Rozdział 1: Własności liniowych modeli ekonometrycznych

Zmienną objaśnianą Y i zmienne objaśniające $X_1, X_2, \ldots, X_k$ traktujemy jako zmienne losowe z pewnej populacji F. Zmienna Y nie jest ograniczona, to znaczy, że realizacje Y należą do zbioru liczb rzeczywistych $\mathbb{R}$. Rozpatrujemy rozkład zmiennej Y warunkowy ze względu na $X_1, X_2, \ldots, X_k$, co zapisujemy $F(Y|X_1, X_2, \ldots, X_k)$. Interesuje nas bowiem, w jaki sposób rozkład zmiennej losowej Y zależy od realizacji zmiennych objaśniających. Na przykład w jaki sposób rozkład zysków w pewnej populacji firm zależy od liczby zatrudnionych w tych firmach pracowników, od wartości zainwestowanego kapitału, czy od aktualnej koniunktury gospodarczej.

W praktyce zwykle koncentrujemy uwagę tylko na warunkowej wartości oczekiwanej $E(Y|X_1, X_2, \ldots, X_k)$ zmiennej Y z takiego rozkładu. Na przykład chcemy wyjaśnić przeciętny wzrost gospodarczy w danym regionie pod warunkiem, że znamy zachowanie innych zmiennych makroekonomicznych w tym regionie. Czasami interesuje nas też warunkowa wariancja zmiennej objaśnianej Y, czyli $Var(Y|X_1, X_2, \ldots, X_k)$, kiedy na przykład chcemy wyjaśnić zmienność stóp zwrotu z akcji na giełdzie przy pomocy innych czynników finansowych. Możliwe jest też analizowanie warunkowych kwantyli, ekspektyli i innych charakterystyk rozkładu zmiennej Y warunkowego ze względu na zmienne objaśniające. W tym rozdziale koncentrujemy uwagę na warunkowej wartości oczekiwanej Y, ale w kolejnych rozdziałach powrócimy do analizy rozkładu $F(Y|X_1, X_2, \ldots, X_k)$.

Ekonomistów interesuje odpowiedź na pytanie, jaki wpływ (w domyśle „przeciętnie") pewna zmienna X ma na zmienną Y. Najłatwiej taki wpływ można zmierzyć, kiedy zależność między zmienną X a zmienną Y jest liniowa. Przy założeniu liniowej zależności między dochodem i wydatkami ekonomista może przyjąć, że wzrost aktualnego dochodu o 100 zł powoduje wzrost wydatków przeciętnie na przykład o 40 zł. Do analizy liniowych zależności między zmiennymi wykorzystywany jest liniowy model wartości oczekiwanej postaci:

$$E(Y|X_1, X_2, \ldots, X_k) = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k.$$

Szczególnym i najbardziej popularnym przykładem takiego modelu jest model regresji liniowej:

$$Y = \beta_1 X_1 + \beta_2 X_2 + \ldots + \beta_k X_k + U,$$

w którym oprócz zmiennych objaśniających $X_1, X_2, \ldots, X_k$ z obserwowalnymi wartościami występuje także nieobserwowalny składnik losowy U. Wartość oczekiwana składnika losowego nie zależy od zmiennych objaśniających i równa jest zero, $E(U|X_1, X_2, \ldots, X_k) = 0$.

Wśród zmiennych objaśniających może też występować stała, czyli wyraz wolny. Na przykład można przyjąć, że $X_1 = 1$ i wtedy X_1 pełni rolę stałej w modelu. Stała pomaga z reguły w dopasowaniu modelu do danych rzeczywistych.

Do analizy zależności w regresji liniowej wykorzystywane są wektory $k+1$ zmiennych losowych $(y_i,\ x_{1,i}, x_{2,i},\ \ldots, x_{k,i}\ )$, wybrane losowo z populacji F jako i-te obserwacje zmiennych Y i X_1, X_2, …, X_k, gdzie $i = 1,\ldots, n$. Wektor i-tej obserwacji zmiennych objaśniających można zapisać jako $\boldsymbol{x}_i = [x_{1,i} \quad x_{2,i} \quad \cdots \quad x_{k,i}]$, a i-tą obserwację zmiennej objaśnianej jako y_i. Natomiast cała próba obserwacji składa się z wektora $n \times 1$ obserwacji zmiennej objaśnianej Y, czyli $\boldsymbol{y} = [y_1 \quad y_2 \quad \cdots \quad y_n]'$, oraz z macierzy $n \times k$ obserwacji zmiennych objaśniających $\boldsymbol{X} = [\boldsymbol{x}_1' \quad \boldsymbol{x}_2' \quad \ldots \quad \boldsymbol{x}_n']'$.

Liniową zależność między obserwacjami y_i pewnej zmiennej objaśnianej Y a obserwacjami $x_{1,i}$, $x_{2,i}$, …, $x_{k,i}$ odpowiednich zmiennych objaśniających X_1, X_2, …, X_k można zapisać przy pomocy następującego modelu regresji liniowej:

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + u_i, \tag{1.1}$$

gdzie y_i oznacza i-tą obserwację zmiennej objaśnianej, $x_{1,i}$, $x_{2,i}$, …, $x_{k,i}$ oznaczają i-te obserwacje zmiennych objaśniających, a u_i oznacza i-tą obserwację składnika losowego dla $i = 1, \ldots, n$. Postać tego modelu w formie macierzowej jest następująca:

$$y_i = \boldsymbol{x}_i \boldsymbol{\beta} + u_i \tag{1.2}$$

lub po uwzględnieniu wszystkich obserwacji modelu jednocześnie

$$\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}, \tag{1.3}$$

gdzie wektor $\boldsymbol{u}$ jest n-elementowym wektorem składników losowych.

Zauważmy, że jest to model liniowy względem parametrów i względem zmiennych. Zmienna objaśniana jest liniową funkcją względem zmiennych objaśniających przy ustalonych wartościach parametrów. Zmienna objaśniana jest też liniową funkcją parametrów przy ustalonych wartościach zmiennych objaśniających.

Definicja 1.1

Nieliniowy model regresji definiujemy jako taki, w którym zmienna objaśniana jest nieliniową funkcją parametrów lub nieliniową funkcją zmiennych objaśniających.

W modelu regresji liniowej wartość parametru β_j stojącego przy zmiennej X_j interpretuje się jako wartość potencjalnej zmiany zmiennej objaśnianej Y pod wpływem krańcowo małej (jednostkowej) zmiany wartości X_j przy założeniu, że wszystkie inne zmienne w modelu pozostają bez zmian (warunek ceteris paribus). Jednak w modelach ekonomicznych wpływ krańcowej zmiany zmiennej objaśniającej X_j na zmienną objaśnianą Y nie zawsze jest liniowy. Nie wystarczy wtedy znajomość parametru β_j stojącego przy zmiennej X_j żeby wyliczyć taki wpływ. Jeśli zależność między Y i X_j (i innymi zmiennymi objaśniającymi należącymi do wektora $\boldsymbol{x}$) zadana jest funkcją $Y = f(\boldsymbol{x})$, to przy wyliczaniu oddziaływania krańcowej zmiany X_j na Y pomocne jest wyznaczenie pochodnej cząstkowej $\partial f(\boldsymbol{x})/\partial x_i$, o ile taka pochodna istnieje.

W modelach liniowych łatwiejszy niż w modelach nieliniowych jest wybór specyfikacji modelu, szacowanie parametrów, testowanie hipotez statystycznych, interpretacja ekonomiczna wyników czy też prognozowanie. W kontekście tego opracowania najważniejszy jest fakt, że poznanie podstawowych założeń i własności liniowych modeli regresji jest ważne do lepszego zrozumienia nieliniowych modeli ekonometrycznych. Dlatego w dalszej części tego rozdziału koncentrujemy uwagę na modelach liniowych, a rozszerzenia modeli liniowych do modeli nieliniowych omówione zostaną w następnych rozdziałach.

1.1. Założenia liniowego modelu regresji

Założenia do liniowego modelu regresji przedstawiamy za Hansenem (2018, str. 88-113; por. także Hayashi, 2000, str. 3-34; Greene, 2012, str. 56).

Założenie 1.1: Obserwacje $\{(y_1, \boldsymbol{x}_1), (y_2, \boldsymbol{x}_2), \ldots, (y_n, \boldsymbol{x}_n)\}$ są niezależne i mają jednakowe rozkłady.

