

Rozdział 3: Metoda największej wiarygodności i nieliniowa metoda najmniejszych kwadratów

W tym rozdziale omówione zostaną dwie najpopularniejsze metody estymacji parametrów w ekonometrycznych modelach nieliniowych, a mianowicie metoda największej wiarygodności (MNW) oraz nieliniowa metoda najmniejszych kwadratów (NMNK). Przedstawimy założenia i własności obu metod, przykładowe zastosowania oraz kody programów. Obie metody estymacji należą do grupy ekstremum-estymatorów (extreme estimators) i mają podobne założenia i własności. Dlatego dokładniej na początku rozdziału jest omówiona MNW, a nieco mniej dokładnie NMNK na końcu rozdziału.

Na własnościach asymptotycznych MNW bazują trzy testy statystyczne, pozwalające zweryfikować poprawność restrykcji nakładanych na parametry modelu – a mianowicie test Walda, test ilorazu wiarygodności i test mnożnika Lagrange’a. Wszystkie trzy testy są często stosowane do sprawdzania specyfikacji nieliniowych modeli regresji i porównania tych specyfikacji z liniowym modelem regresji. Dlatego testy te zostaną opisane w dalszej części rozdziału przed opisem NMNK.

3.1. Metoda największej wiarygodności

Metoda największej wiarygodności (MNW, ang. *maximum likelihood*) to metoda estymacji parametrów w modelu ekonometrycznym. W pewnym uproszczeniu polega ona na wyznaczeniu wartości parametrów, które maksymalizowałyby prawdopodobieństwo wygenerowania przez model takich wartości zmiennej objaśnianej (lub zmiennych objaśnianych), które zaobserwowaliśmy w rzeczywistości.

W celu zastosowania tej metody estymacji buduje się tak zwaną funkcję wiarygodności L , która jest funkcją gęstości $f(\mathbf{y}; \boldsymbol{\theta})$ łącznego rozkładu n obserwacji zmiennej objaśnianej \mathbf{y} przy zadanych wartościach wektora parametrów $\boldsymbol{\theta}$. W przypadku, gdy \mathbf{y} ma rozkład dyskretny, $f(\mathbf{y}; \boldsymbol{\theta})$ oznacza łączny rozkład prawdopodobieństwa. Ponieważ $\boldsymbol{\theta}$ może zawierać różne parametry takie jak na przykład parametr wariancji składnika losowego σ^2 , a nie tylko parametry $\boldsymbol{\beta}$ przy zmiennych objaśniających, to w dalszej części opisu MNW będziemy zapisywali wektor parametrów modelu jako $\boldsymbol{\theta}$.

Przy estymacji parametrów $\boldsymbol{\theta}$ wartości \mathbf{y} są znane, a poszukiwane są właśnie wartości parametrów $\boldsymbol{\theta}$. Dlatego funkcję wiarygodności zapisuje się jako:

$$L(\boldsymbol{\theta}; \mathbf{y}) = f(\mathbf{y}; \boldsymbol{\theta}) \quad (3.1)$$

Proces estymacji, czyli poszukiwania wartości parametrów $\boldsymbol{\theta}$ polega na wyznaczeniu takich ich wartości, które maksymalizują funkcję wiarygodności $L(\boldsymbol{\theta}; \mathbf{y})$. Ze względów obliczeniowych często poszukuje się maksimum logarytmu naturalnego funkcji wiarygodności, czyli $l(\boldsymbol{\theta}) = \ln L(\boldsymbol{\theta}; \mathbf{y})$, zamiast $L(\boldsymbol{\theta}; \mathbf{y})$. Oszacowanie $\hat{\boldsymbol{\theta}}_{MNW}$ (otrzymane MNW) wektora parametrów $\boldsymbol{\theta}$, które maksymalizuje $l(\boldsymbol{\theta})$, maksymalizuje jednocześnie także $L(\boldsymbol{\theta}; \mathbf{y})$:

$$\hat{\boldsymbol{\theta}}_{MNW} = \underset{\boldsymbol{\theta} \in \Theta}{\operatorname{argmax}} \ln L(\boldsymbol{\theta}; \mathbf{y}) . \quad (3.2)$$

Niech $s(\boldsymbol{\theta}; \mathbf{y}) = \frac{\partial l}{\partial \boldsymbol{\theta}}$ oznacza wektor pochodnych cząstkowych logarytmu funkcji wiarygodności, czyli gradient funkcji $l(\boldsymbol{\theta})$ (ang. score). W optimum funkcji jej gradient jest wektorem zerowym, czyli:

$$s(\hat{\boldsymbol{\theta}}_{MNW}; \mathbf{y}) = \left. \frac{\partial l}{\partial \boldsymbol{\theta}} \right|_{\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_{MNW}} = \mathbf{0} . \quad (3.3)$$

Funkcja wiarygodności opisuje łączny rozkład wektora obserwacji \mathbf{y} . W praktyce wygodniej jest analizować pojedyncze (czasami brzegowe) rozkłady poszczególnych elementów (obserwacji) tego wektora. Dlatego wykorzystuje się definicję warunkowej funkcji gęstości by rozbić łączną funkcję gęstości $f(\mathbf{y}; \boldsymbol{\theta})$ na iloczyn n funkcji pojedynczych obserwacji:

$$\begin{aligned} L(\boldsymbol{\theta}; \mathbf{y}) &= f(\mathbf{y}; \boldsymbol{\theta}) = f(y_1, y_2, \dots, y_n | \boldsymbol{\theta}) = \\ &= f(y_1 | \boldsymbol{\theta}) \cdot f(y_2 | y_1, \boldsymbol{\theta}) \cdot \dots \cdot f(y_n | y_1, y_2, \dots, y_{n-1}, \boldsymbol{\theta}) . \end{aligned} \quad (3.4)$$

W szczególności, jeśli obserwacje y_1, y_2, \dots, y_n są wzajemnie niezależne, to można uprościć zapis:

$$L(\boldsymbol{\theta}; \mathbf{y}) = f(\mathbf{y}; \boldsymbol{\theta}) = f(y_1, y_2, \dots, y_n | \boldsymbol{\theta}) = f(y_1 | \boldsymbol{\theta}) \cdot f(y_2 | \boldsymbol{\theta}) \cdot \dots \cdot f(y_n | \boldsymbol{\theta}) \quad (3.5)$$

Zazwyczaj obliczeń dokonuje się wykorzystując logarytm funkcji wiarygodności:

$$\begin{aligned} l(\boldsymbol{\theta}) &= \ln f(\mathbf{y}; \boldsymbol{\theta}) = \\ &= \ln f(y_1 | \boldsymbol{\theta}) + \ln f(y_2 | y_1, \boldsymbol{\theta}) + \dots \\ &+ \ln f(y_n | y_1, y_2, \dots, y_{n-1}, \boldsymbol{\theta}) . \end{aligned} \quad (3.6)$$

W przypadku niezależnych obserwacji y_i możemy ten wzór zapisać w następujący sposób:

$$l(\boldsymbol{\theta}) = \ln f(\mathbf{y}; \boldsymbol{\theta}) = \ln f(y_1 | \boldsymbol{\theta}) + \ln f(y_2 | \boldsymbol{\theta}) + \dots + \ln f(y_n | \boldsymbol{\theta}) . \quad (3.7)$$

W praktyce największa trudność w zastosowaniu estymacji MNW polega właśnie na skonstruowaniu odpowiedniej funkcji wiarygodności. Dlatego w kolejnych przykładach

przedstawiono sposób wyznaczania funkcji wiarygodności dla kilku popularnych modeli ekonometrycznych.

3.2. Funkcja wiarygodności dla modelu regresji liniowej

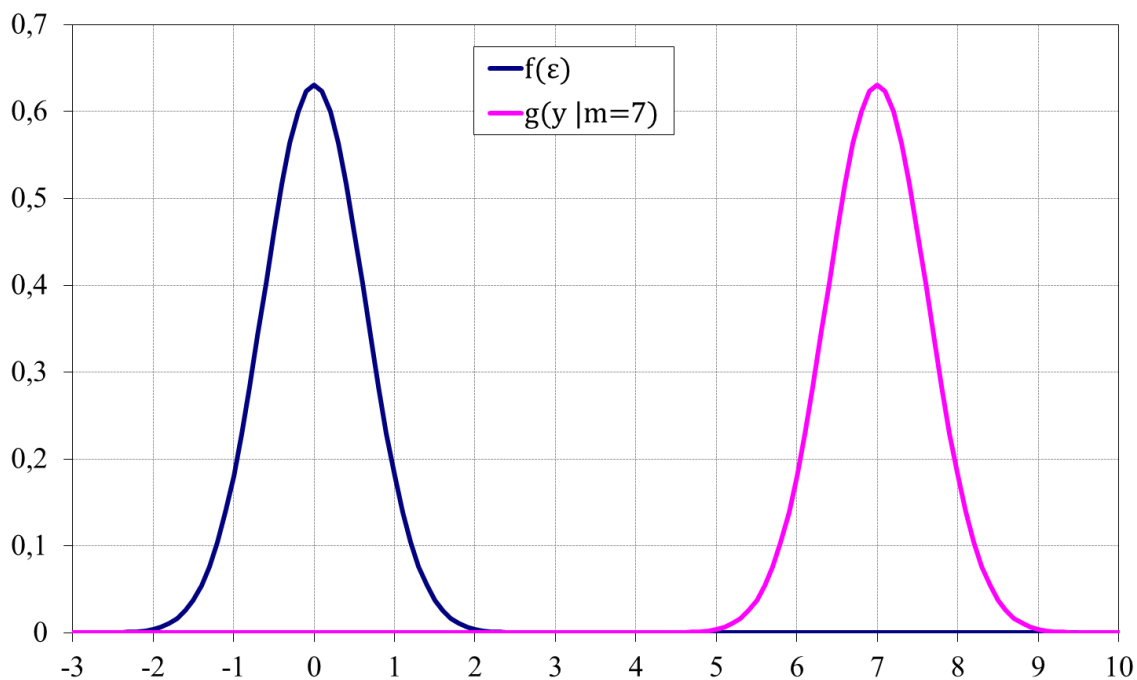
W tym punkcie przedstawiono przykłady konstrukcji funkcji wiarygodności dla modelu regresji liniowej. Zaprezentowano intuicyjne wytłumaczenie, dlaczego funkcja wiarygodności jest często zapisywana jako funkcja zależna składnika losowego modelu. Pokazano również procedurę obliczania funkcji wiarygodności na podstawie konkretnych danych.

