia wariancji potrzebne są reszty:

$$e = \hat{Y} - Y$$

- gdzie:
  - $\hat{Y}$ - wartości teoretyczne zmiennej obajśnianej (uzyskane na podstawie modelu) = wartości przewidywane
  - Y - wartości zmiennej objaśnianej (empiryczne)

Ile wynoszą reszty?

- Do oszacowanego modelu:

$$\hat{Y} = 7{,}941 + 1{,}341 * X_1 + 1{,}800 * X_2$$

- podstawiamy kolejne wartości zmiennych X_1 i X_2

$$\hat{Y}_1 = 7{,}941 + 1{,}341 * 2 + 1{,}800 * 1 = 12{,}423$$
$$\hat{Y}_2 = 7{,}941 + 1{,}341 * 1 + 1{,}800 * 1 = 11{,}082$$
$$\hat{Y}_3 = 7{,}941 + 1{,}341 * 2 + 1{,}800 * 3 = 16{,}023$$
$$\hat{Y}_4 = 7{,}941 + 1{,}341 * 1 + 1{,}800 * 2 = 12{,}882$$
$$\hat{Y}_5 = 7{,}941 + 1{,}341 * 4 + 1{,}800 * 3 = 18{,}705$$

$$\hat{Y} = \begin{bmatrix} 12{,}423 \\ 11{,}082 \\ 16{,}023 \\ 12{,}882 \\ 18{,}705 \end{bmatrix}$$

- Wektor reszt $e = \hat{Y} - Y$ równa się:

$$e_1 = 12{,}423 - 10 = 2{,}423$$
$$e_2 = 11{,}082 - 12 = -0{,}918$$
$$e_3 = 16{,}023 - 13 = 3{,}023$$
$$e_4 = 12{,}882 - 15 = -2{,}118$$
$$e_5 = 18{,}705 - 20 = -1{,}295$$

$$e = \begin{bmatrix} 2{,}423 \\ -0{,}918 \\ 3{,}023 \\ -2{,}118 \\ -1{,}295 \end{bmatrix}$$

$$S^2(e) = \frac{e^T e}{n - (k+1)}$$

- licznik wzoru to:

$$e^T e = \sum e^2$$

$$S^2(e) = \frac{22{,}014}{n-(k+1)} = \frac{22{,}014}{2} = 11{,}007$$

- Odchylenie standardowe składnika resztowego (**błąd estymacji**):

$$S(e) = \sqrt{S^2(e)} = \sqrt{11{,}007} = 3{,}318$$

- **Interpretacja:**
- Poszczególne obserwacje empiryczne Y odchylają się średnio od teoretycznych o ± 3,318 jednostek.

Twierdzenie 3 (Gaussa-Markowa

- Wariancja estymatora parametrów strukturalnych według wzoru:

$$D^2(a) = S^2(e) \cdot (X^T X)^{-1}$$

wynosi:

$$D^2(a) = 11{,}007 \cdot \begin{bmatrix} 1{,}267 & -0{,}133 & -0{,}400 \\ -0{,}133 & 0{,}267 & -0{,}200 \\ -0{,}400 & -0{,}200 & 0{,}400 \end{bmatrix}$$

Obliczając wartości elementów diagonalnych macierzy $D^2(a)$ otrzymamy oceny wariancji poszczególnych parametrów modelu

Wnioskowanie o dokładności szacunku parametrów a_i

- **Błędy średnie** szacunku parametrów strukturalnych:

$$S(a_0)=\sqrt{11{,}007\cdot 1{,}267}=\sqrt{13{,}946}=3{,}734\approx 3{,}7$$

$$S(a_1)=\sqrt{11{,}007\cdot 0{,}267}=\sqrt{2{,}939}=1{,}714\approx 1{,}7$$

$$S(a_2)=\sqrt{11{,}007\cdot 0{,}400}=\sqrt{4{,}403}=2{,}098\approx 2{,}1$$

- **Interpretacja:**

  O ile +- odchylają się wartości ocen parametrów strukturalnych od ich wartości rzeczywistych