To założenie wydaje się racjonalne w przypadku analizy danych przekrojowych, gdzie obserwacje są losowane niezależnie z pewnej populacji. W przypadku danych w postaci szeregów czasowych kolejne obserwacje są rzadko niezależne. Na przykład y_i często zależy od y_{i-1}. Dlatego modele szeregów czasowych wymagają zastosowania nieco innych założeń niż podane tutaj i mają inne własności. Na przykład w miejsce założenia 1.1 przyjmuje się, że $(k+1)$-wymiarowy proces stochastyczny $\{y_i, \boldsymbol{x}_i\}$ jest łącznie stacjonarny i ergodyczny (Hayashi, 2000, str. 109). W tym rozdziale nie rozróżniamy rodzajów danych i dlatego wprowadzamy dość restrykcyjne założenie o niezależności obserwacji w próbie.

Założenie 1.2: Zależności między obserwacjami $(y_i, \boldsymbol{x}_i)$ są objaśniane przez model regresji $y_i = \boldsymbol{x}_i\boldsymbol{\beta} + u_i$, gdzie wartość oczekiwana składnika losowego jest równa zero niezależnie od wartości zmiennych objaśniających $\boldsymbol{x}_i$, $E(u_i|\boldsymbol{x}_i)$=0. Ponieważ obserwacje są niezależne to równoważnie można zapisać założenie, że $E(u_i|\boldsymbol{x}_1, \boldsymbol{x}_2, \dots, \boldsymbol{x}_n) = 0$. Jest to założenie tzw. silnej egzogeniczności (Hayashi, 2000, str. 7).

Czasami w podręcznikach podaje się silniejsze założenie, że składnik losowy jest niezależny od zmiennych objaśniających (por.np. Johnston i DiNardo, 1997, str. 110), ale tak silne założenie nie jest wymagane do zachowania najważniejszych własności modelu regresji i estymatorów takich jak estymator metody najmniejszych kwadratów (MNK, *least squares estimator*), czy też estymator uogólnionej metody najmniejszych kwadratów (UMNK, *generalized least squares estimator*). Inne założenie spotykane w literaturze jest takie, że zmienne objaśniające są nielosowe i dlatego składniki losowe są od nich niezależne (por.np. Johnston i DiNardo, 1997, str. 86). Ponieważ w ekonomii zmienne mają zwykle losowy charakter, to takie założenie wydaje się zbyt restrykcyjne (por. także Hayashi, 2000, str. 13).

Założenie 1.3: Zmienne w modelu regresji mają skończone drugie momenty, czyli $E(y_i^2) < \infty$ oraz $E\|\boldsymbol{x}_i\|^2 < \infty$.

Założenie 1.4: Wartość oczekiwana z iloczynu i-tych obserwacji $\boldsymbol{x}_i$ jest macierzą odwracalną, czyli $E(\boldsymbol{x}_i{}'\boldsymbol{x}_i) > 0$.

Ostatni wzór oznacza, że macierz dodatnio określona jest odwracalna. Ważne jest, żeby rząd macierzy $\boldsymbol{X}$ zawierającej wszystkie obserwacje $\boldsymbol{x}_i$ z próby był równy liczbie zmiennych objaśniających i nie mniejszy od liczby obserwacji w modelu, $r(\boldsymbol{X}) = k \leq n$.

Niektóre własności modeli regresji zależą jeszcze od dwóch dodatkowych założeń.

Założenie 1.5: Składnik losowy jest homoskedastyczny, czyli jego wariancja jest stała i niezależna od $\boldsymbol{x}_i$, czyli $E(u_i^2|\boldsymbol{x}_i) = \sigma^2$. W takim przypadku model regresji nazywamy homoskedastycznym (*homoskedastic regression*).

Założenie 1.5 upraszcza wzory na błędy oszacowań parametrów i dlatego jest szczególnie preferowane przy wyprowadzaniu własności estymatorów i testów statystycznych. Niestety założenie to rzadko jest spełnione w praktyce. Standardem powinno być zatem stosowanie modelu regresji heteroskedastycznej, to znaczy uwzględnianie potencjalnej heteroskedastyczności składnika losowego przy szacowaniu parametrów, liczeniu błędów szacunków, czy też przy testowaniu statystycznym hipotez. Jeśli to założenie nie jest spełnione, to model (i składnik losowy) nazywamy heteroskeastycznym (*heteroskedastic regression, heteroskedastic error term*).

Założenie 1.6: Składnik losowy ma warunkowy rozkład normalny, $(u_i|\boldsymbol{x}_i) \sim N(0, \sigma^2)$. Ponieważ obserwacje są niezależne, to $\boldsymbol{u}|\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$.

Założenie 1.6 jest przydatne, kiedy interesują nas własności estymatorów i statystyk dotyczących modelu regresji w skończonych próbach. Na przykład estymator MNK parametrów $\boldsymbol{\beta}$ w modelu regresji ma (wielowymiarowy) rozkład normalny, kiedy spełnione jest założenie 1.6. Założenie 1.6 umożliwia też szacowanie modelu metodą największej wiarygodności (MNW), ponieważ model jest wtedy w pełni sparametryzowany i znany jest rozkład $\boldsymbol{y}$.

1.2. Własności liniowych modeli regresji i metody najmniejszych kwadratów

Najczęściej wykorzystywaną metodą szacowania parametrów modeli regresji liniowej jest metoda najmniejszych kwadratów (MNK). Polega ona na znalezieniu takich wartości estymatora $\widehat{\boldsymbol{\beta}}$ parametrów $\boldsymbol{\beta}$, które minimalizują sumę kwadratów reszt modelu regresji:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \mathbb{R}^k}{\mathbf{argmin}}\, S(\boldsymbol{\beta}) = \underset{\boldsymbol{\beta} \in \mathbb{R}^k}{\mathbf{argmin}} \left(\sum_{i=1}^{n} (y_i - \boldsymbol{x}_i \boldsymbol{\beta})^2 \right). \quad (1.4)$$

Estymator MNK można wyprowadzić przyrównując pochodną z funkcji $S(\boldsymbol{\beta})$ po parametrach $\boldsymbol{\beta}$ do wektora zerowego:

$$\frac{\partial S(\widehat{\boldsymbol{\beta}})}{\partial \boldsymbol{\beta}} = -2 \sum_{i=1}^{n} y_i \boldsymbol{x}_i + \sum_{i=1}^{n} \boldsymbol{x}_i{}' \boldsymbol{x}_i \widehat{\boldsymbol{\beta}} = \boldsymbol{0}. \quad (1.5)$$

Estymator MNK dany jest zatem wzorem:

$$\widehat{\boldsymbol{\beta}} = \left(\sum_{i=1}^{n} \boldsymbol{x}_i{}' \boldsymbol{x}_i \right)^{-1} \left(\sum_{i=1}^{n} y_i \boldsymbol{x}_i \right) = (\boldsymbol{X}'\boldsymbol{X})^{-1} \boldsymbol{X}'\boldsymbol{y}. \quad (1.6)$$

Estymator MNK parametrów $\boldsymbol{\beta}$ przy spełnieniu założeń 1.1 – 1.4 ma następujące własności.

Przykład 1.1

Zdefiniujmy w programie MATLAB macierze obserwacji zmiennych objaśniających i zmiennej objaśnianej.

```
X = [ones(100,1) randn(100,2)]; % stała i dwie zmienne generowane z N(0,1)
y = randn(100,1);               % niezależnie generowany wektor z N(0,1)
```

Rząd macierzy możemy sprawdzić poleceniem `rank`.

```
r = rank(X);                    % rząd macierzy X
```

W programie MATLAB można oszacować parametry modelu regresji metodą MNK na kilka sposobów.

```
beta1 = inv(X'*X)*(X'*y)        % oryginalny wzór
beta2 = (X'*X)\(X'*y)           % lepszy sposób
beta3 = X\y                     % najlepszy sposób
```

Trzeci sposób jest rekomendowany przez autorów oprogramowania.

Własność 1.1: Przy spełnionych założeniach 1.1 – 1.4 estymator MNK jest nieobciążony, tzn. jego wartość oczekiwana równa jest wartości $\boldsymbol{\beta}$, $E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$. Prawdziwa jest również własność $E(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \boldsymbol{\beta}$.

Przydatne będzie przedstawienie estymatora MNK jako funkcji zależnej od wektora prawdziwych wartości parametrów i odchyleń od tego wektora. Ze wzorów (1.3) i (1.6) wynika:

$$\widehat{\boldsymbol{\beta}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u} \tag{1.7}$$

Ponieważ $E(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = E((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{y}|\boldsymbol{X}) = E((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'(\boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u})|\boldsymbol{X})$, to zachodzi dalej $E((\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{X}\boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{\beta} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'E(\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{\beta}$. Z prawa iterowanych oczekiwań wynika, że $E(\widehat{\boldsymbol{\beta}}) = E\left(E(\widehat{\boldsymbol{\beta}}|\boldsymbol{X})\right) = E(\boldsymbol{\beta}) = \boldsymbol{\beta}$.