Przykład 3.1

Pewna zmienna losowa ε ma rozkład normalny z wartością oczekiwaną 0 i wariancją równą 0,4. W takim razie zmienna losowa $y = \varepsilon + m$, gdzie $m = 7$, też ma rozkład normalny z wartością oczekiwaną 7 i wariancją 0,4 (por. Rysunek 3.1). Zapiszmy wzór na funkcję gęstości zmiennej ε oraz funkcję gęstości zmiennej y przy założeniu $m = 7$:

$$f(\varepsilon) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right)$$
$$g(y|m=7) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y-7)^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) = f(\varepsilon)$$

Rysunek 3.1: Wykresy rozkładów $f(\varepsilon)$ i $g(y|m=7)$



Fakt, że rozkłady y i ε są identyczne jest często wykorzystywany przy zapisywaniu funkcji wiarygodności w modelach, gdzie składnik losowy jest addytywny. Zwykle nie znamy rozkładu zmiennej objaśnianej y , ale możemy założyć, że rozkład składnika losowego ma jakąś znaną postać (np. rozkład normalny). Na ogół nie można też założyć, że kolejne obserwacje y są niezależne (np. w modelach autoregresyjnych kolejne obserwacje y są zależne), ale można założyć, że obserwacje składnika losowego są niezależne. Dlatego zamiast zapisywania funkcji wiarygodności jako funkcji $\mathbf{y} = [y_i]_{N \times 1}$ wygodniej jest ją zapisać jako funkcję obserwacji składnika losowego.

Przykład 3.2

Pewna zmienna losowa ε ma rozkład normalny z wartością oczekiwaną 0 i wariancją równą σ^2 . Rozpatrzmy funkcję gęstości zmiennej y , będącej funkcją liniową wektora losowych zmiennych objaśniających \mathbf{x} i niezależnego składnika losowego ε , $y = \mathbf{x}\boldsymbol{\beta} + \varepsilon$. Warunkowa wartość oczekiwana i wariancja y równe są odpowiednio:

$$E(y|\mathbf{x}) = E(\mathbf{x}\boldsymbol{\beta} + \varepsilon|\mathbf{x}) = \mathbf{x}\boldsymbol{\beta}$$

$$Var(y|\mathbf{x}) = Var(\mathbf{x}\boldsymbol{\beta} + \varepsilon|\mathbf{x}) = Var(\varepsilon|\mathbf{x}) = \sigma^2$$

Funkcja gęstości zmiennej y warunkowa na $\mathbf{x}\boldsymbol{\beta}$ równa jest:

$$g(y|\mathbf{x}\boldsymbol{\beta}) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(y - \mathbf{x}\boldsymbol{\beta})^2}{2\sigma^2}\right) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\varepsilon^2}{2\sigma^2}\right) = f(\varepsilon)$$

Dla funkcji regresji liniowej możemy zatem zapisać logarytm funkcji wiarygodności w następujący sposób:

$$\begin{aligned} l(\boldsymbol{\theta}) &= \ln g(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \\ &= \ln g(y_1|\boldsymbol{\theta}) + \ln g(y_2|y_1, \mathbf{X}, \boldsymbol{\theta}) + \dots \\ &+ \ln g(y_n|y_1, y_2, \dots, y_{n-1}, \mathbf{X}, \boldsymbol{\theta}) = \\ &= \ln g(\mathbf{x}_1\boldsymbol{\beta} + \varepsilon_1|\mathbf{X}, \boldsymbol{\theta}) + \ln g(\mathbf{x}_2\boldsymbol{\beta} + \varepsilon_2|y_1, \mathbf{X}, \boldsymbol{\theta}) + \dots \quad (3.8) \\ &+ \ln g(\mathbf{x}_n\boldsymbol{\beta} + \varepsilon_n|y_1, y_2, \dots, y_{n-1}, \mathbf{X}, \boldsymbol{\theta}) = \\ &= \ln f(\varepsilon_1|\mathbf{X}, \boldsymbol{\theta}) + \ln f(\varepsilon_2|y_1, \mathbf{X}, \boldsymbol{\theta}) + \dots \\ &+ \ln f(\varepsilon_n|y_1, y_2, \dots, y_{n-1}, \mathbf{X}, \boldsymbol{\theta}) \end{aligned}$$

gdzie \mathbf{X} oznacza macierz obserwacji zmiennych objaśniających \mathbf{x}_i , $i = 1, \dots, n$. Jeżeli składniki losowe ε_i są niezależne od \mathbf{X} i od y_1, y_2, \dots, y_{i-1} , to można powyższy wzór jeszcze uprościć:

$$l(\boldsymbol{\theta}) = \ln g(\mathbf{y}|\mathbf{X}, \boldsymbol{\theta}) = \ln f(\varepsilon_1|\mathbf{X}, \boldsymbol{\theta}) + \ln f(\varepsilon_2|\mathbf{X}, \boldsymbol{\theta}) + \dots + \ln f(\varepsilon_n|\mathbf{X}, \boldsymbol{\theta}) \quad (3.9)$$

Procedura szacowania parametrów przy użyciu MNW w modelu regresji liniowej jest następująca.

1. Dla ustalonych „startowych” wartości parametrów $\hat{\boldsymbol{\theta}}_{(0)} = [\hat{\boldsymbol{\beta}}_{(0)}' \hat{\sigma}_{(0)}^2]'$ oblicz reszty regresji $\hat{\varepsilon}_i^{(0)} = y_i - \mathbf{x}_i \hat{\boldsymbol{\beta}}_{(0)}$, gdzie $i = 1, \dots, n$.
2. Wyznacz wartości logarytmu funkcji gęstości dla reszt:

$$\ln f(\hat{\varepsilon}_i^{(0)}) = -\ln \sqrt{2\pi \hat{\sigma}_{(0)}^2} - \frac{(\hat{\varepsilon}_i^{(0)})^2}{2\hat{\sigma}_{(0)}^2}, \text{ gdzie } i = 1, \dots, n.$$
3. Policz logarytm funkcji wiarygodności: $l(\hat{\boldsymbol{\theta}}_{(0)}) = \ln g(\mathbf{y}|\mathbf{X}, \hat{\boldsymbol{\theta}}_{(0)}) = \ln f(\hat{\varepsilon}_1^{(0)}) + \ln f(\hat{\varepsilon}_2^{(0)}) + \dots + \ln f(\hat{\varepsilon}_n^{(0)})$.
4. Manipuluj wartościami szacowanych parametrów tak by wyznaczyć maksimum logarytmu funkcji wiarygodności. Manipulowanie wartościami szacowanych parametrów polega z reguły na stosowaniu numerycznych metod znajdowania optimum funkcji celu (w tym przypadku logarytmu funkcji wiarygodności). Przykłady takich metod przedstawiono w rozdziale 4. Dla każdego nowego wektora parametrów $\hat{\boldsymbol{\theta}}_{(i)}$ należy powtórzyć kroki 1 – 3.
5. Wartości parametrów $\hat{\boldsymbol{\theta}}_{max}$, dla których funkcja wiarygodności osiąga maksimum globalne, stanowią oszacowanie parametrów modelu uzyskane metodą MNW.

Warto zauważyć, że dla modelu regresji liniowej maksimum logarytmu funkcji wiarygodności można wyznaczyć analitycznie. Jednak przedstawione powyżej kroki procedury wyznaczania optimum funkcji wiarygodności są przydatne także przy konstrukcji funkcji wiarygodności bardziej skomplikowanych modeli ekonometrycznych, w tym nieliniowych modeli regresji.

Przykład 3.3

W tym przykładzie przedstawiamy sposób obliczania funkcji wiarygodności dla modelu regresji liniowej w programie MATLAB. Zdefiniowaną funkcję `log_wiaryg` należy zapisać w oddzielnym pliku o nazwie identycznej jak nazwa funkcji, czyli `log_wiaryg.m`.

```
function [l] = log_wiaryg(dane,param) %funkcja zwraca log. fun. wiarygodności
    X      = dane(:,2:end); % macierz obserwacji zmiennych objaśniających
    y      = dane(:,1);    % wektor obserwacji zmiennej objaśnianej
    beta   = param(1:end-1); % wektor parametrów regresji
    sigma  = param(end);   % odchylenie standardowe składnika losowego
```

```

e      = y - X*beta;      % reszty z regresji liniowej
lnf    = - e.^2/(2*sigma^2) - log(sqrt(2*pi*sigma^2));
l = sum(lnf);           % logarytm funkcji wiarygodności
end

```

Następnie możemy wykorzystać dowolne obserwacje zmiennych i przypisane wartości parametrów do obliczenia funkcji wiarygodności. Kod służący do wykonania obliczeń zapiszmy w oddzielnym pliku.

```

X = randn(100,5);      % generujemy losowe wartości X z rozkładu N(0,1)
u = randn(100,1);     % generujemy wartości składnika losowego z N(0,1)
b = [1 2 3 2 1]';     % nadajemy wartości parametrom regresji
y = X*b + u;          % generujemy wartości zmiennej y z modelu
sigma = 1;            % odch. stand. składnika losowego = std(u)
param = [b; sigma];   % przepisujemy parametry do jednego wektora
dane = [y X];         % przepisujemy wszystkie dane do macierzy
[l] = log_wiaryg(dane,param) % oblicza log. funkcji wiarygodności

```

Ostatnie polecenie wykonuje działania zapisane w funkcji `log_wiaryg` i zwraca wartość logarytmu funkcji wiarygodności dla zadanych danych i wartości parametrów.

Gdybyśmy nie znali wartości parametrów, to moglibyśmy je oszacować maksymalizując logarytm funkcji wiarygodności. Wykorzystamy w tym celu następujący kod. Ten fragment kodu wymaga posiadania biblioteki Optimization Toolbox (w Octave biblioteka `optim`; polecenie `pkg load optim`). Poniższe polecenia zostaną dokładnie wytłumaczone w rozdziale 4.

```

f = @(param) -log_wiaryg(dane,param);      % tzw. funkcja anonimowa f
[optymalne_beta_sigma,max_l] = fminunc(f,param) % szukamy minimum funkcji f

```

Oszacowane wartości parametrów znajdują się w wektorze `optymalne_beta_sigma`.

3.3. Metoda największej wiarygodności do szacowania parametrów innych modeli ekonometrycznych

Przykład 3.4. Model autoregresji

W modelu autoregresji postaci $y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t$, gdzie $\varepsilon_t \sim N(0, \sigma^2)$, szacowane są wartości parametrów $\alpha_0, \alpha_1, \dots, \alpha_p$ i σ^2 . Procedura wyznaczania oszacowań jest analogiczna, jak w przypadku modeli regresji liniowej. W pierwszym kroku wyznaczane są reszty $\hat{\varepsilon}_i^{(0)} = y_i - (\hat{\alpha}_0^{(0)} + \hat{\alpha}_1^{(0)} y_{i-1} + \dots + \hat{\alpha}_p^{(0)} y_{i-p})$ dla $t = p + 1, \dots, n$.