- Do interpretacji lepiej posługiwać się średnimi względnymi błędami szacunku parametrów wyznaczonymi ze wzoru:

$$\left|\frac{S(a_0)}{a_0}\right| \cdot 100\% = \frac{3{,}734}{7{,}941} \cdot 100\% = 47{,}02\%$$

$$\left|\frac{S(a_1)}{a_1}\right| \cdot 100\% = \frac{1{,}714}{1{,}341} \cdot 100\% = 127{,}82\%$$

$$\left|\frac{S(a_2)}{a_2}\right| \cdot 100\% = \frac{2{,}089}{1{,}800} \cdot 100\% = 116{,}06\%$$

Błędy średnie stanowią odpowiednio 47,02%, 127,82% oraz 116,06% wartości kolejnych parametrów.

Współczynnik zbieżności dany wzorem:

$$\varphi^2 = \frac{Y^T Y - Y^T X \cdot a}{(Y - \bar{Y})^T (Y - \bar{Y})} = \frac{[n - (k+1)] \cdot S^2(e)}{(Y - \bar{Y})^T (Y - \bar{Y})}$$

wynosi: $$\varphi^2 = \frac{(5-3) \cdot 11{,}007}{58} = 0{,}380$$

bowiem:

$$\bar{Y} = \frac{70}{5} = 14; \quad (Y - \bar{Y}) = \begin{bmatrix} 10-14 \\ 12-14 \\ 13-14 \\ 15-14 \\ 20-14 \end{bmatrix} = \begin{bmatrix} -4 \\ -2 \\ -1 \\ 1 \\ 6 \end{bmatrix}$$

$$(Y - \bar{Y})^T (Y - \bar{Y}) = \begin{bmatrix} -4 & -2 & -1 & 1 & 6 \end{bmatrix} \cdot \begin{bmatrix} -4 \\ -2 \\ -1 \\ 1 \\ 6 \end{bmatrix} = 58$$

- Współczynnik zbieżności $\varphi^2 = 0{,}380$ oznacza, iż **38%** zmienności zmiennej objaśnianej Y nie zostało wyjaśnione przez model.
- **Współczynnik determinacji R^2 :**

$$R^2 = 1 - \varphi^2 = 1 - 0{,}380 = 0{,}620$$

co oznacza, iż **62%** zmienności zmiennej objaśnianej Y zostało wyjaśnione przez model

- **Współczynnik zmienności losowej:**

$$V = \frac{S(e)}{\bar{Y}} \cdot 100\% = \frac{3{,}318}{14} \cdot 100\% = 23{,}7\%$$

- **Interpretacja:**
- Odchylenia losowe stanowią 23,7% wartości średniej zmiennej objaśnianej Y.

- W ekonometrii przyjęta jest konwencja podawania średnich błędów szacunku parametrów strukturalnych łącznie z oszacowaniem modelu.
- Oszacowany model ekonometryczny jest postaci:

$$\hat{Y} = \underset{(3,734)}{7{,}941} + \underset{(1,714)}{1{,}341}X_1 + \underset{(2,089)}{1{,}800}X_2 \qquad R^2 = 0{,}62$$

Istotność zmiennych objaśniających

Test T-Studenta

Przy badaniu istotności wpływu zmian j-tej zmiennej objaśniającej na zmiany wartości zmiennej objaśnianej (lub **istotności parametru a_j**) zastosujemy **test istotności**.

Testowana jest:

hipoteza zerowa **H_0 :** $\alpha_j = 0$ wobec

hipotezy alternatywnej **H_1 :** $\alpha_j \neq 0$.

Jeśli spełnione jest założenie 7 - że składnik losowy ϵ w każdym z okresów ma rozkład normalny o wartości oczekiwanej zero i skończonej, stałej wariancji $\epsilon_i = N(0,\sigma^2)$ dla i=1,2...n - oraz jeśli prawdziwa jest hipoteza H_0 to zmienna losowa:

$$t = \frac{a_j}{S(a_j)}$$

ma **rozkład t-Studenta** z **n–(k+1)** stopniami swobody.

Hipoteza H_0 zakłada, że parametr α_j nieistotnie różni się od zera, tzn. że zmienna X_j przy której on stoi, wywiera nieistotny wpływ na Y (lub, że zmienna X_j jest nieistotna dla rozpatrywanego modelu).