Własność 1.2: Przy spełnionych założeniach 1.1 – 1.5 wariancja estymatora MNK warunkowa ze względu na $\boldsymbol{X}$ ma postać $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = Var(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$.

Na podstawie wzoru (1.7) można wyznaczyć wariancję oszacowań parametrów:

$$\begin{aligned} \boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} &= Var(\widehat{\boldsymbol{\beta}}|\boldsymbol{X}) = E[(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})'|\boldsymbol{X}] \\ &= E[(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{u}\boldsymbol{u}'\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}|\boldsymbol{X}] = \\ &= (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1} = \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}. \end{aligned} \tag{1.8}$$

Ostatnia równość jest prawdziwa, ponieważ macierz wariancji składników losowych w modelu homoskedastycznym ma postać $\boldsymbol{D} = E[\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X}] = \sigma^2\boldsymbol{I}$ i ma wymiary $n \times n$. Macierz $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ ma wymiary $k \times k$. Ponieważ wariancja składnika losowego σ^2 zwykle nie jest znana, to zamiast niej podstawia się do wzoru (1.8) jej nieobciążony i zgodny szacunek $\hat{\sigma}^2 = (\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}})/(n-k)$, gdzie $\widehat{\boldsymbol{u}}$ oznacza wektor reszt regresji oszacowanej metodą MNK. Wtedy estymator wariancji oszacowań parametrów ma wzór:

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}} = \hat{\sigma}^2(\boldsymbol{X}'\boldsymbol{X})^{-1} = \frac{\widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}}{n-k}(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{1.9}$$

Ponadto z twierdzenia Gaussa-Markowa wynika, że estymator MNK parametrów $\boldsymbol{\beta}$ w modelu homoskedastycznym (czyli przy spełnieniu założeń 1.1 – 1.5) ma następującą własność.

Własność 1.3: (twierdzenie Gaussa-Markowa) Przy spełnionych założeniach 1.1 – 1.5 estymator MNK jest efektywny w klasie liniowych nieobciążonych estymatorów (BLUES, *best linear unbiased estimator*). To znaczy, że dla każdego liniowego nieobciążonego estymatora $\widetilde{\boldsymbol{\beta}}$ zachodzi $Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) \geq \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$, czyli różnica $Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) - \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1}$ jest macierzą dodatnio półokreśloną.

Ta własność oznacza, że estymator MNK jest najbardziej precyzyjnym estymatorem w klasie estymatorów liniowych, czyli tych będących liniową funkcją $\boldsymbol{y}$, i nieobciążonych. Dowód twierdzenia można znaleźć na przykład w podręcznikach Hayashiego (2000, str. 29) i Hansena (2018, str. 94-95).

Przykład 1.2

Wykorzystując dane z przykładu 1.1 oszacujmy wariancję składnika losowego modelu regresji.

```
beta = X\y;                     % szacunek parametrów modelu regresji
u = y - X*beta;                 % obliczanie reszt z modelu regresji
[n k] = size(X)                 % liczba obserwacji n i liczba parametrów k
sigma2 = u'*u/(n-k)             % szacunek wariancji składnika losowego
```

Następnie możemy policzyć szacunek wariancji oszacowań parametrów.

```
V = sigma2*inv(X'*X)           % macierz wariancji oszacowań parametrów
```

Własność 1.4: Jeśli spełnione jest dodatkowo założenie 1.6 i wektor składników losowych ma n-wymiarowy rozkład normalny $\boldsymbol{u}|\boldsymbol{X} \sim N(\boldsymbol{0}, \sigma^2\boldsymbol{I})$, to rozkład $\widehat{\boldsymbol{\beta}}$ warunkowy ze względu na $\boldsymbol{X}$ jest także normalny, czyli:

$$\widehat{\boldsymbol{\beta}}|\boldsymbol{X} \sim N\left(\boldsymbol{\beta}, \boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\right).$$

Dla każdego j-tego elementu wektora $\widehat{\boldsymbol{\beta}}$, gdzie $j = 1, \ldots, k$, mamy $\hat{\beta}_j \sim N\left(\beta_j, \sigma^2(\hat{\beta}_j)\right)$, gdzie $\sigma^2(\hat{\beta}_j)$ jest j-tym parametrem leżącym na diagonalnej macierzy $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$.

Jeśli znana jest parametru $\beta_j = \beta^*$ i znana jest wariancja $\sigma^2(\hat{\beta}_j)$ szacunku β_j (i jej pierwiastek, czyli odchylenie standardowe $\sigma(\hat{\beta}_j)$), to statystyka $z = (\hat{\beta}_j - \beta_j)/\sigma(\hat{\beta}_j)$ ma standardowy rozkład normalny, $N(0,1)$. Ta własność wykorzystywana jest często do testowania istotności zmiennych stojących przy poszczególnych parametrach w modelu regresji (1) lub do testowania hipotez zakładających konkretną wartość danego parametru, to znaczy $\beta_j = \beta^*$. Bezpośrednie użycie statystyki z w praktyce nie jest możliwe, gdy nie jest znana wariancja składnika losowego ani tym bardziej nie jest znane odchylenie standardowe $\sigma(\hat{\beta}_j)$ szacunku parametru β_j. Wykorzystuje się wtedy szacunek wariancji oszacowań parametrów $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}$ do przybliżenia $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$. Szacunek odchylenia standardowego estymatora $\hat{\beta}_i$ ma wtedy wzór $\hat{\sigma}(\hat{\beta}_j) = \sqrt{v_{jj}}$, gdzie v_{jj} oznacza j-ty element diagonalnej macierzy $\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}$. Obliczana jest wtedy analogiczna do z statystyka t:

$$t = (\hat{\beta}_j - \beta^*)/\hat{\sigma}(\hat{\beta}_j)$$

i przy założeniu, że $\beta_j = \beta^*$, ma ona rozkład $t(n-k)$. Dla dużych prób, gdzie $n - k > 30$ rozkład t jest bardzo podobny do standardowego rozkładu normalnego.

Przykład 1.3

Statystyka t jest często wykorzystywana do sprawdzenia, czy zmienna objaśniająca jest statystycznie istotna w danym modelu (to znaczy, czy parametr przy tej zmiennej jest statystycznie istotnie różny od zera). Wylicza się wtedy statystykę postaci $t = \hat{\beta}_j/\hat{\sigma}(\hat{\beta}_j)$ i porównuje ją z odpowiednią wartością krytyczną, np. dla przyjętego poziomu istotności 0.05 taką wartością krytyczną będzie 1,96.

Obliczenia błędów szacunku i statystyk t dla każdej zmiennej można dokonać w programie MATLAB w następujący sposób. Wykorzystano tutaj dane i wcześniejsze obliczenia z przykładów 1.1 i 1.2.

```
s_beta = sqrt(diag(V));         % błędy standardowe oszacowań parametrów
t = beta./s_beta                % obliczone statystyki t
(abs(t) > 1.96)                 % czy zmienna istotna? 1=prawda, 0=fałsz
```

W modelu, w którym założenia 1.5 i 1.6 nie są spełnione, macierz wariancji oszacowań parametrów $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} = (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}\boldsymbol{X}(\boldsymbol{X}'\boldsymbol{X})^{-1}$ nie daje się łatwo skrócić tak jak we wzorze (1.8). Macierz $\boldsymbol{D}$ można wtedy zapisać w następujący sposób:

$$\boldsymbol{D} = diag(\sigma_1^2, \sigma_n^2, \dots, \sigma_n^2) = \begin{bmatrix} \sigma_1^2 & 0 & \dots & 0 \\ 0 & \sigma_2^2 & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & \sigma_n^2 \end{bmatrix}.$$

Wartości σ_i^2 z reguły nie są znane i macierze $\boldsymbol{D}$ i $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ mogą zostać jedynie oszacowane. W praktyce często stosowany jest estymator White'a wariancji oszacowań parametrów $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ (White, 1980):

$$\widehat{\boldsymbol{V}}_{\widehat{\boldsymbol{\beta}}}^* = \frac{n}{n-k}(\boldsymbol{X}'\boldsymbol{X})^{-1}\left(\sum_{i=1}^{n} \hat{u}_i^2 \boldsymbol{x}_i{}'\boldsymbol{x}_i\right)(\boldsymbol{X}'\boldsymbol{X})^{-1}. \tag{1.10}$$

Ten oraz inne alternatywne estymatory wariancji szacunku parametrów, odporne na heteroskedastyczność składnika losowego, zostały dokładniej opisane między innymi w podręczniku Hansena (2018, str.101-104).

W modelu heteroskedastycznym przestaje też działać twierdzenie Gaussa-Markowa i prawdziwa jest następująca bardziej ogólna własność.