Następnie wyliczane są wartości $\ln f(\hat{\varepsilon}_t^{(0)}) = -\ln \sqrt{2\pi\hat{\sigma}_{(0)}^2} - \frac{(\hat{\varepsilon}_t^{(0)})^2}{2\hat{\sigma}_{(0)}^2}$ dla $t = p + 1, \dots, n$ oraz logarytm funkcji wiarygodności $l(\hat{\theta}_{(0)}) = \ln f(\hat{\varepsilon}_{p+1}^{(0)}) + \ln f(\hat{\varepsilon}_{p+2}^{(0)}) + \dots + \ln f(\hat{\varepsilon}_n^{(0)})$. Znalezienie wartości oszacowań parametrów polega tutaj na wyznaczeniu takich ich wartości, które maksymalizują wartość (logarytmu) funkcji wiarygodności.

Przykład 3.5. Model ARMA

W modelu autoregresyjnym ze średnią ruchomą (ang. autoregressive moving average), postaci $y_t = \alpha_0 + \alpha_1 y_{t-1} + \dots + \alpha_p y_{t-p} + \varepsilon_t + \delta_1 \varepsilon_{t-1} + \dots + \delta_q \varepsilon_{t-q}$, gdzie $\varepsilon_t \sim N(0, \sigma^2)$, szacowane są wartości parametrów $\alpha_0, \alpha_1, \dots, \alpha_p, \delta_1, \dots, \delta_q$ i σ^2 .

Procedura wyznaczania oszacowań jest analogiczna, jak w przypadku modeli regresji liniowej. W pierwszym kroku wyznaczone są rekurencyjnie reszty $\hat{\varepsilon}_t^{(0)} = y_t - (\hat{\alpha}_0^{(0)} + \hat{\alpha}_1^{(0)} y_{t-1} + \dots + \hat{\alpha}_p^{(0)} y_{t-p} + \delta_1^{(0)} \hat{\varepsilon}_{t-1}^{(0)} + \dots + \delta_q^{(0)} \hat{\varepsilon}_{t-q}^{(0)})$ dla $t = \max(p + 1, q + 1), \dots, n$, gdzie przyjmuje się często $\hat{\varepsilon}_t^{(0)} = 0$ dla $t = 1, \dots, q$. Następnie wyliczane są wartości

$$\ln f(\hat{\varepsilon}_t^{(0)}) = -\ln \sqrt{2\pi\hat{\sigma}_{(0)}^2} - \frac{(\hat{\varepsilon}_t^{(0)})^2}{2\hat{\sigma}_{(0)}^2} \text{ dla } t = \max(p + 1, q + 1), \dots, n$$

oraz logarytm funkcji wiarygodności

$$l(\hat{\theta}_{(0)}) = \ln f(\hat{\varepsilon}_i^{(0)}) + \ln f(\hat{\varepsilon}_{i+1}^{(0)}) + \dots + \ln f(\hat{\varepsilon}_n^{(0)}), \text{ gdzie } i = \max(p + 1, q + 1).$$

Znalezienie wartości oszacowań parametrów polega również tutaj na wyznaczeniu takich ich wartości, które maksymalizują wartość (logarytmu) funkcji wiarygodności.

Przykład 3.6

W tym przykładzie przedstawiamy sposób obliczania funkcji wiarygodności dla modelu ARMA w programie MATLAB. Zdefiniowaną funkcję `l_ARMA` należy zapisać w oddzielnym pliku o nazwie identycznej jak nazwa funkcji, czyli `l_ARMA.m`. Warto zwrócić uwagę, że reszty są wyliczane w pętli w sposób rekurencyjny. Dla ułatwienia zakładamy, że pierwsze obserwacje reszt są równe zero.

```
function [l] = l_ARMA(y,p,q,param) %funkcja zwraca log. fun. wiarygodności
    alfy = param(1:p+1); % wartości parametrów beta
    delty = param(p+2:end-1); % wartości parametrów delta
    sigma = param(end); % odchylenie standardowe składnika losowego
    start = max([p+1;q+1]); % wybierz moment rozpoczęcia iteracji
```

```

n      = length(y);          % liczba obserwacji zmiennej objaśnianej
e      = zeros(n,1);        % ustaw wartości reszt na zero
l      = 0;                 % początkowa wartość logarytmu funkcji wiarygodności
for i=start:n
    x = [1; y(i-p:i-1); e(i-q:i-1)]; % definiuje zmienne objaśniające
    e(i) = y(i) - [alfy' delty']*x; % obliczanie reszt
    lnf = - e(i)^2/(2*sigma^2) - log(sqrt(2*pi*sigma^2));
    l = l + lnf; % dodawanie kolejnych wartości do log. fun. wiarygod.
end
end

```

Następnie możemy wykorzystać dowolne obserwacje zmiennych i przypisane wartości parametrów do obliczenia funkcji wiarygodności. Kod służący do wykonania obliczeń zapiszmy w oddzielnym pliku. Załóżmy, że prawdziwy model to ARMA(1,1).

```

p = 1; % liczba opóźnień y w modelu ARMA(1,1)
q = 1; % liczba opóźnień e w modelu ARMA(1,1)
alfy = [0; 0.5]; % wartości parametrów alfa
delty = [0.3]; % wartości parametrów delta
y = zeros(100,1); % deklarujemy wektor y z samymi zerami
e = randn(100,1); % generujemy obserwacje losowe z N(0,1)
for t=2:100
    y(t) = [1 y(t-1)]*alfy + e(t) + e(t-1)*delty; % generujemy obserwacje y
end
sigma = 1; % odch. stand. składnika losowego = std(e)
param = [alfy; delty; sigma]; % przepisujemy parametry do jednego wektora
[1] = l_ARMA(y,p,q,param) % obliczamy log. funkcji wiarygodności

```

Ostatnie polecenie wykonuje działania zapisane w funkcji l_ARMA i zwraca wartość logarytmu funkcji wiarygodności dla zadanych danych i wartości parametrów.

Gdybyśmy nie znali wartości parametrów, to moglibyśmy je oszacować maksymalizując logarytm funkcji wiarygodności. Wykorzystamy w tym celu następujący kod. Ten fragment kodu wymaga posiadania biblioteki Optimization Toolbox (w Octave biblioteka optim; polecenie pkg load optim). Poniższe polecenia zostaną dokładnie wytłumaczone w rozdziale 4.

```

f = @(param) -l_ARMA(y,p,q,param); % tzw. funkcja anonimowa f
[optymalne_parametry,max_l] = fminunc(f,param) % szukamy minimum funkcji f

```

Oszacowane wartości parametrów znajdują się w wektorze optymalne_parametry.

Przykład 3.7 Model regresji liniowej z efektem GARCH(1,1)

W modelu regresji z efektem GARCH (ang. generalized autoregressive conditional heteroskedasticity) wariancja warunkowa składnika losowego zależy od swoich historycznych wartości od historycznych wartości kwadratu składnika losowego. Model ma następującą postać:

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + \varepsilon_t, \quad \varepsilon_t \sim N(0, \sigma_t^2),$$

$$\sigma_t^2 = \alpha_0 + \alpha_1 \varepsilon_{t-1}^2 + \alpha_2 \sigma_{t-1}^2, \quad \alpha_0 > 0, \alpha_1 > 0, \alpha_2 > 0, \alpha_1 + \alpha_2 < 1.$$

W modelu tym szacowane są wartości parametrów $\boldsymbol{\beta}$, α_0 , α_1 , α_2 . W pierwszym kroku wyznaczane są reszty $\hat{\varepsilon}_t^{(0)} = y_t - \mathbf{x}_t \hat{\boldsymbol{\beta}}_{(0)}$ oraz oszacowania wariancji warunkowej $\hat{\sigma}_t^2 = \hat{\alpha}_0^{(0)} + \hat{\alpha}_1^{(0)} (\hat{\varepsilon}_t^{(0)})^2 + \hat{\alpha}_2^{(0)} \hat{\sigma}_{t-1}^2$ kolejno dla $t = 2, \dots, n$. Zwykle przyjmuje się $\hat{\sigma}_1^2 = \hat{\alpha}_0 / (1 - \hat{\alpha}_1 - \hat{\alpha}_2)$. Następnie wyliczane są wartości $\ln f(\hat{\varepsilon}_t^{(0)}) = -\ln \sqrt{2\pi \hat{\sigma}_t^2} - \frac{(\hat{\varepsilon}_t^{(0)})^2}{2\hat{\sigma}_t^2}$ dla $t = 2, \dots, n$ oraz logarytm funkcji wiarygodności $l(\hat{\boldsymbol{\theta}}_{(0)}) = \ln f(\hat{\varepsilon}_2^{(0)}) + \ln f(\hat{\varepsilon}_3^{(0)}) + \dots + \ln f(\hat{\varepsilon}_n^{(0)})$. Znalezienie wartości oszacowań parametrów polega tutaj również na wyznaczeniu takich ich wartości, które maksymalizują wartość (logarytmu) funkcji wiarygodności.

Przykład 3.8

W tym przykładzie przedstawiamy sposób obliczania funkcji wiarygodności dla modelu GARCH(1,1) w programie MATLAB. Zdefiniowaną funkcję `l_GARCH11` należy zapisać w oddzielnym pliku o nazwie identycznej jak nazwa funkcji, czyli `l_GARCH11.m`. W wektorze parametrów `param` przechowywane są wartości parametrów bety z równania regresji oraz wartości parametrów alfy równania wariancji warunkowej.

```
function [l] = l_GARCH11(y,X,param) %funkcja zwraca log. fun. wiarygodności
    k = size(X,2); % liczba parametrów beta w regresji
    bety = param(1:k); % wartości parametrów beta
    alfy = param(k+1:end); % wartości parametrów alfa
    n = length(y); % liczba obserwacji zmiennej objaśnianej
    e = y - X*bety; % obliczamy wartości reszt
    l = 0; % początkowa wartość logarytmu funkcji wiarygodności
    sigma2 = exp(alfy(1))/(1-exp(alfy(2))-exp(alfy(3))); % wariancja e
    for t=2:n
        sigma2 = [1 e(t-1)^2 sigma2]*exp(alfy); % warunkowa wariancja e
        lnf = - e(t)^2/(2*sigma2) - 0.5*log(2*pi) - 0.5*log(sigma2);
        l = l + lnf; % dodawanie kolejnych wartości do log. fun. wiarygod.
    end
end
```

Zakładamy, że parametry równania wariancji warunkowej muszą być dodatnie, żeby zachować dodatnie wartości wariancji warunkowej `sigma2`. Dlatego wykorzystujemy w tym równaniu wyrażenia `exp(alfy)` zamiast `alfy`.