Odrzucenie hipotezy H_0, oznacza przyjęcie hipotezy alternatywnej, że wartość parametru α_j istotnie różni się od zera (zmienna X_j wywiera istotny wpływ na zmienną objaśnianą).

Wartość krytyczną testu t-Studenta t^* odczytujemy z tablic rozkładu t-Studenta przy poziomie istotności α=0.05 i n-(k+1) stopniach swobody.
Hipotezę H_0 odrzucamy, jeśli $|t|>t^*$.

Weryfikujemy istotność parametrów strukturalnych oszacowanego modelu

- Stawiamy hipotezę:
- $H_0: \alpha_i = 0$ (parametr α_i nieistotnie różni się od zera tzn. że zmienna X_i przy której parametr stoi wywiera nieistotny wpływ na zmienną objaśnianą);
- $H_1: \alpha_i \neq 0$ (parametr α_i istotnie różni się od zera);
- Test istotności pozwalający na weryfikację hipotezy $H_0: \alpha_i = 0$ oparty jest na rozkładzie statystyki t-Studenta określonej wzorem:

$$t_{\alpha_i} = \frac{a_i - \alpha_i}{S(a_i)}$$

- Dla każdego parametru obliczamy wartości empiryczne statystyki t:

$$t_{(a_0)} = \frac{7{,}941}{3{,}734} = 2{,}127$$

$$t_{(a_1)} = \frac{1{,}341}{1{,}714} = 0{,}782$$

$$t_{(a_2)} = \frac{1{,}800}{2{,}089} = 0{,}862.$$

- Z tablic t-Studenta dla przyjętego poziomu istotności $\alpha = 0{,}01$ oraz dla $n-(k+1)= 5-(2+1)=2$ stopnie swobody odczytujemy wartość krytyczną $t^* = 4{,}303$.

- Jeżeli spełniona jest nierówność:

$$\left|t_{(\alpha)}\right| > t^{*}$$

to hipoezę H_0 należy odrzucić na korzyśćalternatywnej hipotezy H_1, czyli dany parametr jest statystycznie istotny.

- W przypadku, gdy:

$$\left|t_{(\alpha)}\right| \le t^{*}$$

nie ma odstaw do odrzucenia hipotezy H_0 o nieistotności parametru.

- Z naszych obliczeń wynika m.in., iż:

$$t_{(a_0)} < t^*$$

więc hipotezę H_1 odrzucamy, a parametr a_0 jest statystycznie nieistotny.

- Dla parametrów a_1 i a_2 spełniona jest również nierówność:

$$t_{(a_1)}, t_{(a_2)} < t^*$$

co oznacza, iż w tym przypadku również nie ma podstaw do odrzucenia hipotezy H_0.

- **Interpretacja:**

Parametry a_0, a_0 i a_2 są statystycznie nieistotne. A zatem zmienne objaśniające X_1 i X_2 wywierają nieistotny wpływ na zmienną objaśnianą Y.

Własność koincydecji.

Jeśli modelujemy zjawisko, o którym teoria ekonomii nie stanowi, wówczas analizę merytoryczną można oprzeć na obliczeniu i interpretacji współczynników korelacji między zmiennymi występującymi w modelu.

Zestawiamy następujące macierze:

- współczynników korelacji między zmiennymi Y i X_i $\mathbf{R_0} = \{r_i\}$ o wymiarach k x 1;
- współczynników korelacji między zmiennymi X_i oraz X_j $\mathbf{R} = (r_{ij})$ o wymiarach k x k.

Żądamy, by szacowany model był koincydentny.

Mówimy, że oszacowany model jest koincydentny, jeśli dla każdej zmiennej objaśniającej modelu spełniony jest warunek:

$$\textbf{signum } r_i = \textbf{signum } b_i$$

gdzie: signum - funkcja określająca znak swojego argumentu.

Jeśli model nie jest koincydentny, wtedy należy zmienić zestaw zmiennych objaśniających albo analityczną postać modelu.