Własność 1.5: Jeśli w modelu regresji $\boldsymbol{y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{u}$ składnik losowy spełnia warunki $E(\boldsymbol{u}|\boldsymbol{X}) = \boldsymbol{0}$ i $E[\boldsymbol{u}\boldsymbol{u}'|\boldsymbol{X}] = \boldsymbol{D}$, to dla każdego liniowego nieobciążonego estymatora $\widetilde{\boldsymbol{\beta}}$ zachodzi $Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) \geq (\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}$, czyli różnica $Var(\widetilde{\boldsymbol{\beta}}|\boldsymbol{X}) - (\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}$ jest macierzą dodatnio półokreśloną.

Oznacza to, że w modelu heteroskedastycznym (lub w modelu ze wzajemnie skorelowanymi składnikami losowymi) estymator MNK nie jest najbardziej precyzyjny, czyli efektywny.

Efektywny jest następujący estymator uogólnionej metody najmniejszych kwadratów (UMNK):

$$\widehat{\boldsymbol{\beta}}_{UNMK} = (\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{X})^{-1}\boldsymbol{X}'\boldsymbol{D}^{-1}\boldsymbol{y}. \tag{1.11}$$

W praktyce jednak ze względu na trudności z ustaleniem wartości $\boldsymbol{D}$ stosuje się estymator MNK, a jedynie wariancję i błędy estymatora liczy się uwzględniając heteroskedastyczność składnika losowego. Romano i Wolf (2017) proponują z kolei zastosowanie ważonej MNK (WMNK) w sytuacji, gdy testy statystyczne potwierdzą występowanie heteroskedastyczności składnika losowego ze względu na większą precyzję oszacowań otrzymanych przy użyciu WMNK.

Przykład 1.4

Szacunku błędów standardowych oszacowań parametrów metodą White'a można dokonać w programie MATLAB w następujący sposób. Wykorzystano tutaj obliczenia dokonane wcześniej w przykładach od 1.1 do 1.3.

```
D = diag(u.^2);                  % macierz diagonalna z kwadratami reszt
Vhet = inv(X'*X)*(X'*D*X)*inv(X'*X);  % macierz wariancji White'a
V_White = Vhet*(n/(n-k));        % poprawienie liczby stopni swobody
s_White = sqrt(diag(V_White))    % odporne błędy oszacowań parametrów
```

Rozkład estymatora $\widehat{\boldsymbol{\beta}}$ w skończonych próbach zwykle nie jest znany, w szczególności kiedy nie jest spełnione założenie 1.6. Dlatego przydatna jest informacja o asymptotycznym rozkładzie estymatora $\widehat{\boldsymbol{\beta}}$, to znaczy wtedy gdy $n \to \infty$. Liczymy na to, że asymptotyczny rozkład estymatora będzie dobrym przybliżeniem rozkładu $\widehat{\boldsymbol{\beta}}$ w skończonej próbie. W tym kontekście ważnymi cechami estymatora MNK są jego zgodność i asymptotyczna normalność.

Własność 1.6: (zgodność estymatora MNK) Jeżeli spełnione są warunki: (1) obserwacje $(y_i, \boldsymbol{x}_i)$, $i = 1,\ldots,n$, mają identyczne niezależne rozkłady, (2) $E(y^2) < \infty$, (3) $E\|\boldsymbol{x}\|^2 < \infty$, (4) macierz $\boldsymbol{Q}_{\boldsymbol{xx}} = E(\boldsymbol{xx}')$ jest dodatnio określona, to estymator MNK jest **zgodny**, czyli $\lim_{n\to\infty} P\left(\left|\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}\right| \le \boldsymbol{\delta}\right) = 1$ dla każdego $\boldsymbol{\delta} > \boldsymbol{0}$, co w skrócie zapisujemy jako $\widehat{\boldsymbol{\beta}}_n \xrightarrow{p} \boldsymbol{\beta}$ przy $n \to \infty$.

Własność 1.7: (asymptotyczna normalność estymatora MNK) Jeżeli spełnione są warunki: (1) obserwacje $(y_i, \boldsymbol{x}_i)$, $i = 1,\ldots,n$, mają identyczne niezależne rozkłady, (2) $E(y^4) < \infty$, (3) $E\|\boldsymbol{x}\|^4 < \infty$, (4) macierz $\boldsymbol{Q}_{xx} = E(\boldsymbol{x}\boldsymbol{x}')$ jest dodatnio określona, to dla $n \to \infty$

$$\sqrt{n}(\widehat{\boldsymbol{\beta}}_n - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \boldsymbol{V}_{\boldsymbol{\beta}}),$$

gdzie $\boldsymbol{V}_{\boldsymbol{\beta}} = \boldsymbol{Q}_{xx}^{-1}\boldsymbol{\Omega}\boldsymbol{Q}_{xx}^{-1}$ oraz $\boldsymbol{\Omega} = E(x_i x_i' e_i^2)$. Symbol $\xrightarrow{d}$ oznacza zbieżność z dystrybuantą (słabą zbieżność).

Macierz $\boldsymbol{V}_{\boldsymbol{\beta}}$ jest nazywana asymptotyczną wariancją $\widehat{\boldsymbol{\beta}}$. Przy $n \to \infty$ wariancja oszacowań $\widehat{\boldsymbol{\beta}}$ z próby, $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$, pomnożona przez n dąży z prawdopodobieństwem do $\boldsymbol{V}_{\boldsymbol{\beta}}$, czyli $n\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}} \xrightarrow{p} \boldsymbol{V}_{\boldsymbol{\beta}}$.

1.3. Nakładanie restrykcji liniowych na parametry modelu regresji liniowej

W modelach ekonometrycznych nakłada się czasami warunki na parametry, żeby móc następnie sprawdzić przy pomocy testów statystycznych, czy te warunki są spełnione, lub wykorzystać nałożone warunki do dokładniejszego oszacowania parametrów modelu.

Przykład 1.5

Chcemy sprawdzić, czy w modelu regresji $y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_k x_{k,i} + u_i$ z k zmiennymi objaśniającymi spełniony jest warunek $\beta_1 = \beta^*$. Hipoteza zerowa zakłada, że $\beta_1 = \beta^*$ (na przykład $\beta_1 = 0$**)**, a hipoteza alternatywna zakłada, że $\beta_1 \neq \beta^*$. Statystyka testowa służąca do weryfikacji hipotezy zerowej ma następującą postać (analogiczną do wzoru na statystykę z omawianą wcześniej):

$$t_1 = \frac{\hat{\beta}_1 - \beta^*}{\hat{\sigma}(\hat{\beta}_1)}$$

i przy założeniu prawdziwości hipotezy zerowej oraz spełnionych założeń 1.1 – 1.4 i 1.6 ma ona rozkład $t - Studenta$ z $n - k$ stopniami swobody, $t_1 \sim t(n-k)$. Warto wiedzieć, że przy założeniu prawdziwości hipotezy zerowej oraz spełnieniu założeń 1.1 – 1.4 ta sama statystyka ma asymptotyczny (to znaczy przy $n \to \infty$) standardowy rozkład normalny.

Przykład 1.6

Często nakłada się na parametry liniowego modelu regresji restrykcje liniowe postaci $\boldsymbol{R\beta} = \boldsymbol{r}$. Na przykład restrykcje zerowe na grupę trzech parametrów z pięciu w modelu regresji można zapisać następująco: $\beta_1 = 0$, $\beta_2 = 0$ i $\beta_3 = 0$ lub stosując zapis macierzowy: $\begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \end{bmatrix} \begin{bmatrix} \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_1 \\ \beta_1 \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \end{bmatrix}$.

Przykład 1.7

Można też nakładać bardziej złożone warunki na liniowe funkcje parametrów w tym samym modelu regresji, np. $2 \cdot \beta_1 - 3 \cdot \beta_2 = 1$, co przedstawia się stosując zapis macierzowy jako $[2 \quad -3 \quad 0 \quad 0 \quad 0] \cdot [\beta_1 \quad \beta_2 \quad \beta_3 \quad \beta_4 \quad \beta_5]' = 1$.

Przykład 1.8

Załóżmy, że chcemy przetestować jednocześnie dwa niezależne warunki ($m = 2$), $\beta_1 + \beta_3 = 5$ oraz $\beta_2 = \beta_4$, nałożone na parametry w następującym modelu regresji:

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{3,i} + \beta_4 x_{4,i} + u_i$$

Liczba wszystkich parametrów β_j wynosi zatem 4. Odpowiednie macierze $\boldsymbol{R}$ i $\boldsymbol{r}$ służące do zapisania warunków w formie macierzowej $\boldsymbol{R\beta} = \boldsymbol{r}$ mają postać:

$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & -1 \end{bmatrix}$ oraz $\boldsymbol{r} = \begin{bmatrix} 5 \\ 0 \end{bmatrix}$.