Następnie możemy wykorzystać dowolne (np. losowo wygenerowane) obserwacje zmiennych i przypisane wartości parametrów do obliczenia funkcji wiarygodności. Kod służący do wykonania obliczeń zapiszmy w oddzielnym pliku. Załóżmy, że prawdziwy jest

model regresji z tzw. efektem GARCH(1,1).

```
bety = [1; 2; 3];           % wartości parametrów beta
alfy = [0.1; 0.05; 0.90]; % wartości parametrów alfa
X = randn(1000,3);        % generujemy losowe wartości X z rozkładu N(0,1)
e = randn(1000,1);       % generujemy obserwacje losowe z N(0,1)
y = zeros(1000,1);       % deklarujemy wektor y z samymi zerami
sigma2 = ones(1000,1);    % deklarujemy wartości wariancji warunkowej e
sigma2(1) = alfy(1)/(1-alfy(2)-alfy(3)); % wariancja długookresowa e
for t=2:1000
    sigma2(t) = [1 e(t-1)^2 sigma2(t-1)]*alfy; % wariancja warunkowa e
    y(t) = X(t,:)*bety + e(t)*sqrt(sigma2(t)); % generujemy obserwacje y
end
param = [bety; log(alfy)]; % przepisujemy parametry do jednego wektora
[1] = l_GARCH11(y,X,param) % obliczamy log. funkcji wiarygodności
```

Ostatnie polecenie wykonuje działania zapisane w funkcji `l_GARCH11` i zwraca wartość logarytmu funkcji wiarygodności dla zadanych danych i wartości parametrów. Zauważmy, że do funkcji `l_GARCH11` podstawiamy między innymi wartości `log(alfy)` zamiast `alfy`.

Gdybyśmy nie znali wartości parametrów, to moglibyśmy je oszacować maksymalizując logarytm funkcji wiarygodności. Wykorzystamy w tym celu następujący kod. Ten fragment kodu wymaga posiadania biblioteki Optimization Toolbox (w Octave biblioteka `optim`; polecenie `pkg load optim`). Poniższe polecenia zostaną dokładnie wytłumaczone w rozdziale 4.

```
f = @(param) -l_GARCH11(y,X,param); % tzw. funkcja anonimowa f
[optymalne_parametry,max_l] = fminunc(f,param); % szukamy minimum funkcji f
beta_est = optymalne_parametry(1:3)
alfa_est = exp(optymalne_parametry(4:6))
Oszacowane wartości parametrów znajdują się w wektorach beta_est oraz alfa_est.
```

Przykład 3.9 Model logitowy

W modelu logitowym zmienna objaśniana Y może przyjmować wartości ze zbioru $\{0, 1\}$. Prawdopodobieństwo, że i -ta obserwacja Y (czyli y_i) równa jest 1 dane jest wzorem $\Pr(y_i = 1) = \exp(\mathbf{x}_i\boldsymbol{\beta}) / (1 + \exp(\mathbf{x}_i\boldsymbol{\beta}))$, natomiast $\Pr(y_i = 0) = 1 / (1 + \exp(\mathbf{x}_i\boldsymbol{\beta}))$. Przy założeniu, że obserwacje y_i są wzajemnie niezależne, można zapisać logarytm funkcji wiarygodności:

$$l(\boldsymbol{\theta}) = \ln f(\mathbf{y}; \boldsymbol{\theta}) = \ln f(y_1|\mathbf{x}_1\boldsymbol{\beta}) + \ln f(y_2|\mathbf{x}_2\boldsymbol{\beta}) + \dots + \ln f(y_n|\mathbf{x}_n\boldsymbol{\beta})$$

gdzie $f(y_i|\mathbf{x}_i\boldsymbol{\beta}) = \mathbf{1}(y_i = 1) \cdot \Pr(y_i = 1) + \mathbf{1}(y_i = 0) \cdot \Pr(y_i = 0)$ dla $i = 1, \dots, n$; $f(y_i|\mathbf{x}_i\boldsymbol{\beta})$ oznacza prawdopodobieństwo z jakim model wskaże taką wartość y_i , która została zaobserwowana w rzeczywistości. Funkcja $\mathbf{1}(\cdot)$ przyjmuje wartość 1, gdy jej argument jest zdaniem prawdziwym, oraz 0, gdy argument jest zdaniem fałszywym.

Przykład 3.10

W tym przykładzie przedstawiamy sposób obliczania funkcji wiarygodności dla modelu logitowego w programie MATLAB. Zdefiniowaną funkcję `l_logit` należy zapisać w oddzielnym pliku o nazwie identycznej jak nazwa funkcji, czyli `l_logit.m`. W wektorze parametrów `param` przechowywane są wartości parametrów bety z równania logitowego.

```
function [l] = l_logit(y,X,param) %funkcja zwraca log. fun. wiarygodności
    p = exp(X*param)./(1 + exp(X*param)); % wektor P(y=1|...)
    f = y.*p + (1-y).*(1-p); % wektor f(y|...) dla każdej obserwacji
    l = sum(log(f)); % logarytm funkcji wiarygodności
end
```

Następnie możemy wykorzystać dowolne (np. losowo wygenerowane) obserwacje zmiennych i przypisane wartości parametrów do obliczenia funkcji wiarygodności. Kod służący do wykonania obliczeń zapiszmy w oddzielnym pliku. Załóżmy, że prawdziwy jest model logitowy z trzema zmiennymi objaśniającymi.

```
X = randn(100,3); % generujemy losowe wartości X z rozkładu N(0,1)
beta = [1; 2; 3]; % wartości parametrów beta
Fu = rand(100,1); % losujemy liczby z rozkładu jednostajnego
u = -log(1./Fu - 1); % generujemy u z rozkładu logistycznego (0,1)
y = (X*beta + u > 0); % generujemy losowe y z modelu logitowego
[l] = l_logit(y,X,beta) % obliczamy logarytm funkcji wiarygodności
```

Ostatnie polecenie wykonuje działania zapisane w funkcji `l_logit` i zwraca wartość logarytmu funkcji wiarygodności dla zadanych danych i wartości parametrów.

Gdybyśmy nie znali wartości parametrów, to moglibyśmy je oszacować maksymalizując logarytm funkcji wiarygodności. Wykorzystamy w tym celu następujący kod. Ten fragment kodu wymaga posiadania biblioteki Optimization Toolbox (w Octave biblioteka `optim`; polecenie `pkg load optim`). Poniższe polecenia zostaną dokładnie wytłumaczone w rozdziale 4.

```
f = @(param) -l_logit(y,X,param); % tzw. funkcja anonimowa f=-l_logit()
[beta_est,max_l] = fminunc(f,beta) % szukamy minimum funkcji f
Oszacowane wartości parametrów znajdują się w wektorze beta_est.
```

3.4. Założenia i własności metody największej wiarygodności

W przypadku modeli regresji liniowej przyjmowane są zwykle identyczne założenia potrzebne do stosowania metody największej wiarygodności jak dla metody najmniejszych kwadratów. Spełnione powinno być założenie o normalności składnika losowego, ponieważ wtedy estymator MNW parametrów modelu regresji jest identyczny z estymatorem MNK.

Jeżeli rozkład składnika losowego jest inny, ale znany (np. rozkład t-Studenta), to możemy wyznaczyć oszacowania parametrów stosując procedurę opisaną w punkcie 3.2.

W przypadku bardziej ogólnym, obejmującym nieliniowe modele regresji, założenia do stosowania MNW będą zależały od konkretnego modelu. Zestaw założeń dla przypadku, gdzie obserwacje y_i są niezależne i pochodzą z ustalonego rozkładu, w przystępny sposób przedstawili na przykład Newey i McFadden (1994). Poniższy zestaw założeń przedstawiono za tymi autorami.

Założenie 3.1: Niech y_i , $i = 1, 2, \dots, n$, będą niezależne i pochodzą z jednakowego rozkładu $f(y_i|\theta_0)$ oraz dla $\theta \neq \theta_0$ zachodzi $f(y_i|\theta) \neq f(y_i|\theta_0)$.

Założenie 3.2: Zbiór $\Theta \subset \mathbb{R}^k$ parametrów θ_0 modelu jest zwarty.

Założenie 3.3: Funkcja $\ln f(y_i|\theta)$ jest ciągła dla każdego $\theta \in \Theta$ z prawdopodobieństwem 1.

Założenie 3.4: $E \left(\sup_{\theta \in \Theta} |\ln f(y|\theta)| \right) < \infty$.

Jeśli spełnione są powyższe warunki, to istnieje unikalne maksimum funkcji wiarygodności w punkcie θ_0 , istnieje estymator MNW $\hat{\theta}_n$ wektora dla parametrów θ_0 oraz estymator $\hat{\theta}_n$ jest zgodny, czyli $\text{plim}_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$. Dalsze założenia są potrzebne do wykazania „asymptotycznej normalności estymatora” MNW (Newey i McFadden, 1994).

Założenie 3.5: Wektor θ_0 jest punktem wewnętrznym zbioru Θ .

Założenie 3.6: Funkcja $f(y|\theta)$ jej dwukrotnie różniczkowalna względem θ i jej druga pochodna jest ciągła oraz $f(y|\theta) > 0$ w otoczeniu $N(\theta_0)$ punktu θ_0 .

Założenie 3.7: Niech $\int \sup_{\theta \in N(\theta_0)} \|\nabla_{\theta} f(y|\theta)\| dy < \infty$ oraz $\int \sup_{\theta \in N(\theta_0)} \|\nabla_{\theta\theta} f(y|\theta)\| dy < \infty$,

gdzie $\|A\|$ oznacza pierwiastek sumy kwadratów wszystkich elementów A .

Założenie 3.8: Istnieje nieosobliwa macierz $I(\theta_0) = E[(\nabla_{\theta} f(y|\theta_0))(\nabla_{\theta} f(y|\theta_0))']$.

Założenie 3.9: $E \left(\sup_{\theta \in N(\theta_0)} \|\nabla_{\theta\theta} \ln f(y|\theta)\| \right) < \infty$.

Jeśli spełnione są założenia 3.1 – 3.9, to odpowiednio przeskalowany estymator MNW ma asymptotyczny rozkład normalny, to znaczy $\sqrt{n}(\hat{\theta}_n - \theta_0) \xrightarrow{L} N(0, I(\theta_0)^{-1})$. Symbol \xrightarrow{L}

oznacza słabą zbieżność (zbieżność z dystrybuantą). Zwróćmy uwagę, że $\hat{\theta}_n \xrightarrow{L} \theta_0$, ponieważ $\text{plim}_{n \rightarrow \infty} \hat{\theta}_n = \theta_0$ (jest to estymator zgodny, a zbieżność z prawdopodobieństwem implikuje słabą zbieżność). Mimo że to rozkład przeskalowanego estymatora $\sqrt{n}\hat{\theta}_n$ zbiega do rozkładu normalnego przy $n \rightarrow \infty$, to i tak tę własność estymatora MNW określa się jako jego asymptotyczną normalność. Podobne założenia, uwzględniające możliwość istnienia zależności między obserwacjami zmiennych w modelu, tak jak na przykład w modelu autoregresyjnym, podali między innymi Heijmans i Magnus (1986).

Przedstawione wyżej założenia 3.1 – 3.9 implikują szereg własności estymatorów MNW.