Przykład . Współczynnik korelacji liniowej ceny mieszkania i jego powierzchni ($r_{C,POW}$)=0.75, współczynnik korelacji liniowej ceny mieszkania i liczby sypialni ($r_{C,SYP}$)=0.01 oraz ceny mieszkania i liczby łazienek ($r_{C,WAN}$)=0.55.

Model opisujący wysokość ceny mieszkania nie jest koincydentny, bo znak przy ocenie parametru stojącego przy zmiennej SYP jest różny od znaku odpowiedniego współczynnika korelacji.

Badanie **koincydencji**

- Model jest koincydentny, jeżeli dla każdej zmiennej objaśniającej model zachodzi:

$$\operatorname{sgn} r_i = \operatorname{sgn} a_i$$

gdzie:

- a_i – jest oceną parametru strukturalnego α_i;
- r_i – jest współczynnikiem korelacji między zmienną Y a zmienną X_i.

$$R_0 = \begin{bmatrix} r_{YX_1} \\ r_{YX_2} \end{bmatrix} = \begin{bmatrix} 0{,}697 \\ 0{,}722 \end{bmatrix}$$

Model jest koincydentny.

Współliniowość – **czy zmienne są katalizatorami**?

- Zmienna X_i z pary zmiennych (X_i, X_j) jest katalizatorem jeżeli:

$$r_{ij} < 0 \qquad \text{lub} \qquad r_{ij} > \frac{r_i}{r_j}$$

- Z obliczeń wynika, iż:

$$r_{12} < \frac{r_1}{r_2} \qquad bo \qquad \frac{r_1}{r_2} = \frac{0{,}697}{0{,}722} = 0{,}965$$

$$r_{21} < \frac{r_2}{r_1} \qquad bo \qquad \frac{r_2}{r_1} = \frac{0{,}722}{0{,}697} = 1{,}036$$

Żadna ze zmiennych nie jest katalizatorem.

Weryfikacja statystyczna modelu

Drugą częścią weryfikacji modelu jest weryfikacja statystyczna. Trudno jest podać algorytm postępowania, niełatwo bowiem ustalić zakres, sposób i kolejność testowania własności statystycznych modelu.

Podczas weryfikacji statystycznej głównym przedmiotem zainteresowania jest wektor reszt modelu e = y - Xa.

Uważa się go za empiryczną realizację składnika losowego modelu.

Przy pomocy odpowiednich testów przeprowadza się m.in. następujące badania:

- badanie losowości składnika losowego;
- badanie symetrii rozkładu składnika losowego;
- badanie stacjonarności składnika losowego;
- badanie wartości oczekiwanej składnika losowego;
- badanie autokorelacji składnika losowego;
- badanie homoscedastyczności składnika losowego;
- badanie normalności składnika losowego.

Weryfikacja składnika losowego modelu ekonometrycznego składa się więc z wielu kroków. Niepowodzenie w badaniu jakiejkolwiek pożądanej cechy składnika losowego powinno spowodować powrót do wcześniejszych etapów modelowania ekonometrycznego. W praktyce, kompromis polega na przyjęciu modelu gorzej oszacowanego, ale mającego inne korzystne cechy.

Badanie losowości

- Badanie losowości ma związek z wyborem postaci analitycznej modelu.
- W standardowym modelu liniowym zmienna objaśniana jest liniową funkcją zmiennych objaśniających plus korekta.
- W przypadku, gdy korekty mają przez dłuższy okres jednakowe znaki można przypuszczać, że został popełniony **błąd specyfikacji**:
  - nietrafny wybór postaci analitycznej modelu;
  - nietrafny wybór zmiennych objaśniających

Badanie losowości reszt.

Weryfikację losowości reszt przeprowadzamy za pomocą testu serii.

Wyznaczonym resztom $\mathbf{e_t}$ przypisujemy symbole: **a - jeżeli $\mathbf{e_t > 0}$**, oraz **b - jeżeli $\mathbf{e_t < 0}$**. Reszt dokładnie równych zeru nie bierze się pod uwagę. Otrzymamy w ten sposób ciąg złożony z symboli a i b , w których można zauważyć serie (**seria to podciąg złożony z kolejnych elementów jednego rodzaju**).

Stąd określamy **liczbę serii tzw. k** empiryczne.