Przykład 1.9

Przy założeniu stałych korzyści skali w funkcji produkcji Cobba-Douglasa suma elastyczności produkcji względem czynników produkcji jest równa 1, co można zapisać następująco: $Y = \exp(\alpha_0) \cdot X_1^{\alpha_1} \cdot X_2^{\alpha_2} \cdot X_3^{\alpha_3} \cdot U$, gdzie U to czynnik losowy, a $\alpha_1 + \alpha_2 + \alpha_3 = 1$. Taki model można przekształcić do postaci liniowej względem parametrów redefiniując zmienne: $y = \ln(Y)$, $x_1 = \ln(X_1)$, $x_2 = \ln(X_2)$, $x_1 = \ln(X_3)$, $u = \ln(U)$. Model regresji służący do testowania stałych korzyści skali będzie miał wzór:

$$y = \alpha_0 + \alpha_1 x_1 + \alpha_1 x_2 + \alpha_1 x_3 + u_t$$

a restrykcje zapisane w macierzach $\boldsymbol{R}$ i $\boldsymbol{r}$ mają wtedy postać $\boldsymbol{R} = [0 \quad 1 \quad 1 \quad 1]$, $\boldsymbol{r} = [1]$.

Przykład 1.10

Przypuśćmy, że chcemy testować tak zwane prawo jednej ceny, które mówi, że cena towaru na jednym rynku powinna być równa cenie towaru na drugim rynku po uwzględnieniu kursu walutowego między oboma rynkami. Zachodzi wtedy zależność $S = \frac{P}{P^*}$, gdzie S oznacza kurs walutowy między oboma rynkami (np. kurs EUR/PLN równy około 4,2, czyli 4,2 złote za 1 euro), P oczacza cenę towaru na jednym rynku (np. w walucie PLN), a P^* oznacza cenę towaru na drugim rynku (np. w walucie EUR). Do zbadania prawdziwości prawa jednej ceny można zbudować model ekonometryczny i wykorzystać dane o cenach i kursach walutowych w postaci szeregów czasowych. Wzór $S = \frac{P}{P^*}$ można zlinearyzować poprzez obustronne zlogarytmowanie przy założeniu, że ceny są dodatnie, $s = p - p^*$. Model ekonometryczny służący do ustalenia faktycznej zależności między zmiennymi może mieć wtedy postać:

$$s_t = \beta_0 + \beta_1 \cdot p_t + \beta_2 \cdot p_t^* + \varepsilon_t.$$

Restrykcje na parametry powinny być wtedy następujące $\beta_0 = 0$, $\beta_1 = -\beta_2 = 1$. Odpowiednie macierze $\boldsymbol{R}$ i $\boldsymbol{r}$ mają wzory $\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix}$, $\boldsymbol{r} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}$.

Przykład 1.11

Zgodnie z teorią nieubezpieczonego parytetu stóp procentowych oczekiwane dynamika kursu walutowego powinna być równa różnicy stóp procentowych w kraju i za granicą, $E_t(\ln(S_{t+1}) - \ln(S_t)) \equiv E_t(\Delta s_{t+1}) = i_t - i_t^*$. Taką teorię można testować przy pomocy modelu regresji

$$\Delta s_{t+1} = \alpha_0 + \alpha_1 i_t + \alpha_2 i_t^* + u_t$$

i nałożonych na nią restrykcji $\alpha_0 = 0$, $\alpha_1 = 1$, $\alpha_2 = -1$. Odpowiednie macierze $\boldsymbol{R}$ i $\boldsymbol{r}$ mają wtedy postać:

$$\boldsymbol{R} = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & -1 \end{bmatrix}, \boldsymbol{r} = \begin{bmatrix} 0 \\ 1 \\ -1 \end{bmatrix}.$$

Przykład 1.12

Na rynku giełdowym, na którym inwestują jedynie gracze neutralni wobec ryzyka kurs futures pewnego instrumentu finansowego (np. akcji) powinien w sposób nieobciążony prognozować przyszły kurs spot tego instrumentu. Taką własność można analizować przy użyciu odpowiedniego modelu ekonometrycznego oraz danych w postaci szeregów czasowych

$$s_{t+h} = \beta_0 + \beta_1 \cdot f_t + \varepsilon_t.$$

s_{t+h} oznacza tutaj kurs spot (lub jego logarytm naturalny) w okresie $t + h$, f_t oznacza kurs futures obserwowany w okresie t. Restrykcje nakładane na parametry powinny mieć wtedy postać $\beta_0 = 0$, $\beta_1 = 1$. Odpowiednie macierze $\boldsymbol{R}$ i $\boldsymbol{r}$ mają wzory $\boldsymbol{R} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix}$, $\boldsymbol{r} = \begin{bmatrix} 0 \\ 1 \end{bmatrix}$.

Z własności 1.4 można także wyprowadzić statystyki przydatne do testowania bardziej złożonych hipotez, dotyczących jednocześnie wielu parametrów i wielu warunków. Pomożenie estymatora $\widehat{\boldsymbol{\beta}}$ przez pewną macierz znanych parametrów $\boldsymbol{R}$ o rozmiarach $m \times K$ powoduje, że iloczyn $\boldsymbol{R}\widehat{\boldsymbol{\beta}}$ ma rozkład (dla uproszczenia zapisu opuszczamy warunkowanie po $\boldsymbol{X}$):

$$\boldsymbol{R}\widehat{\boldsymbol{\beta}} \sim N\left(\boldsymbol{R}\boldsymbol{\beta}, \boldsymbol{R}\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\right)$$

Z kolei odjęcie od tego iloczynu wektora znanych parametrów $\boldsymbol{r}$ o rozmiarach $m \times 1$ zmienia rozkład wyrażenia w następujący sposób:

$$\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r} \sim N\left(\boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r}, \boldsymbol{R}\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\right)$$

Wykorzystuje się fakt, że dla każdego wektora $\boldsymbol{z} \sim N(\boldsymbol{0}, \boldsymbol{\Omega})$ o wymiarach $m \times 1$, gdzie macierz wariancji $\boldsymbol{\Omega}$ jest nieosobliwa, forma kwadratowa $\boldsymbol{z}'\,\boldsymbol{\Omega}^{-\boldsymbol{1}}\boldsymbol{z}$ ma rozkład χ^2 z m stopniami swobody. Dlatego przy założeniu hipotezy, że $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$ (czyli $\boldsymbol{R}\boldsymbol{\beta} - \boldsymbol{r} = \boldsymbol{0}$), następująca forma kwadratowa ma także rozkład χ^2 z m stopniami swobody:

$$\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\left(\boldsymbol{R}\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right) \sim \chi^2(m)$$

W analogiczny sposób budowana jest statystyka F, służąca do testowania hipotezy zerowej H_0: $\boldsymbol{R}\boldsymbol{\beta} = \boldsymbol{r}$. Hipoteza alternatywna zakłada, że równość nie jest spełniona, H_1: $\boldsymbol{R}\boldsymbol{\beta} \neq \boldsymbol{r}$. Ponieważ wartość $\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}$ nie jest znana, to należy ją oszacować, na przykład stosując wzór (1.9) lub (1.10). Wtedy następująca statystyka testowa F ma rozkład $Fishera - Snedecora$ (zwanego też rozkładem F):

$$F = \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'\left(\boldsymbol{R}\boldsymbol{V}_{\widehat{\boldsymbol{\beta}}}\boldsymbol{R}'\right)^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)/m \sim F(m, n-k). \tag{1.12}$$

We wzorze tym m oznacza liczbę niezależnych warunków, czyli wierszy wektora $\boldsymbol{r}$.

Przy założeniu modelu homoskedastycznego wykorzystuje się następującą własność wariancji składnika losowego $\frac{\boldsymbol{u}'\boldsymbol{u}}{\sigma^2} \sim \chi^2(n-k)$ oraz własność, że iloraz dwóch niezależnych zmiennych losowych z rozkładów $\chi^2(m)$ i $\chi^2(n-k)$ pomnożony przez $\frac{n-k}{m}$ ma rozkład $F(m, n-k)$. Wtedy prawdziwy jest wzór:

$$\frac{\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)'(\sigma^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)}{\frac{\boldsymbol{u}'\boldsymbol{u}}{\sigma^2}}\cdot\frac{n-k}{m}$$

$$=\frac{\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)'(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)/m}{\frac{\boldsymbol{u}'\boldsymbol{u}}{n-k}}$$

który po podstawieniu $\hat{\sigma}^2$ prowadzi do wzoru statystyki testowej F:

$$F=\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)'(\hat{\sigma}^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)/m \sim F(m,n-k). \tag{1.13}$$

Warto dodać, że bardzo podobny wzór ma statystyka Walda, służąca do testowania tej samej hipotezy zerowej ($\boldsymbol{R\beta}=\boldsymbol{r}$). Przy spełnionych założeniach 1.1 – 1.4 oraz spełnionym warunku $\boldsymbol{R\beta}=\boldsymbol{r}$ statystyka Walda dla modelu liniowego ma wzór:

$$W=\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)'(\hat{\sigma}^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}}-\boldsymbol{r}\right)=m\cdot F\xrightarrow{d}\chi^2(m). \tag{1.14}$$

Ważne jest, że statystyka Walda ma asymptotyczny (to znaczy przy $n\to\infty$) rozkład χ^2 nawet, gdy składnik losowy nie ma rozkładu normalnego. Statystyka Walda zostanie omówiona dokładniej w rozdziale 2.