Własność 3.1: Estymator MNW jest zgodny, to znaczy $\text{plim}_{n \rightarrow \infty} \hat{\theta}_n = \theta$. Ciąg wartości estymatora $\hat{\theta}_n$ zbiega z prawdopodobieństwem do prawdziwej wartości parametru θ przy $n \rightarrow \infty$.

Definicja 3.1: Macierz informacji to wartość oczekiwana iloczynu gradientów logarytmu funkcji wiarygodności dla pojedynczej obserwacji zmiennej y :

$$I(\theta) = E[(\nabla_{\theta} \ln f(y|\theta))(\nabla_{\theta} \ln f(y|\theta))'].$$

Należy dodać, że w podręcznikach często pojawia się definicja macierzy informacji dla funkcji wiarygodności zdefiniowanej dla n obserwacji zmiennej y , czyli $I^*(\theta) = E\left[\left(\frac{\partial l}{\partial \theta}\right)' \left(\frac{\partial l}{\partial \theta}\right)\right]$, gdzie l to logarytm funkcji wiarygodności $L(\theta; \mathbf{y})$ liczonej na całej próbie, zdefiniowanej wzorem (3.1). Ponieważ obserwacje y_i są niezależne (zgodnie z założeniem 3.1), to $I^*(\theta) = I(\theta) \cdot n$.

Własność 3.2: Dla estymatora MNW macierz informacyjna Fishera równa jest ujemnej wartości oczekiwanej hesjanu logarytmu funkcji wiarygodności dla pojedynczej obserwacji zmiennej y , co możemy zapisać

$$I(\theta) = E[(\nabla_{\theta} \ln f(y|\theta))(\nabla_{\theta} \ln f(y|\theta))'] = -E[\nabla_{\theta\theta} \ln f(y|\theta)] = H(\theta).$$

Zachodzi również analogiczna równość dla logarytmu funkcji wiarygodności na całej próbie $I^*(\theta) = E\left[\left(\frac{\partial l}{\partial \theta}\right)' \left(\frac{\partial l}{\partial \theta}\right)\right] = -E\left[\left(\frac{\partial^2 l}{\partial \theta \partial \theta'}\right)\right] = H^*(\theta)$.

Macierz informacji jest wykorzystywana często do wyliczania statystyk (np. statystyki Walda mnożnika Lagrange'a) i wariancji asymptotycznej estymatorów MNW. Dlatego warto przyjrzeć się dokładniej konstrukcji tej macierzy. Zrobimy to na przykładzie macierzy $\mathbf{I}^*(\boldsymbol{\theta})$ i $\mathbf{H}^*(\boldsymbol{\theta})$. Do obliczenia $\mathbf{I}^*(\boldsymbol{\theta})$ potrzebne jest policzenie gradientu logarytmu funkcji

wiarygodności (score), $\mathbf{s}(\mathbf{y}, \boldsymbol{\theta}) = \frac{\partial l}{\partial \boldsymbol{\theta}} = \begin{bmatrix} \partial l / \partial \theta_1 \\ \vdots \\ \partial l / \partial \theta_k \end{bmatrix}$. Macierz $\mathbf{I}^*(\boldsymbol{\theta})$ może być obliczona jako

wartość oczekiwana iloczynu $\mathbf{s}(\mathbf{y}, \boldsymbol{\theta})\mathbf{s}(\mathbf{y}, \boldsymbol{\theta})'$, ale może też być liczona jako ujemna wartość oczekiwana macierzy drugich pochodnych (hesjanu) logarytmu funkcji wiarygodności, $\mathbf{H}^*(\boldsymbol{\theta}) = -E \left[\left(\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \right]$. Ponieważ $\frac{\partial [\partial l / \partial \boldsymbol{\theta}]}{\partial \theta_i} = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta_i \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_i \partial \theta_2} & \dots & \frac{\partial^2 l}{\partial \theta_i \partial \theta_k} \end{bmatrix}$, to hesjan ma postać:

$$\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = \begin{bmatrix} \frac{\partial^2 l}{\partial \theta_1 \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_2} & \dots & \frac{\partial^2 l}{\partial \theta_1 \partial \theta_k} \\ \frac{\partial^2 l}{\partial \theta_2 \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_2 \partial \theta_2} & \dots & \frac{\partial^2 l}{\partial \theta_2 \partial \theta_k} \\ \vdots & \vdots & \ddots & \vdots \\ \frac{\partial^2 l}{\partial \theta_k \partial \theta_1} & \frac{\partial^2 l}{\partial \theta_k \partial \theta_2} & \dots & \frac{\partial^2 l}{\partial \theta_k \partial \theta_k} \end{bmatrix}. \quad (3.10)$$

W praktyce łatwiej jest policzyć pierwsze pochodne logarytmu funkcji wiarygodności niż drugie pochodne, dlatego $E \left[\left(\frac{\partial l}{\partial \boldsymbol{\theta}} \right)' \left(\frac{\partial l}{\partial \boldsymbol{\theta}} \right) \right]$ łatwiej jest policzyć niż $-E \left[\left(\frac{\partial^2 l}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) \right]$. Przy numerycznym przybliżaniu tych wyrażeń, mogą się one między sobą różnić.

Własność 3.3: Estymator MNW jest asymptotycznie normalny, to znaczy $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{L} N(\mathbf{0}, \mathbf{H}(\boldsymbol{\theta})^{-1} \cdot \mathbf{I}(\boldsymbol{\theta}) \cdot \mathbf{H}(\boldsymbol{\theta})^{-1}) = N(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta})^{-1}) = N(\mathbf{0}, \mathbf{H}(\boldsymbol{\theta})^{-1})$, gdzie $\mathbf{I}(\boldsymbol{\theta})$ jest macierzą informacji Fishera, a $\mathbf{H}(\boldsymbol{\theta}) = -E[\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ln f(\mathbf{y}|\boldsymbol{\theta})]$.

Własność 3.3 można wykorzystać do oszacowania tak zwanej „asymptotycznej wariancji” estymatora $\hat{\boldsymbol{\theta}}_n$, $avar(\hat{\boldsymbol{\theta}}_n)$, jako przybliżenia prawdziwej wariancji tego estymatora pomnożonej przez n . Asymptotyczna wariancja estymatora MNW przy założeniu poprawnej specyfikacji modelu jest równa odwrotności macierzy informacji Fishera, $avar(\hat{\boldsymbol{\theta}}_n) = \mathbf{I}(\boldsymbol{\theta})^{-1}$. Zatem zgodne oszacowania $avar(\hat{\boldsymbol{\theta}}_n)$ można otrzymać szacując $\mathbf{I}(\boldsymbol{\theta})$:

$$\hat{\mathbf{I}}_1 = n^{-1} \sum_{i=1}^n \left[\mathbf{s}(y_i, \hat{\boldsymbol{\theta}}_n) \cdot \mathbf{s}(y_i, \hat{\boldsymbol{\theta}}_n)' \right], \quad (3.11)$$

gdzie $\mathbf{s}(y_i, \hat{\boldsymbol{\theta}}_n) = \nabla_{\boldsymbol{\theta}} f(y|\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}$. Drugi sposób oszacowania $\mathbf{I}(\boldsymbol{\theta})$ wykorzystuje hesjan logarytmu funkcji wiarygodności:

$$\hat{\mathbf{I}}_2 = n^{-1} \sum_{i=1}^n [\nabla_{\boldsymbol{\theta}\boldsymbol{\theta}} \ln f(y|\boldsymbol{\theta})|_{\boldsymbol{\theta}=\hat{\boldsymbol{\theta}}_n}]. \quad (3.12)$$

Przy spełnieniu założeń 3.1-3.9, estymatory $\widehat{avar}_1(\hat{\boldsymbol{\theta}}_n) = \hat{\mathbf{I}}_1^{-1}$ i $\widehat{avar}_2(\hat{\boldsymbol{\theta}}_n) = \hat{\mathbf{I}}_2^{-1}$ są zgodne.

Kiedy model jest źle wyspecyfikowany, na przykład prawdziwy rozkład zmiennej objaśnianej y różni się od zakładanego $f(y|\boldsymbol{\theta})$, to wtedy własność 3.2 nie zachodzi i w związku z tym nie zachodzi też własność 3.3. Nie można wykorzystać wtedy $\hat{\mathbf{I}}_1^{-1}$ ani $\hat{\mathbf{I}}_2^{-1}$ jako zgodnych oszacowań wariancji estymatora $\hat{\boldsymbol{\theta}}_n$ parametrów $\boldsymbol{\theta}$. Gourieroux, Monfort i Trognon (1984) zauważyli, że jeśli prawdziwy rozkład y należy do szerokiej klasy rozkładów eksponencjalnych, to zgodnym estymatorem asymptotycznej wariancji będzie $\widehat{avar}_3(\hat{\boldsymbol{\theta}}_n) = \hat{\mathbf{I}}_2^{-1} \hat{\mathbf{I}}_1 \hat{\mathbf{I}}_2^{-1}$. Czasami taki estymator nazywany jest estymatorem wariancji odpornym na błędy specyfikacji.

Skoro estymator $\widehat{avar}(\hat{\boldsymbol{\theta}}_n)$ asymptotycznej wariancji $avar(\hat{\boldsymbol{\theta}}_n)$ aproksymuje wariancję wyrażenia $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta})$, to $\mathbf{V}_{\hat{\boldsymbol{\theta}}_n}^{MNW} = avar(\hat{\boldsymbol{\theta}}_n)/n$ przybliża wariancję estymatora $\hat{\boldsymbol{\theta}}_n$. Czasami spotyka się alternatywną definicję wariancji asymptotycznej estymatora zapisaną właśnie jako $avar(\hat{\boldsymbol{\theta}}_n) = \mathbf{I}(\boldsymbol{\theta})^{-1}/n$. Jednak przy $n \rightarrow \infty$ wyrażenie $\mathbf{I}(\boldsymbol{\theta})^{-1}/n$ dąży do zera (Hayashi, 2000, str. 95). Dlatego tutaj preferowana jest pierwsza definicja wariancji asymptotycznej $avar(\hat{\boldsymbol{\theta}}_n) = \mathbf{I}(\boldsymbol{\theta})^{-1}$.