Z tablic liczby serii odczytujemy dla $\mathbf{n_1}$ **(liczba symboli a)** i $\mathbf{n_2}$ **(liczba symboli b)** oraz przyjętego poziomu istotności α, taką wartość krytyczną $\mathbf{k_\alpha}$, że

$P\{ k \le k_\alpha \} = \alpha$.

Jeżeli $k_{emp} \le k_\alpha$, to hipotezę o losowości reszt należy odrzucić (jest to jednoznaczne z koniecznością zmiany postaci analitycznej modelu).

Jeżeli $k_{emp} > k_\alpha$ nie ma podstaw do odrzucenia hipotezy o losowości reszt.

Czy reszty są losowe?

Wektor reszt **Reguły testu** (dla prób małych ($n \le 30$)

$$e = \begin{bmatrix} 2{,}423 \\ -0{,}918 \\ 3{,}023 \\ -2{,}118 \\ -1{,}295 \end{bmatrix}$$

- Przypisujemy resztom e_k symbole a, gdy $e_k > 0$, oraz b gdy $e_k < 0$
- Otrzymujemy ciąg złożony z symboli a i b
  - a, b, a, b, b.
- Określamy liczbę serii k_{emp}

$$k_{emp} = 4$$

Z tablic liczby serii dla n_1 = liczba symboli a i n_2 = liczba symboli b oraz przyjętego $\alpha = 0{,}05$ odczytujemy wartośc $t_\alpha = 2$

Wobec $k_{emp} > k_\alpha$ nie ma podstaw do odrzucenia hipotezy, że rozkłsd reszt jest losowy

Wartości krytyczne testu serii

Rozkład serii

$\alpha = 0,05$.

n_2 \ n_1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20
2																			
3																			
4			2																
5		2	2	3															
6		2	3	3	3														
7		2	3	3	4	4													
8	2	2	3	3	4	4	5												
9	2	2	3	4	4	5	5	6											
10	2	3	3	4	5	5	6	6	6										
11	2	3	3	4	5	5	6	6	7	7									
12	2	3	4	4	5	6	6	7	7	8	8								
13	2	3	4	4	5	6	6	7	8	8	9	9							
14	2	3	4	5	5	6	7	7	8	8	9	9	10						
15	2	3	4	5	6	6	7	8	8	9	9	10	10	11					
16	2	3	4	5	6	6	7	8	8	9	10	10	11	11	11				
17	2	3	4	5	6	7	7	8	9	9	10	10	11	11	12	12			
18	2	3	4	5	6	7	8	8	9	10	10	11	11	12	12	13	13		
19	2	3	4	5	6	7	8	8	9	10	10	11	12	12	13	13	14	14	
20	2	3	4	5	6	7	8	9	9	10	11	11	12	12	13	13	14	14	15

Badanie symetrii składnika resztowego.

Obserwacje odchylając się in plus (in minus) powinny stanowić połowę (w sensie probabilistycznym) wszystkich obserwacji.

Formułujemy hipotezę:

H_0 : składnik resztowy ma rozkład symetryczny:

$$H_0 : \left(\frac{m}{n} = \frac{1}{2} \right)$$

wobec hipotezy alternatywnej:

H_1 : że rozkład składnika resztowego nie jest symetryczny,

$$H_1 = \left(\frac{m}{n} \neq \frac{1}{2} \right)$$

przy czym: m - liczba reszt odchylających się in plus,

n - liczba wszystkich obserwacji.

Do weryfikacji hipotezy H_0 służy statystyka:

$$t_{emp} = \frac{\left| \frac{m}{n} - \frac{1}{2} \right|}{\sqrt{\frac{\frac{m}{n}\left(1 - \frac{m}{n}\right)}{n-1}}}$$

Jeśli $n \leq 30$, statystyka t ma przy założeniu prawdziwości hipotezy H_0 rozkład Studenta o n-1 stopniach swobody, natomiast dla $n > 30$ statystyka ma rozkład normalny.

Czy rozkład reszt modelu jest symetryczny?