Przykład 1.13

Policzmy w programie MATLAB przykładowe statystyki F i W służące do testowania restrykcji nałożonych na modele regresji z przykładu 1.8. Chcemy przetestować jednocześnie dwa niezależne warunki ($m=2$), $\beta_1+\beta_3=5$ oraz $\beta_2=\beta_4$, nałożone na parametry w następującym modelu regresji $y_i=\beta_1x_{1,i}+\beta_2x_{2,i}+\beta_3x_{3,i}+\beta_4x_{4,i}+u_i$.

Zdefiniujmy macierze obserwacji zmiennych objaśniających i zmiennej objaśnianej.

```
X = randn(100,4);               % cztery zmienne generowane z N(0,1)
y = randn(100,1);               % niezależnie generowany wektor z N(0,1)
```

Zapiszmy restrykcje nałożone na parametry w macierzach $\boldsymbol{R}$ i $\boldsymbol{r}$.

```
R = [ 1 0 1 0; 0 1 0 -1];       % macierze restrykcji na parametry
r = [5; 0];
m = size(r,1);                  % liczba niezależnych restrykcji
```

Oszacujmy model i policzmy statystyki F i W.

```
beta = X\y;                     % szacunek parametrów modelu regresji
u = y - X*beta;                 % obliczanie reszt z modelu regresji
[n k] = size(X);                % liczba obserwacji n i liczba parametrów k
```

```
sigma2 = u'*u/(n-k);                % szacunek wariancji składnika losowego
V = sigma2*inv(X'*X);               % macierz wariancji oszacowań parametrów
F = (R*beta-r)'*inv(R*V*R')*(R*beta-r)/m    % statystyka F
W = (R*beta-r)'*inv(R*V*R')*(R*beta-r)      % statystyka Walda
```

Warto zwrócić uwagę, że obserwacje zmiennej y w tym przykładzie zostały wygenerowane niezależnie ze standardowego rozkładu normalnego i nie zależą od zmiennych objaśniających w modelu. Z tego powodu oszacowania parametrów będą niedokładne. Dlatego niskie wartości statystyk F i W mogą wskazywać na brak podstaw do odrzucenia hipotezy zerowej o prawdziwości założonych restrykcji. Nie oznacza to, że restrykcje te są spełnione, ale raczej że dane nie pozwalają zbyt dokładnie ocenić, czy hipoteza jest spełniona czy nie. Wartości p ($p - value$, tzw. empiryczne poziomy istotności) obu statystyk można policzyć w następujący sposób.

```
p_F = 1-fcdf(F,m,n-k)               % p-value statystyki F
p_W = 1-chi2cdf(W,m)                % p-value statystyki W
```

Stosowanie funkcji `fcdf` i `chi2cdf` wymaga zainstalowania pakietu Statistics and Machine Learning Toolbox w programie MATLAB.

Tabela 1.1. Własności estymatora MNK w modelach regresji

Założenia modelu	Oszacowania $\boldsymbol{\beta}$	Statystyka t_l	Statystyka F
$\boldsymbol{X}$ nielosowe, $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$	$\widehat{\boldsymbol{\beta}} \sim N(\boldsymbol{\beta}, \sigma^2 (\boldsymbol{X}'\boldsymbol{X})^{-1})$	$t_l \sim t(n-k)$	$F \sim F(m, n-k)$ $m \cdot F = W \xrightarrow{d} \chi^2(m)$
$\boldsymbol{X}$ losowe, ale niezależne od $\boldsymbol{u}$, $\boldsymbol{u} \sim N(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$	$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, rozkład niegaussowski (*)	$t_l \sim t(n-k)$	$F \sim F(m, n-k)$ $m \cdot F = W \xrightarrow{d} \chi^2(m)$
$\boldsymbol{X}$ losowe, ale niezależne od $\boldsymbol{u}$, $(\sum_{i=1}^{n} \boldsymbol{x}_i \boldsymbol{x}_i')/N \xrightarrow{p} \boldsymbol{Q}$ $\boldsymbol{u} \sim niegaussowski(\boldsymbol{0}, \sigma^2 \boldsymbol{I})$	$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, $\sqrt{N}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q}^{-1})$	$t_l \xrightarrow{d} N(0,1)$	$m \cdot F = W \xrightarrow{d} \chi^2(m)$
Model autoregresji ze stacjonarnymi zmiennymi	$E(\widehat{\boldsymbol{\beta}}) = \boldsymbol{\beta}$, $\sqrt{N}(\widehat{\boldsymbol{\beta}}_N - \boldsymbol{\beta}) \xrightarrow{d} N(\boldsymbol{0}, \sigma^2 \boldsymbol{Q}^{-1})$	$t_l \xrightarrow{d} N(0,1)$	$m \cdot F = W \xrightarrow{d} \chi^2(m)$

Źródło: opracowanie na podstawie pracy Hamiltona (1994), Tabela 8.1, str. 209.

Uwagi: $\boldsymbol{Q}$ oznacza $E(\boldsymbol{x}_i\boldsymbol{x}_i{}')$. Symbol $\xrightarrow{d}$ oznacza zbieżność z dystrybuantą (słabą zbieżność), a symbol $\xrightarrow{p}$ oznacza zbieżność z prawdopodobieństwem. (*) rozkład $\widehat{\boldsymbol{\beta}}|\boldsymbol{X}$ jest jednak normalny, czyli $\widehat{\boldsymbol{\beta}}|\boldsymbol{X} \sim N(\boldsymbol{\beta}, \sigma^2(\boldsymbol{X}'\boldsymbol{X})^{-1})$.

W Tabeli 1.1 przedstawiono najważniejsze własności estymatorów MNK i podstawowych statystyk t i F dla różnych rodzajów modeli. Warto zwrócić uwagę na trzeci wiersz, który przedstawia najczęściej spotykaną sytuację, że obserwacje $\boldsymbol{X}$ są losowe, obserwacje składnika losowego nie mają rozkładu normalnego. W takiej sytuacji korzysta się przede wszystkim z własności asymptotycznych estymatorów i statystyk lub wykorzystuje metody symulacyjne (np. Monte-Carlo lub bootstrap) do przybliżania interesujących nas rozkładów.

Przykład 1.14

Dla tej samej hipotezy zerowej, co w przykładzie 1.5 ($\beta_i = \beta^*$), można skonstruować odpowiednią statystykę testową F. Wzór (1.13) znacznie się uprości do postaci:

$$F = \left(\hat{\beta}_i - \beta^*\right)^2 / \hat{\sigma}^2\left(\hat{\beta}_i\right) \sim F(1, N-K)$$

Zwróćmy uwagę, że $t_i = \sqrt{F}$. Z własności rozkładu F wynika, że pierwiastek zmiennej z tego rozkładu z parametrami 1 i $n-k$ ma rozkład $t(n-k)$.

1.4. Metoda najmniejszych kwadratów przy warunkach pobocznych

W poprzednim punkcie omówiono testowanie restrykcji liniowych postaci $\boldsymbol{R\beta} = \boldsymbol{r}$, nałożonych na parametry modelu regresji liniowej. Jeżeli wyniki odpowiedniego testu statystycznego wskazują, że nie ma podstaw do odrzucenia tych restrykcji, to sensowne będzie oszacowanie parametrów modelu regresji przy jednoczesnym nałożeniu warunku $\boldsymbol{R\beta} = \boldsymbol{r}$ na szacowane parametry. Taki szacunek będzie dokładniejszy niż szacunek otrzymany metodą MNK.

Jedna z metod służących do estymacji parametrów przy jednoczesnym nałożeniu warunków ograniczających na te parametry to **MNK przy warunkach pobocznych** (constrained least squares method). Wzór estymatora można wyprowadzić w podobny sposób jak w klasycznej metodzie najmniejszych kwadratów, to znaczy minimalizując sumę kwadratów błędów regresji, ale przy jednoczesnym zachowaniu warunku $\boldsymbol{R\beta} = \boldsymbol{r}$:

$$J(\boldsymbol{\beta}) = (\boldsymbol{y} - \boldsymbol{X\beta})'(\boldsymbol{y} - \boldsymbol{X\beta}) - 2\lambda'(\boldsymbol{R\beta} - \boldsymbol{r}) \rightarrow min$$

Rozwiązaniem jest estymator postaci:

$$\widehat{\boldsymbol{\beta}}^* = \widehat{\boldsymbol{\beta}} + (\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}'(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}\left(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}\right) \qquad (1.15)$$

gdzie $\widehat{\boldsymbol{\beta}}$ oznacza estymator MNK.