Przykład 3.11 Normalny model regresji

W homoskedastycznym modelu regresji liniowej $\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{u}$, gdzie $\mathbf{u} \sim N(\mathbf{0}, \sigma^2 \mathbf{I})$, inaczej zwanym normalnym modelem regresji (ang. normal regression), logarytm funkcji wiarygodności można zapisać następującym wzorem:

$$\begin{aligned} l(\boldsymbol{\theta}) = l(\boldsymbol{\beta}, \sigma^2) &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{\mathbf{u}'\mathbf{u}}{2\sigma^2} = \\ &= -\frac{N}{2} \ln(2\pi) - \frac{N}{2} \ln(\sigma^2) - \frac{(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})'(\mathbf{y} - \mathbf{X}\boldsymbol{\beta})}{2\sigma^2} \end{aligned} \quad (3.13)$$

Gradient logarytmu funkcji wiarygodności $\mathbf{s}(\boldsymbol{\theta})$ i macierz informacji Fishera $\mathbf{I}(\boldsymbol{\theta})$ mają w modelu regresji następujące wzory

$$\mathbf{s}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \mathbf{X}'\mathbf{u} \\ -\frac{n}{\sigma^2} + \frac{\mathbf{u}'\mathbf{u}}{2\sigma^2} \end{bmatrix}. \quad (3.14)$$

oraz

$$I(\theta) = \begin{bmatrix} \frac{1}{\sigma^2} E(X'X) & \mathbf{0} \\ \mathbf{0} & 1/(2\sigma^4) \end{bmatrix}. \quad (3.15)$$

Łatwo wykazać korzystając ze wzorów (3.2), (3.3), (3.9) i (3.13), że estymator MNW dla wektora β jest w modelu normalnej regresji liniowej identyczny jak estymator MNK:

$$\hat{\beta}_{MNW} = (X'X)^{-1}X'y. \quad (3.16)$$

Podobnie można pokazać, że estymator MNW wariancji składnika losowego σ^2 jest równy $\hat{\sigma}_{MNW}^2 = \hat{u}'\hat{u}/n$, gdzie $\hat{u} = y - X\hat{\beta}$ to wektor reszt. Jest to estymator obciążony, ale zgodny.

Estymator „wariancji asymptotycznej” $\hat{\beta}_{MNW}$ dany jest wzorem $V_{\hat{\beta}}^{MNW} = \sigma^2 E(X'X)^{-1}$, natomiast estymator wariancji $\hat{\beta}_{MNW}$ to

$$\hat{V}_{\hat{\beta}}^{MNW} = \hat{\sigma}_{MNW}^2 (X'X)^{-1}. \quad (3.17)$$

Przykład 3.12

W tym przykładzie przedstawiamy sposób numerycznego obliczania macierzy informacji na przykładzie modelu regresji liniowej w programie MATLAB. Na podstawie obliczonej macierzy informacji obliczymy błędy szacunku parametrów i porównamy je z analitycznie wyliczonymi błędami szacunku. Zaczniemy od zaprogramowania funkcji obliczającej numerycznie gradient pewnej funkcji.

```
function [g] = Gradient(Funkcja,theta,przyrost)
    F0 = Funkcja(theta); % wartość funkcji w punkcie theta
    k = size(theta, 1); % długość wektora parametrów
    g = zeros(k,1); % definiujemy wektor gradientu
    for i = 1:k % dla każdego elementu wektora parametrów
        theta0 = theta(i); % przechowuje wartość parametru
        theta(i)= theta0 + przyrost; % nowa wartość i-tego parametru
        F1 = Funkcja(theta);
        theta(i)= theta0 - przyrost; % nowa wartość i-tego parametru
        g(i) = (F1 - Funkcja(theta))/(przyrost*2); % pochodna po theta(i)
        theta(i)= theta0; % stara wartość i-tego parametru
    end
end
```

Funkcja zwraca wektor g będący gradientem funkcji Funkcja w punkcie theta. Dokładność przybliżenia numerycznego zależy od parametru przyrost.

Następnie zdefiniujemy funkcję logarytmu wiarygodności dla modelu regresji liniowej podobnie jak w przykładzie 3.3.

```
function logf = log_likelihood(theta,y,X)
    alfa = theta(1:end-1);
    sigma2 = theta(end);
    logf = -0.5*log(2*pi)-0.5*log(sigma2) - ((y - X*alfa).^2)/(2*sigma2);
end
```


Należy zwrócić uwagę, że funkcja ta zwraca wartość logarytmu funkcji wiarygodności $f(y_i|\theta)$ dla każdej obserwacji oddzielnie, to znaczy dla $i = 1, 2, \dots, n$. W przypadku gdy x i y są macierzami zawierającymi wszystkie obserwacje z próby, funkcja `log_likelihood` zwraca wektor `logf` o długości n . Zakładamy, że składniki losowe mają rozkład normalny. Każdą ze zdefiniowanych funkcji `Gradient` i `log_likelihood` warto zapisać w oddzielnym pliku o nazwie identycznej z nazwą funkcji.

W końcu generujemy główny kod programu. Definiujemy symulowane obserwacje x i y oraz parametry `alfa` i `sigma2`. Obliczamy wektor `logf` logarytmu funkcji wiarygodności $f(y_i|\theta)$ dla każdej obserwacji. W pętli obliczamy gradient logarytmu funkcji wiarygodności $f(y_i|\theta)$ dla każdej obserwacji, który jest następnie wykorzystywany do obliczania szacunku macierzy informacji. Po ukończeniu pętli wartość macierzy informacji znajduje się w macierzy `ss`.

```
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
sigma2 = 1; % wariancja składnika losowego równa 1
theta = [alfa; sigma2]; % wektor parametrów
logf = log_likelihood(theta,y,X); % oblicza log.f. wiarygodności
przyrost = 0.00001; % ustalamy przybliżenie numerycznego gradientu
k = length(theta); % liczba parametrów w modelu
ss = zeros(k,k); % wartość startowa iloczynu gradientów
n = length(y); % liczba obserwacji w próbie
for i=1:n % dla każdej obserwacji liczy ...
    f = @(theta) log_likelihood(theta,y(i),X(i,:)); % funkcję log. wiaryg.
    g = Gradient(f,theta,przyrost); % gradient funkcji
    ss = g*g'/n + ss; % kumuluje iloczynny grad.
end
V = inv(ss)/n % wariancja oszacowań parametrów
Sa1fa = sqrt(diag(V(1:3,1:3))) % błąd oszacowań parametrów MNW
Va = sigma2*inv(X'*X) % wariancja oszacowań parametrów (analityczna)
Sa = sqrt(diag(Va)) % błąd oszacowań parametrów MNW
```

Macierz V zawiera oszacowania wariancji estymatorów MNW parametrów θ . Na jej podstawie wyliczono błędy szacunku parametrów `alfa`, czyli `Sa1fa`. Macierz `Va` zawiera wariancję \hat{V}_{β}^{MNV} oszacowań parametrów `alfa`, policzoną na podstawie wzoru (3.17) z przykładu 3.11. Błędy oszacowań parametrów `Sa1fa` i `Sa` powinny być podobne.

Definicja 3.2: Zgodny estymator nazywamy asymptotycznie efektywnym, jeśli jego

asymptotyczna wariancja wynosi $-\left[\lim_{n \rightarrow \infty} E \frac{1}{n} \left(\frac{\partial^2 l(\theta_0)}{\partial \theta \partial \theta'} \right) \right]^{-1}$.

Amemiya (1985, str. 124) definiuje asymptotycznie efektywny estymator jako taki, który spełnia warunek $\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}) \xrightarrow{L} N\left(\mathbf{0}, -\left[\lim_{n \rightarrow \infty} E \frac{1}{n} \left(\frac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}\right)\right]^{-1}\right)$. Estymator MNW spełnia obie te definicje. Newey i McFadden (1994) zauważają, że estymator MNW jest asymptotycznie efektywny między innymi w klasie estymatorów uogólnionej metody momentów.

Własność 3.4: Estymator MNW jest asymptotycznie efektywny w klasie estymatorów uogólnionej metody momentów.

Własność 3.4 oznacza, że różnica macierzy wariancji dowolnego estymatora z tej klasy estymatorów oraz macierzy wariancji estymatora MNW jest macierzą dodatnio półokreśloną. W przypadku estymatorów pojedynczego parametru estymator MNW ma po prostu równą lub mniejszą wariancję niż inny estymator tego parametru w tej klasie estymatorów. Można spotkać twierdzenie, że estymator MNW jest asymptotycznie efektywny w całej klasie estymatorów zgodnych i asymptotycznie normalnych (np. Johnston i DiNardo, 1997, str. 144). Jednak jest to pewne uproszczenie, ponieważ znane są szczególne (i niepraktyczne) przypadki estymatorów w tej klasie, które mają mniejszą wariancję (Amemiya, 1985, str. 124).

Własność 3.5: Estymator MNW jest niezmienniczy względem reparametryzacji (przekształcenia) estymowanych parametrów w tym sensie, że jeśli $\hat{\boldsymbol{\theta}}$ jest estymatorem MNW wektora $\boldsymbol{\theta}$ i $c(\boldsymbol{\theta})$ jest dowolną funkcją, to $c(\hat{\boldsymbol{\theta}})$ jest estymatorem $c(\boldsymbol{\theta})$.

Własność 3.5 jest bardzo praktyczna, ponieważ często trudno jest oszacować bezpośrednio interesujące nas parametry i wolimy szacować ich przekształcenia. Na przykład, kiedy chcemy nałożyć na szacowany parametr ograniczenie dodatniości, to zamiast $\alpha \in \mathbb{R}$ możemy szacować $\beta = \exp(\alpha)$, który zawsze będzie dodatni.

W niektórych podręcznikach dodawane są wymagania odnośnie funkcji $c(\cdot)$. Jeśli funkcja ta jest wzajemnie jednoznaczna (bijekcja), to każdemu estymatorowi $\hat{\boldsymbol{\theta}}$ będzie odpowiadał dokładnie jeden estymator $c(\hat{\boldsymbol{\theta}})$ (i na odwrót). Czasami dodawane jest założenie, że funkcja c jest ciągła i jej pierwsza pochodna jest ciągła po to by spełnione były założenia MNW (np. Greene, 2012, str. 554 i 561).

Własność 3.6: Dla pojedynczej obserwacji y_i gradient logarytmu jej funkcji wiarygodności $\mathbf{s}(y_i, \boldsymbol{\theta}) = \nabla_{\boldsymbol{\theta}} \ln f(y|\boldsymbol{\theta})$ jest zmienną losową i ma wartość oczekiwaną równą wektorowi zerowemu i wariancję $\mathbf{I}(\boldsymbol{\theta})$, czyli $E[\mathbf{s}(y_i, \boldsymbol{\theta})] = \mathbf{0}$ oraz $E[\mathbf{s}(y_i, \boldsymbol{\theta})\mathbf{s}(y_i, \boldsymbol{\theta})'] = \mathbf{I}(\boldsymbol{\theta})$.