W celu zweryfikowania hipotezy

- przyjęto poziom istotności testu **α = 0,05:**
- **m = 2** - liczba reszt dodatnich
- **n = 5** – całkowita liczba reszt
- następnie obliczono wartość statystyki testowej t_{emp} = 1,67
- Dla n-1=5-1=4 stopni swobody wartość t*= 2,776.

- Odp. Rozkład reszt jest losowy, bowiem 1,67<2,776

Z tablic testu t Studenta

- dla przyjętego poziomu istotności α oraz dla n-1 stopni swobody odczytuje się wartość krytyczną t*
- . Jeżeli $|t_{emp}| \leq t^*$, nie ma podstaw do odrzucenia hipotezy H_0 i rozkład reszt modelu jest symetryczny.

Czy występuje autokorelacja skladnika losowego?

- Jednym z założeń dotyczących modelu regresji jest niezależność błędów obserwacji, czyli fakt, czy występujące reszty w predykcji zmiennej zależnej są ze sobą skorelowane.
- Dobrze dopasowane modele regresji zakładają, że otrzymywane reszty (e) - błędy przewidywania rzeczywistej wartości zmiennej zależnej na podstawie utworzonego przez nas modelu regresji - są niezależne od siebie,
- Oznacza to, że rozkład reszt jest losowy, przypadkowy, bez stale występującego wzorca.

- Sposobem określenia niezależności błędów obserwacji jest wyznaczenie **autokorelacji składnika resztowego**, czyli korelacji r-Pearsona pomiędzy kolejnymi resztami, powstałymi z nieidealnego dopasowania modelu.

$$\rho_i = \frac{\mathrm{cov}(\varepsilon_t, \varepsilon_{t-i})}{D(\varepsilon_t)D(\varepsilon_{t-i})}$$

zależność korelacyjna składników losowych ε_t oraz ich pierwszych opóźnień ε_{t-i}

Współczynnik korelacji Pearsona

- **r_{xy}** jest miernikiem związku liniowego między dwiema cechami (zmiennymi) mierzalnymi
- jest wyznaczany poprzez standaryzację kowariancji
- **kowariancja** (wariancja wspólna cech x i y) jest średnią arytmetyczną iloczynu odchyleń wartości liczbowych tych cech (zmiennych) x i y od ich średnich arytmetycznych

$$r_{xy} = \frac{\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y})}{n \cdot S(x) \cdot S(y)}$$

$$r_{xy} = \frac{\operatorname{cov}(x,y)}{S(x) \cdot S(y)}$$

$$\operatorname{cov}(x,y) = \operatorname{cov}(y,x) = \frac{1}{n}\sum_{i=1}^{n}(x_i - \overline{x})(y_i - \overline{y}) = \overline{x \cdot y} - \overline{x} \cdot \overline{y}$$

Proces autokorelacji rzędu I

- Załóżmy, że składniki losowe ε_t związane są zależnością:

$$\varepsilon_t = \rho\,\varepsilon_{t-1} + \eta_t \qquad |\rho| < 1$$

gdzie: (t=1...,n-1)

zmienne losowe η są niezależne i mają jednakowy rozkład

Test Durbina-Watsona

- Test Durbina-Watsona (statystyka) służy do oceny występowania korelacji pomiędzy resztami (błędami, składnikami resztowymi).
- Sprawdzamy, czy składniki losowe modelu pochodzą z procesu autokorelacji rzędu I.
- Przyczyną występowania zjawiska autokorelacji składnika losowego w modelu są:
  - natura procesów ekonomicznych (skutki pewnych zdarzeń albo decyzji rozciągaja sie na wiele okresów;
  - niepoprawna postać analityczna modelu;
  - niepełny zestaw zmiennych objasniających.

Badanie autokorelacji reszt.

Test Durbina-Watsona.

Weryfikowaną hipotezą jest:

$H_0 : \rho = 0$ (brak autokorelacji reszt modelu)

$H_1 : \rho > 0$ (autokorelacja reszt występuje)

Sprawdzianem hipotezy jest jest statystyka d zdefiniowana wzorem:

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

Wartości krytyczne tej statystyki d_L i d_U odczytuje się z tablic Durbina-Watsona dla przyjętego poziomu istotności α oraz n i k stopni swobody (n-liczba obserwacji, k-liczba zmiennych objaśniających w modelu).