Reszty modelu regresji otrzymujemy odejmując od prawdziwych wartości y_i wartości teoretyczne, $\boldsymbol{e}^* = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^*$. Sumę kwadratów reszt regresji z ograniczeniami możemy wtedy zapisać jako:

$$\boldsymbol{e}^{*\prime}\boldsymbol{e}^* = \boldsymbol{e}'\boldsymbol{e} + (\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}})'\boldsymbol{X}'\boldsymbol{X}(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}), \qquad (1.16)$$

ponieważ $\boldsymbol{e}^* = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}^* = \boldsymbol{y} - X\widehat{\boldsymbol{\beta}} - \boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right) = \boldsymbol{e} - \boldsymbol{X}\left(\widehat{\boldsymbol{\beta}}^* - \widehat{\boldsymbol{\beta}}\right)$. Reszty z modelu regresji bez restrykcji zapisano jako $\boldsymbol{e} = \boldsymbol{y} - \boldsymbol{X}\widehat{\boldsymbol{\beta}}$.

Po podstawieniu ze wzoru (1.15) różnicy między estymatorami $\widehat{\boldsymbol{\beta}}^*$ i $\widehat{\boldsymbol{\beta}}$ do wzoru (1.16) otrzymujemy po uproszczeniu następujące wyrażenie:

$$\boldsymbol{e}^{*\prime}\boldsymbol{e}^* = \boldsymbol{e}'\boldsymbol{e} + (\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}})'(\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}(\boldsymbol{r} - \boldsymbol{R}\widehat{\boldsymbol{\beta}}) \qquad (1.17)$$

W poprzednim punkcie wyprowadzono wzór (1.13) na statystykę F postaci:

$$F = \left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)'(\hat{\sigma}^2\boldsymbol{R}(\boldsymbol{X}'\boldsymbol{X})^{-1}\boldsymbol{R}')^{-1}\left(\boldsymbol{R}\widehat{\boldsymbol{\beta}} - \boldsymbol{r}\right)/m \sim F(m, n-k).$$

Wykorzystując wzór (1.17) można tę samą statystykę zapisać wzorem:

$$F = \frac{(\boldsymbol{e}^{*\prime}\boldsymbol{e}^* - \boldsymbol{e}'\boldsymbol{e})/m}{\boldsymbol{e}'\boldsymbol{e}/(n-k)} \sim F(m, n-k). \qquad (1.18)$$

Przykład 1.15

W tym przykładzie oszacujmy model regresji liniowej przy pomocy MNK przy warunkach pobocznych oraz zapiszmy statystykę F, służącą do weryfikacji restrykcji nałożonych na parametry modelu.

Wygenerujmy obserwacje zmiennych objaśniających i objaśnianej oraz wektor „prawdziwych" wartości parametrów.

```
X =[ones(100,1) randn(100,4)];  % stała i cztery zmienne generowane z N(0,1)
e = randn(100,1);               % wektor składników losowych z N(0,1)
alfa = [1; 2; 3; 4; 5];         % prawdziwe parametry modelu
y = X*alfa+e;                   % wartości y wygenerowane z modelu
```

Udając, że nie znamy wartości parametrów modelu, oszacujmy je przy pomocy MNK oraz przy pomocy MNK z restrykcją $\alpha_1 + \alpha_2 + \alpha_3 = 3$. Taką restrykcję zapiszemy przy pomocy macierzy $\boldsymbol{R}$ i $\boldsymbol{r}$ jako $\boldsymbol{R} = [1 \quad 1 \quad 1 \quad 0 \quad 0]$, $\boldsymbol{r} = [3]$.

```
alfa_MNK = X\y                  % oszacowania MNK
R = [1 1 1 0 0];                % restrykcje na parametry
r = 3;
alfa_RESTR = alfa_MNK + inv(X'*X)*R'*inv(R*inv(X'*X)*R')*(r-R*alfa_MNK)
                                % oszacowanie MNK z restrykcją
sum(alfa_RESTR(1:3))            % sprawdzenie warunku a1+a2+a3=3
```

Możemy także policzyć statystykę F testującą ten warunek zgodnie ze wzorem (1.13)

```
u   = y - X*alfa_MNK;           % reszty z regresji szacowanej MNK
u_r = y - X*alfa_RESTR;         % reszty z regresji z restrykcjami
m = 1;                          % liczba niezależnych restrykcji
[n k] = size(X);                % liczba obserwacji n i liczba parametrów k
F = (u_r'*u_r - u'*u)/(u'*u)*(n-k)/m    % statystyka F
p_F = 1-fcdf(F,m,n-k)           % p-value statystyki F
```

1.5. Nieliniowe restrykcje w liniowym modelu regresji

Rozważamy model regresji liniowej postaci:

$$y_i = \beta_1 x_{1,i} + \beta_2 x_{2,i} + \ldots + \beta_K x_{K,i} + u_i$$

gdzie y_i oznacza i-tą obserwację zmiennej objaśnianej, $x_{1,i}$, $x_{2,i}$, …, $x_{k,i}$ oznaczają i-te obserwacje zmiennych objaśniających, a u_i oznacza i-tą obserwację składnika losowego dla $i = 1,\ldots,n$. Wśród zmiennych objaśniających może występować stała, czyli wyraz wolny. Postać tego modelu w formie macierzowej jest następująca:

$$\boldsymbol{y} = \boldsymbol{X\beta} + \boldsymbol{u}$$

Wektor $\boldsymbol{y}$ zawiera n obserwacji zmiennej objaśnianej, macierz $\boldsymbol{X}$ o wymiarach $n \times k$ zawiera obserwacje zmiennych objaśniających, a $\boldsymbol{u}$ jest n-elementowym wektorem składników losowych.

Sprawdźmy, czy parametry takiego modelu regresji spełniają pewne, potencjalnie nieliniowe restrykcje, na przykład $\beta_1/(\beta_2 + \beta_3) = 1$.

Niech $\mathbf{g}(\boldsymbol{\beta}): \Re^k \to \Re^m$, funkcja wektora parametrów regresji, spełnia równanie $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}_{m\times 1}$. Zwróćmy uwagę, że szczególnym przypadkiem $\mathbf{g}(\boldsymbol{\beta})$ jest liniowa funkcja parametrów zadana wzorem $\mathbf{g}(\boldsymbol{\beta}) = \boldsymbol{R\beta} - \boldsymbol{r}$. Zdefiniujmy następnie macierz $\boldsymbol{G}(\boldsymbol{\beta})$ o wymiarach $m \times k$

jako macierz pierwszych pochodnych z funkcji $\mathbf{g}(\boldsymbol{\beta})$ po wektorze parametrów $\boldsymbol{\beta}$, $\boldsymbol{G}(\boldsymbol{\beta}) = \frac{\partial \boldsymbol{g}}{\partial \boldsymbol{\beta}} = \begin{bmatrix} \frac{\partial \boldsymbol{g}}{\partial \beta_1} & \cdots & \frac{\partial \boldsymbol{g}}{\partial \beta_k} \end{bmatrix}$.

Przykład 1.16

Zdefiniujmy funkcję $\mathbf{g}(\boldsymbol{\beta})$ i zapiszmy macierz $\boldsymbol{G}(\boldsymbol{\beta})$ dla warunku $\beta_1/(\beta_2 + \beta_3) = 1$ w liniowym modelu regresji z trzema parametrami, $\boldsymbol{\beta} = [\beta_1 \quad \beta_2 \quad \beta_3]'$:

$$\mathbf{g}(\boldsymbol{\beta}) = \beta_1/(\beta_2 + \beta_3) - 1$$

$$\boldsymbol{G}(\boldsymbol{\beta}) = \begin{bmatrix} \frac{1}{\beta_2 + \beta_3} & -\frac{\beta_1}{(\beta_2 + \beta_3)^2} & -\frac{\beta_1}{(\beta_2 + \beta_3)^2} \end{bmatrix}$$

W tym przypadku $m = 1$ i $k = 3$.