Analogicznie gradient logarytmu funkcji wiarygodności dla całej próby, $\mathbf{s}(\mathbf{y}, \boldsymbol{\theta}) = \frac{\partial l}{\partial \boldsymbol{\theta}}$, jest zmienną losową z wartością oczekiwaną równą wektorowi zerowemu i wariancją równą $\mathbf{I}^*(\boldsymbol{\theta})$, czyli $E\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right) = \mathbf{0}$ oraz $E\left[\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right)\left(\frac{\partial l}{\partial \boldsymbol{\theta}}\right)'\right] = \mathbf{I}^*(\boldsymbol{\theta})$. Ponadto z centralnego twierdzenia granicznego Lindeberga-Lévy'ego wynika, że $\sum_{i=1}^n \mathbf{s}(y_i, \boldsymbol{\theta})/\sqrt{n} \xrightarrow{L} N(\mathbf{0}, \mathbf{I}(\boldsymbol{\theta}))$ (np. Hayashi, 2000, str. 474). Te własności są przydatne przy konstrukcji testów statystycznych, w szczególności testu mnożnika Lagrange'a.

3.5. Testowanie restrykcji nałożonych na parametry modelu

W tym punkcie przedstawiono trzy popularne testy statystyczne, służące do weryfikacji hipotez nakładających restrykcje liniowe na parametry w modelach oszacowanych metodą największej wiarygodności. Te testy to test ilorazu wiarygodności (likelihood ratio test; LR), test mnożnika Lagrange'a (Lagrange multiplier test; LM) i test Walda. We wszystkich trzech testach hipoteza zerowa zakłada, że parametry modelu spełniają ogólny warunek $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{0}_{m \times 1}$. Szczególnym przypadkiem jest warunek liniowy $\mathbf{R}\boldsymbol{\theta} - \mathbf{r} = \mathbf{0}_{m \times 1}$. Liczba m wierszy wektora $\mathbf{g}(\boldsymbol{\theta})$ (oraz macierzy \mathbf{R} i \mathbf{r}) oznacza liczbę niezależnych warunków nałożonych na parametry modelu. Hipoteza alternatywna przewiduje, że warunek $\mathbf{g}(\boldsymbol{\theta}) = \mathbf{0}$ nie jest spełniony. Szczegółowe własności testów zostaną tutaj przedstawione w poszczególnych przykładach przy założeniu, że dotyczą liniowego modelu regresji szacowanego MNW, ponieważ w ten sposób łatwiej jest przedstawić ideę testu oraz podobieństwa i różnice między poszczególnymi testami. Rozszerzenia tych trzech testów dla przypadków nieliniowej regresji i innych modeli nieliniowych, szacowanych także przy pomocy innych metod estymacji niż MNW, zostały omówione między innymi w pracach Engle'a (1984) oraz Neweya i McFaddena (1994). Przykłady testów Walda, LM i LR dla modeli nieliniowych omówimy w kolejnych rozdziałach.

Przykład 3.12

W normalnym modelu regresji (por. przykład 3.7) estymator MNW z nałożonymi

restrykcjami $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ można wyprowadzić maksymalizując następujące wyrażenie $\ln(L(\boldsymbol{\beta}^*)) = \ln(L(\boldsymbol{\beta})) - \boldsymbol{\mu}' \cdot (\mathbf{R}\boldsymbol{\beta} - \mathbf{r})$ względem parametrów $\boldsymbol{\beta}$ i $\boldsymbol{\mu}$. Okazuje się, że estymator MNW parametrów modelu regresji liniowej przy nałożonych restrykcjach jest identyczny jak odpowiedni estymator MNK przy warunkach pobocznych (por. wzór (1.15) w punkcie 1.4):

$$\hat{\boldsymbol{\beta}}^* = \hat{\boldsymbol{\beta}} + (\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}'(\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1}(\mathbf{r} - \mathbf{R}\hat{\boldsymbol{\beta}}), \quad (3.18)$$

gdzie $\hat{\boldsymbol{\beta}} = (\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\mathbf{y}$ to estymator MNW bez nałożonych restrykcji.

3.5.1 Test ilorazu wiarygodności

Niech $\lambda = \frac{L(\boldsymbol{\theta}^*)}{L(\boldsymbol{\theta})}$ oznacza iloraz wiarygodności, to znaczy iloraz funkcji wiarygodności modelu z nałożonymi na parametry restrykcjami oraz funkcji wiarygodności dla modelu bez nałożonych restrykcji. Statystyka testowa ma wó

$$LR = -2 \cdot \ln(\lambda) = 2 \cdot [\ln(L(\boldsymbol{\theta})) - \ln(L(\boldsymbol{\theta}^*))]. \quad (3.19)$$

W przypadku, kiedy hipoteza zerowa jest prawdziwa, statystyka LR ma asymptotyczny rozkład $\chi^2(m)$. Warto zwrócić uwagę na fakt, że w celu przeprowadzenia testu wymagane jest oszacowanie parametrów zarówno modelu ogólnego (bez restrykcji) jak i modelu z restrykcjami. W przypadku niektórych modeli nieliniowych oszacowanie parametrów modelu z restrykcjami lub bez nich może okazać się kłopotliwe. Wtedy pomocne okazują się testy mnożnika Lagrange'a lub Walda.

Przykład 3.13.

W normalnym modelu regresji (por. przykład 3.7) statystyka LR służąca do weryfikacji warunku $\mathbf{R}\boldsymbol{\beta} - \mathbf{r} = \mathbf{0}$ ma następujący wzór

$$\begin{aligned} LR &= 2 \cdot [\ln(L(\hat{\boldsymbol{\beta}}, \hat{\sigma})) - \ln(L(\hat{\boldsymbol{\beta}}^*, \hat{\sigma}^*))] = \\ &= 2 \cdot \left[\ln \left(\left(\frac{2\pi e}{n} \right)^{-n/2} \cdot (\hat{\mathbf{u}}'\hat{\mathbf{u}})^{-\frac{n}{2}} \right) - \ln \left(\left(\frac{2\pi e}{n} \right)^{-n/2} \cdot (\hat{\mathbf{u}}^*\hat{\mathbf{u}}^*)^{-\frac{n}{2}} \right) \right] = \\ &= n[\ln(\hat{\mathbf{u}}^*\hat{\mathbf{u}}^*) - \ln(\hat{\mathbf{u}}'\hat{\mathbf{u}})] \approx n \left(\frac{\hat{\mathbf{u}}^*\hat{\mathbf{u}}^* - \hat{\mathbf{u}}'\hat{\mathbf{u}}}{\hat{\mathbf{u}}'\hat{\mathbf{u}}} \right), \end{aligned} \quad (3.20)$$

gdzie $\hat{\mathbf{u}}^*\hat{\mathbf{u}}^*$ jest sumą kwadratów reszt z oszacowanego modelu z restrykcjami $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ jest sumą kwadratów reszt z oszacowanego modelu bez restrykcji.

3.5.2 Test Walda

Niech $I^*(\theta)$ będzie macierzą informacji Fishera dla wektora parametrów θ szacowanych MNW, obliczoną dla całej próby n obserwacji y (por. własność 3.2). Jeśli specyfikacja modelu jest prawidłowa, to macierz $I^*(\theta)^{-1}$ równa jest macierzy $V_{\hat{\theta}}^{MNV}$ macierzy wariancji-kowariancji oszacowań parametrów θ . Statystyka Walda może być zapisana jako następująca forma kwadratowa

$$W = \mathbf{g}(\hat{\theta})' [\hat{\mathbf{G}} \cdot I^*(\theta)^{-1} \cdot \hat{\mathbf{G}}']^{-1} \mathbf{g}(\hat{\theta}). \quad (3.21)$$

gdzie $\hat{\mathbf{G}} = \frac{\partial \mathbf{g}(\hat{\theta})}{\partial \theta}$ to macierz pochodnych cząstkowych funkcji $\mathbf{g}(\theta)$ po parametrach θ policzona w punkcie $\hat{\theta}$. W przypadku restrykcji liniowych, takich jak $\mathbf{R}\theta - \mathbf{r} = \mathbf{0}$, statystykę Walda można zapisać wzorem

$$W = (\mathbf{R}\hat{\theta} - \mathbf{r})' (\mathbf{R}I^*(\theta)^{-1}\mathbf{R}')^{-1} (\mathbf{R}\hat{\theta} - \mathbf{r}) = (\mathbf{R}\hat{\theta} - \mathbf{r})' (\mathbf{R}V_{\hat{\theta}}\mathbf{R}')^{-1} (\mathbf{R}\hat{\theta} - \mathbf{r})$$

Ponieważ zwykle nie jest znana wartość macierzy $V_{\hat{\theta}}$, to wykorzystuje się jej zgodne oszacowanie $\hat{V}_{\hat{\theta}}$

$$W = (\mathbf{R}\hat{\theta} - \mathbf{r})' (\mathbf{R}\hat{V}_{\hat{\theta}}\mathbf{R}')^{-1} (\mathbf{R}\hat{\theta} - \mathbf{r}). \quad (3.22)$$

W przypadku, kiedy hipoteza zerowa jest prawdziwa, statystyka W ma asymptotyczny rozkład $\chi^2(m)$. W celu przeprowadzenia testu Walda wymagane jest – przynajmniej teoretycznie – oszacowanie jedynie parametrów modelu ogólnego (bez restrykcji). Nie trzeba szacować parametrów modelu z restrykcjami, chociaż czasami jest to przydatne (por. przykład 3.13).

Przykład 3.13

W normalnym modelu regresji (por. przykład 3.7) statystyka W służąca do weryfikacji warunku $\mathbf{R}\beta = \mathbf{r}$ ma następujący wzór

$$\begin{aligned} W &= (\mathbf{R}\hat{\beta} - \mathbf{r})' (\mathbf{R}\hat{V}_{\hat{\beta}}\mathbf{R}')^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) = \\ &= (\mathbf{R}\hat{\beta} - \mathbf{r})' (\mathbf{R}\hat{\sigma}^2(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) = \\ &= (\mathbf{R}\hat{\beta} - \mathbf{r})' (\mathbf{R}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{R}')^{-1} (\mathbf{R}\hat{\beta} - \mathbf{r}) / \hat{\sigma}^2 = n \left(\frac{\hat{\mathbf{u}}^{*\prime} \hat{\mathbf{u}}^* - \hat{\mathbf{u}}' \hat{\mathbf{u}}}{\hat{\mathbf{u}}' \hat{\mathbf{u}}} \right), \end{aligned} \quad (3.23)$$

gdzie $\hat{\mathbf{u}}^{*\prime} \hat{\mathbf{u}}^*$ jest sumą kwadratów reszt z oszacowanego modelu z restrykcjami $\mathbf{R}\beta = \mathbf{r}$, $\hat{\mathbf{u}}' \hat{\mathbf{u}}$ jest sumą kwadratów reszt z oszacowanego modelu bez restrykcji, a $\hat{\sigma}^2 = \hat{\mathbf{u}}' \hat{\mathbf{u}} / n$ jest zgodnym estymatorem wariancji składnika losowego w modelu bez restrykcji. Ostatnia równość zachodzi, ponieważ oszacowania MNW z nałożonymi restrykcjami liniowymi są identyczne jak oszacowania MNK przy warunkach pobocznych i w szczególności zastosowanie ma wzór

(3.18).