Na podstawie tablicy 1. wyznaczono empiryczną wartość $d_{emp}=2.499$

Hipotezę alternatywną trzeba doprecyzować w zależności od otrzymanej wartości d_{emp}.

Jeżeli $d < 2$, to $H_1 : \rho_1 > 0$ (autokorelacja dodatnia),

jeżeli zaś $d > 2$, to $H_1 : \rho_1 < 0$ (autokorelacja ujemna).

Przy autokorelacji ujemnej należy wyznaczyć d' według wzoru:

$d' = 4 - d_{emp}$

W przypadku postawienia hipotezy alternatywnej o istnieniu autokorelacji dodatniej przy:

$d_{emp} < d_L$ hipotezę H_0 należy odrzucić,

$d_{emp} > d_U$ to nie ma podstaw do odrzucenia hipotezy H_0,

$d_L \le d_{emp} \le d_U$ wtedy nie możemy podjąć żadnej decyzji.

W przypadku, gdy stawiamy hipotezę alternatywną o istnieniu autokorelacji ujemnej analogiczne wnioskowanie dotyczy d'.

Tablice testu Durbina-Watsona prezentują wartości krytyczne d_L oraz d_U dla odpowiedniej liczby obserwacji *n* oraz liczby zmiennych objaśniających *k*

Tablica 12. Wartości krytyczne testu Durbina-Watsona

n	α = 0,01				n	α = 0,05			
	k = 1		k = 2			k = 1		k = 2	
	d_L	d_U	d_L	d_U		d_L	d_U	d_L	d_U
7	0,435	1,036	0,294	1,675	7	0,700	1,356	0,467	1,896
8	0,497	1,003	0,345	1,489	8	0,763	1,332	0,467	1,896
9	0,554	0,998	0,409	1,389	9	0,824	1,320	0,629	1,699
10	0,504	1,001	0,466	1,333	10	0,879	1,320	0697	1,641
11	0,653	1,010	0,519	1,297	11	0,927	1,324	0,758	1,604
12	0,696	1,023	0,569	1,274	12	0,971	1,331	0,812	1,579
13	0,738	1,038	0,616	1,261	13	1,010	1,340	0,861	1,562
14	0,776	1,054	0,660	1,254	14	1,045	1,350	0,905	1,551
15	0,811	1,070	0,700	1,252	15	1,077	1,361	0,946	1,543
16	0,844	1,086	0,737	1,252	16	1,106	1,371	0,982	1,539
17	0,874	1,102	0,772	1,255	17	1,133	1,381	1,015	1,536
18	0,902	1,118	0,805	1,259	18	1,158	1,391	1,046	1,535
19	0,928	1,132	0,834	1,265	19	1,180	1,401	1,074	1,536
20	0,952	1,147	0,863	1,271	20	1,201	1,411	1,100	1,537
21	0,975	1,161	0,890	1,277	21	1,221	1,420	1,125	1,538
22	0,997	1,174	0,914	1,284	22	1,239	1,429	1,147	1,541

Czy występuje autokorelacja reszt?

Statystyka d

- Dla modelu wartość:

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

- d= 53,501/22,015=2,430

Obliczenia:

- Zasadą jest, że **wartości statystyk testowych w zakresie od 1,5 do 2,5 są stosunkowo normalne.**
- Każda wartość spoza tego zakresu może być powodem do obaw.
- Statystyka Durbina – Watsona, chociaż wyświetlana przez wiele programów analizy regresji, nie ma zastosowania w niektórych sytuacjach.
- Np. gdy opóźnione zmienne zależne są zawarte w zmiennych objaśniających, niewłaściwe jest użycie tego testu.

Konstrukcja i weryfikacja modelu ekonometrycznego

Krok 1: Określić zmienną objaśnianą Y i zbiór kandydatek na zmienne objaśniające $X_1, X_2, \ldots X_k$. Zgromadzić niezbędne dane statystyczne.

Krok 2: Przeprowadzić procedurę doboru zmiennych objaśniających (metodą Hellwiga, metodą grafu, metodą ekspercką).