Do testowania hipotezy $\mathcal{H}_0$ zakładającej, że $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$, służy statystyka Walda postaci

$$W = n \cdot \mathbf{g}(\boldsymbol{b})'(\boldsymbol{G}(\boldsymbol{b})\boldsymbol{V}_n\boldsymbol{G}(\boldsymbol{b})')^{-\mathbf{1}}\mathbf{g}(\boldsymbol{b})$$

gdzie $\boldsymbol{V}_n = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}\boldsymbol{\Omega}_n\,(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}$. Macierz $\boldsymbol{\Omega}_n$ jest zgodnym estymatorem macierzy wariancji $E(x_i x_i' u_i^2)$. Na przykład w modelu z warunkową homoskedastycznością $\boldsymbol{\Omega}_n^0 = \boldsymbol{X}'\boldsymbol{X}s^{\mathbf{2}}$, $s^{\mathbf{2}} = \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}/(n-k)$, a w modelu heteroskedastycznym można użyć $\boldsymbol{\Omega}_n = \sum_{i=1}^{\boldsymbol{n}} \boldsymbol{x}_i\boldsymbol{x}_i'\hat{u}_i^2$. W modelu homoskedastycznym $\boldsymbol{V}_n = s^{\mathbf{2}}(\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}$. Symbolem $\boldsymbol{b}$ oznaczono oszacowania parametrów $\boldsymbol{\beta}$ otrzymane metodą najmniejszych kwadratów.

Przy założeniu prawdziwości $\mathcal{H}_0$ statystyka W ma asymptotyczny rozkład χ^2 z m stopniami swobody, gdzie m oznacza liczbę niezależnych restrykcji, czyli liczbę wierszy w układzie równań $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}_{m\times 1}$.

Przykład 1.17

Wykorzystajmy wysymulowane dane i obliczenia z przykładu 1.16 do konstrukcji odpowiedniej statystyki W weryfikującej warunek $\beta_1/(\beta_2 + \beta_3) = 1$.

Wygenerujmy obserwacje zmiennych objaśniających i objaśnianej oraz wektor „prawdziwych” wartości parametrów.

```
X =[ones(100,1) randn(100,2)];  % stała i dwie zmienne generowane z N(0,1)
e = randn(100,1);               % wektor składników losowych z N(0,1)
alfa = [3; 2; 1];               % prawdziwe parametry modelu
y = X*alfa+e;                   % wartości y wygenerowane z modelu
```

Policzmy oszacowania parametrów i macierzy wariancji oszacowań parametrów.

```
beta = X\y                      % oszacowania MNK
u = y - X*beta;                 % obliczanie reszt z modelu regresji
[n k] = size(X);                % liczba obserwacji n i liczba parametrów k
sigma2 = u'*u/(n-k);            % szacunek wariancji składnika losowego
V = sigma2*inv(X'*X);           % macierz wariancji oszacowań parametrów
```

Zapiszmy restrykcje na parametry $\mathbf{g}(\boldsymbol{\beta})$ i wektor $\boldsymbol{G}(\boldsymbol{\beta})$ pochodnych $\mathbf{g}(\boldsymbol{\beta})$ po parametrach.

```
g = beta(1)/(beta(2)+beta(3))-1         % nieliniowa restrykcja na parametry
G = [1/(beta(2)+beta(3)) -beta(1)/sqrt(beta(2)+beta(3))...
    -beta(1)/sqrt(beta(2)+beta(3))]     % pochodne po parametrach z g()
m = 1;                                  % liczba niezależnych restrykcji
```

W powyższym kodzie zwróćmy uwagę na możliwość przeniesienia części długiego wyrażenia do następnej linii. W końcu policzmy statystykę W i jej wartość p.

```
W = n*g'*inv(G*V*G')*g          % statystyka Walda
p_W = 1-chi2cdf(W,m)            % p-value statystyki W
```

Restrykcje nieliniowe można zapisać na różny sposób, tak by statystyka Walda miała różną wartość. Na przykład warunek $\beta_1/\beta_2 + \beta_3 - 1 = 0$ jest identyczny jak $\beta_1 + \beta_2 \cdot \beta_3 - \beta_2 = 0$ przy założeniu $\beta_2 \neq 0$, ale statystyka Walda ma w każdym przypadku inną postać i inną wartość. W obu przypadkach asymptotyczne wartości krytyczne statystyki W są jednak identyczne, ponieważ $W \xrightarrow{d} \chi_1^2$. W badaniu empirycznym może się zatem zdarzyć, że nie będzie podstaw do odrzucenia warunku pierwszego, a odrzucony zostanie identyczny warunek drugi. Możliwość otrzymania różnych wyników testu dla identycznych, ale różnie zapisanych nieliniowych restrykcji jest poważnym problemem przy analizowaniu restrykcji nieliniowych w modelach ekonometrycznych (Gregory i Veall, 1985).

Możliwym rozwiązaniem jest zastosowanie testu minimalnej odległości (ang. minimum-distance test) Neweya i Westa (1987). Test ten polega na porównaniu statytystyk $J(\boldsymbol{\beta})$ uogólnionej metody momentów (UMM) policzonych, odpowiednio, dla modelu z restrykcjami $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$ i bez restrykcji. Statystyki $J(\boldsymbol{\beta})$ stanowią kryteria minimalizacji otrzymywane w procesie estymacji parametrów modeli. W modelu regresji liniowej oszacowanie parametrów $\boldsymbol{\beta}$ przy pomocy UMM polega na znalezieniu takich wartości parametrów $\widehat{\boldsymbol{\beta}}$, dla których minimum osiąga statystyka $J(\boldsymbol{\beta})$ postaci:

$$J(\boldsymbol{\beta}) = n(\boldsymbol{y} - \boldsymbol{X\beta})'\boldsymbol{X}\boldsymbol{\Omega}_n^{-1}\boldsymbol{X}'(\boldsymbol{y} - \boldsymbol{X\beta})$$

gdzie $\mathbf{\Omega}_n$ jest zgodnym estymatorem macierzy wariancji $E(x_i x_i' u_i^2)$. Na przykład w modelu z warunkową homoskedastycznością $\mathbf{\Omega}_n^0 = \boldsymbol{X}'\boldsymbol{X}s^{\mathbf{2}}$, $s^{\mathbf{2}} = \widehat{\boldsymbol{u}}'\widehat{\boldsymbol{u}}/(n-k)$, a w modelu heteroskedastycznym można użyć $\mathbf{\Omega}_n = \sum_{i=1}^{\boldsymbol{n}} \boldsymbol{x}_i \boldsymbol{x}_i' \hat{u}_i^2$.

W przypadku modelu bez restrykcji estymator UMM ma identyczną postać jak estymator MNK:

$$\widehat{\boldsymbol{\beta}} = \underset{\boldsymbol{\beta} \in \Re^k}{\operatorname{argmin}} J(\boldsymbol{\beta}) = (\boldsymbol{X}'\boldsymbol{X})^{-\mathbf{1}}(\boldsymbol{X}'\boldsymbol{y})$$

Natomiast dla modelu regresji z nieliniowymi restrykcjami estymator UMM nie ma ustalonej formuły i należy wykorzystać metody numeryczne do znalezienia minimalnej wartości $J(\boldsymbol{\beta})$:

$$\widetilde{\boldsymbol{\beta}} = \underset{\mathbf{g}(\boldsymbol{\beta})=\mathbf{0}}{\operatorname{argmin}} J(\boldsymbol{\beta})$$

Więcej informacji na temat numerycznych metod optymalizacji przedstawiono w rozdziale 2.

Statystyka testowa ma postać:

$$D = J(\widetilde{\boldsymbol{\beta}}) - J(\widehat{\boldsymbol{\beta}}) \sim \chi_m^2$$

i jest ona odporna na algebraiczny sposób zapisania nieliniowego warunku $\mathbf{g}(\boldsymbol{\beta}) = \mathbf{0}$. Co ciekawe wartości statystyk D i W są identyczne, gdy testowane są restrykcje liniowe i używane są identyczne estymatory $E(x_i x_i' u_i^2)$.

Literatura

Greene, W. (2012) Econometric Analysis, wyd. 7, Pearson.

Gregory, A., Veall, M. (1985) Formulating Wald tests of nonlinear restrictions, Econometrica, 53 (6), 1465-1468.

Hamilton J. (1994) Time Series Analysis, Princeton University Press.

Hansen, B. (2018) Econometrics, University of Wisconsin.

Hayashi, F. (2000) Econometrics, Princeton University Press.

Johnston, J., DiNardo, J. (1997) Econometric Methods, wyd. 4, McGraw-Hill.

Newey, W., West, K. (1987) Hypothesis testing with efficient method of moments estimation. International Economic Review, 28 (3), 777-787.

Romano, J., Wolf, M. (2017) Resurrecting weighted least squares. Journal of Econometrics, 197(1), 1–19.

White, H. (1980) A Heteroskedasticity-Consistent Covariance Matrix Estimator and a Direct Test for Heteroskedasticity, Econometrica, 48(4), 817-838.