Krok 3: Zdefiniować jednorównaniowy liniowy model ekonometryczny (jeśli nieliniowy, to linearyzujemy).

Krok 4: Oszacować parametry modelu metodą najmniejszych kwadratów.

Krok 5: Wyznaczyć reszty modelu.

Krok 6: Czy reszty mają rozkład normalny?
TAK→ krok 9 NIE → krok 7

Krok 7: Czy reszty mają inny znany rozkład?
TAK → krok 8 NIE → STOP

Krok 8: Oszacować parametry modelu metodą największej wiarygodności → STOP

Krok 9: Czy występuje zjawisko autokorelacji składnika losowego modelu?
TAK →krok 10 NIE → krok11

Krok 10: Oszacować parametry modelu metodą Cochrane-Orcutta i przejść do kroku 5.

Krok 11: Czy występuje zjawisko heteroskedastyczności składnika losowego modelu?
TAK → krok 12 NIE → krok 13

Krok 12: Oszacować parametry modelu ważoną metodą najmniejszych kwadratów i przejść do kroku 5.

Krok 13: Czy model ekonometryczny jest liniowy?
TAK → krok 15 NIE → krok 14

Konstrukcja i weryfikacja modelu ekonometrycznego (c.d.)

Krok 14: Zmienić postać analityczną modelu ekonometrycznego. Jeśli jest to niezbędne dokonać linearyzacji modelu i przejść do kroku 4. Jeśli nowy model jest ściśle nieliniowy, to dalsze postępowanie nie mieści się w tej procedurze → STOP

Krok 15: Czy występuje zjawisko współliniowości?
TAK →krok 16 NIE → krok 17

Krok 16: Oszacować parametry modelu metodą regresji grzbietowej i przejść do kroku 5.

Krok 17: Czy wszystkie zmienne objaśniające są istotne statystycznie?
TAK → krok 19 NIE → krok 18

Krok 18: Zmienić zestaw zmiennych objaśniających i przejść do kroku 4.

Krok 19: Czy można zaakceptować wartość współczynnika determinacji?
TAK → krok 20 NIE → krok 18

Krok 20: Czy występuje efekt katalizy?
TAK → krok 18 NIE → krok 21

Krok 21: Czy można zaakceptować interpretację wartości oszacowań parametrów modelu?
TAK → krok 22 NIE → krok 18

Krok 22: Wykorzystać oszacowany i zweryfikowany model ekonometryczny → STOP

Model zweryfikowany?

1. Model ekonometryczny, który:

- wywodzi się z określonej teorii i modelu matematycznego;
- został oszacowany na podstawie właściwie dobranych danych liczbowych;
- przeszedł "zwycięsko" podstawowe próby weryfikacji merytorycznej i statystycznej;

może być wykorzystany do celu, w jakim został "stworzony".

2. Takim modelem posługujemy się na ogół jednorazowo.

3. Upływ czasu powoduje konieczność skorzystania z nowych informacji o modelowanej zależności:

- ulega wzbogaceniu zbiór danych liczbowych, bowiem pojawiają się nowe obserwacje;
- obserwowane w przeszłości związki między zmiennymi objaśniającymi a zmienną objaśnianaą mogą zmienić siłę lub kierunek itp.

4. Profesjonalne modele ekonometryczne są okresowo uaktualniane i modyfikowane. Ich twórcy "monitorują" zachowanie się modelu w zmieniających się warunkach i dokonują odpowiednich korekt.

5. **A zatem, w przedsięwzięciu ekonometrycznym nie ma praktycznie etapu, w którym na dłużej jesteśmy przekonani o tym, że dany model jest poprawny i "zweryfikowany".**

Ostatni VII etap to praktyczne wykorzystanie modelu.

Model może być wykorzystany:

- **do analizy** , jak wyglądały w przeszłości prawidłowości ilościowe w tej sferze zjawisk ekonomicznych, które opisuje model,

- **dla celów wnioskowania w przyszłość**, tj. dla celów predykcji (proces wnioskowania w przyszłość na podstawie modelu ekonometrycznego nazywamy predykcją, natomiast konkretny wynik tego procesu nazywać będziemy prognozą).