3.5.3 Test mnożnika Lagrange'a

Test mnożnika Lagrange'a wykorzystuje własności MNW dotyczące gradientu logarytmu funkcji wiarygodności. Pomysł jest następujący. Jeśli restrykcje nałożone na model ekonometryczny są prawdziwe, to nie tylko funkcja wiarygodności w modelu z restrykcjami powinna być podobna do wartości funkcji wiarygodności modelu bez restrykcji, ale także gradienty logarytmów obu funkcji powinny być wektorami podobnych zmiennych losowych, to znaczy z zerową wartością oczekiwaną i wariancją równą macierzy informacyjnej Fishera (por. własność 3.6). Gradient funkcji wiarygodności modelu z restrykcjami powinien mieć też rozkład asymptotycznie normalny. Numerycznie, oszacowany gradient funkcji wiarygodności w modelu bez restrykcji jest zawsze wektorem zerowym $\mathbf{s}(\hat{\boldsymbol{\theta}}) = \mathbf{0}$. Jeśli restrykcje nałożone na parametry modelu są prawdziwe, to także $\mathbf{s}(\hat{\boldsymbol{\theta}}^*) \approx \mathbf{0}$. Dzięki tym podobieństwom można skonstruować statystykę testową jako następującą formę kwadratową

$$LM = \mathbf{s}'(\hat{\boldsymbol{\theta}}^*) \cdot \mathbf{I}^*(\hat{\boldsymbol{\theta}}^*)^{-1} \cdot \mathbf{s}(\hat{\boldsymbol{\theta}}^*). \quad (3.24)$$

Statystyka taka będzie miała asymptotyczny rozkład $\chi^2(m)$, gdzie m oznacza liczbę niezależnych restrykcji nałożonych na parametry modelu. W celu przeprowadzenia tego testu wymagane jest oszacowanie jedynie parametrów modelu z restrykcjami (np. liniowego), który często okazuje się łatwiejszy do szacowania niż model bez restrykcji (np. model nieliniowy będący uogólnieniem modelu liniowego).

Przykład 3.14

W normalnym modelu regresji liniowej gradient logarytmu funkcji wiarygodności względem szacowanych parametrów ma następujący wzór

$$\mathbf{s}(\boldsymbol{\theta}) = \begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \sigma^2} \end{bmatrix} = \begin{bmatrix} \frac{1}{\sigma^2} \mathbf{X}'\mathbf{u} \\ -\frac{n}{\sigma^2} + \frac{\mathbf{u}'\mathbf{u}}{2\sigma^2} \end{bmatrix}. \quad (3.25)$$

Natomiast gradient funkcji wiarygodności przy założeniu prawdziwości restrykcji $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ po oszacowaniu parametrów ma postać

$$\mathbf{s}(\hat{\boldsymbol{\theta}}^*) = \begin{bmatrix} \frac{\partial l}{\partial \boldsymbol{\beta}} \\ \frac{\partial l}{\partial \sigma^2} \end{bmatrix} \Big|_{\boldsymbol{\beta}=\hat{\boldsymbol{\beta}}^*, \sigma^2=\hat{\sigma}^2} = \begin{bmatrix} \frac{1}{\hat{\sigma}^2} \mathbf{X}'\hat{\mathbf{u}}^* \\ 0 \end{bmatrix}. \quad (3.26)$$

Statystykę LM można w takim przypadku zapisać następującym wzorem

$$LM = \begin{bmatrix} \frac{1}{\hat{\sigma}_*^2} \mathbf{X}'\hat{\mathbf{u}}^* \\ 0 \end{bmatrix}' \begin{bmatrix} \hat{\sigma}_*^2 (\mathbf{X}'\mathbf{X})^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\hat{\sigma}_*^4}{n} \end{bmatrix} \begin{bmatrix} \frac{1}{\hat{\sigma}_*^2} \mathbf{X}'\hat{\mathbf{u}}^* \\ 0 \end{bmatrix} = \frac{n\hat{\mathbf{u}}^{*'}\mathbf{X}(\mathbf{X}'\mathbf{X})^{-1}\mathbf{X}'\hat{\mathbf{u}}^*}{\hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^*}.$$

Wzór ten da się przekształcić, podobnie jak w testach Walda i ilorazu wiarygodności, do postaci zależnej od reszt z modeli z restrykcjami i bez restrykcji

$$LM = \frac{n(\hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^* - \hat{\mathbf{u}}'\hat{\mathbf{u}})}{\hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^*}, \quad (3.27)$$

gdzie $\hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^*$ jest sumą kwadratów reszt z oszacowanego modelu z restrykcjami $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$, $\hat{\mathbf{u}}'\hat{\mathbf{u}}$ jest sumą kwadratów reszt z oszacowanego modelu bez restrykcji, a $\hat{\sigma}_*^2 = \hat{\mathbf{u}}^{*'}\hat{\mathbf{u}}^*/n$ jest zgodnym estymatorem wariancji składnika losowego w modelu z restrykcjami.

Interesujące jest spostrzeżenie, że w przypadku restrykcji $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ nałożonej na parametry modelu regresji liniowej wszystkie trzy omówione tutaj statystyki, czyli statystyka Walda, ilorazu wiarygodności i mnożnika Lagrange'a, mają podobne – ale nie identyczne – wzory. Ponadto, kiedy restrykcja $\mathbf{R}\boldsymbol{\beta} = \mathbf{r}$ jest prawdziwa, to wszystkie trzy statystyki mają ten sam asymptotyczny rozkład $\chi^2(m)$.

Przykład 3.15

Ten przykład jest kontynuacją przykładów 3.9 i 3.10. Przedstawiamy tutaj zastosowanie testu Walda, testu LM i testu LR do zbadania hipotezy, że w modelu logitowym skonstruowanym w przykładzie 3.10 suma parametrów przy zmiennych objaśniających równa jest 6, to znaczy $[1 \ 1 \ 1] \cdot \boldsymbol{\beta} = 6$. W tym celu definiujemy funkcję służącą do szacowania macierzy informacji $\mathbf{I}^*(\boldsymbol{\beta})$.

```
function [I] = I_logit(y,X,param) %funkcja zwraca mac. informacji modelu logit
    p = exp(X*param)./(1 + exp(X*param)); % wektor P(y=1|...)
    lnf = log(y.*p + (1-y).*(1-p)); % wektor f(y|..) dla każdej obserwacji
    n = length(lnf); % liczba obserwacji
    k = length(param); % liczba parametrów modelu
    przyrost = 0.0001; % dh do liczenia gradientu
    I = zeros(k,k); % startowa wartość macierzy informacji
    for i=1:n
        lnf = @(param) l_logit(y(i),X(i,:),param); % log. f. wiar. dla obs. i
        g = Gradient(lnf,param,przyrost); % gradient log. f. wiar. dla obs. i
        I = g*g' + I; % szacuje macierz informacji I*
    end
end
```

Ta funkcja powinna zostać zapisana w oddzielnym pliku o nazwie `I_logit.m`. Funkcja `Gradient`, wykorzystywana tutaj do obliczania gradientu logarytmu funkcji wiarygodności

dla i -tej obserwacji, została zdefiniowana w przykładzie 3.12. Macierz informacji $I^*(.)$ jest wykorzystywana do konstrukcji testów Walda i LM.

Kod z przykładu 3.10 należy uzupełnić o następujące polecenia w celu oszacowania modelu z restrykcjami nałożonymi na parametry.

```
Aeq = [1 1 1]; % restrykcje na parametry
beq = 6; % suma parametrów = 6
[beta_res,res_1] = fmincon(f,beta,[],[],Aeq,beq) % min. fun. f z restrykcją
```

Po oszacowaniu wektor `beta_res` zawiera oszacowania parametrów z restrykcjami, a liczba `res_1` jest równa logarytmowi funkcji wiarygodności dla tych oszacowań. Statystyki Walda, LM i LR mogą być policzone przy pomocy następujących poleceń.

```
g = Aeq*beta_est - beq; % zapis restrykcji do testu
G = Aeq; % pochodna z g() po beta
V = inv(I_logit(y,X,beta_est)) % wariancja oszacowania bety
Wald = g'*inv(G*V*G')*g % statystyka Walda
przyrost = 0.0001; % dh do liczenia gradientu
s = -Gradient(f,beta_res,przyrost) % gradient z log. f. wiarygodności
Vres = inv(I_logit(y,X,beta_res)) % war. oszac. bety z restrykcją
LM = s'*Vres*s % statystyka LM
LR = 2*(-max_l+res_1) % statystyka LR
p_Wald = 1-chi2cdf(Wald,1) % p-value statystyki Walda
p_LM = 1-chi2cdf(LM,1) % p-value statystyki LM
p_LR = 1-chi2cdf(LR,1) % p-value statystyki LR
```

Wynikowe wartości p dla trzech statystyk powinny być w tym przykładzie dalekie od zera.

3.6. Nieliniowa metoda najmniejszych kwadratów

Model nieliniowy względem parametrów i względem zmiennych często przyjmuje formę równania z nieliniową funkcją zmiennych objaśniających $h(x)$ i z addytywnym składnikiem losowym ε . Na przykład w modelu

$$y = h(x) + \varepsilon = \alpha + \beta \frac{\exp(x)^{\gamma}-1}{\gamma} + \varepsilon, \quad (3.28)$$

α , β i γ to parametry modelu, a y i x to zmienne objaśniana i objaśniająca. Składnik losowy jest dodawany do funkcji $h(x)$, co oznacza, że odchylenia rzeczywistych wartości y od wartości wskazywanych (prognozowanych) przez model mają charakter addytywny.

Przypuśćmy, że chcemy oszacować wartości parametrów takiego nieliniowego modelu ekonometrycznego. W tym przypadku możemy (przy spełnionych pewnych warunkach teoretycznych, omówionych w punkcie 3.7) do oszacowania parametrów wykorzystać nieliniową metodę najmniejszych kwadratów (NMNK).

Estymator NMNK minimalizuje sumę kwadratów odchylenia obserwacji y_i od $h(\mathbf{x}_i; \boldsymbol{\beta})$:

$$Q_n(\boldsymbol{\beta}) = \frac{1}{n} \sum_{i=1}^n [y_i - h(\mathbf{x}_i; \boldsymbol{\beta})]^2, \quad (3.29)$$

czyli

$$\hat{\boldsymbol{\beta}}_{NMNK} = \underset{\boldsymbol{\beta}}{\operatorname{argmin}} Q_n(\boldsymbol{\beta}). \quad (3.30)$$

Podobnie jak klasyczna MNK, NMNK polega na minimalizowaniu kwadratów reszt modelu poprzez wyznaczenie optymalnych wartości jego parametrów. Reszty reprezentują tę część „zmienności” zmiennej objaśnianej, której model nie jest w stanie wyjaśnić. Okazuje się, że estymator tak zdefiniowany jak we wzorze (3.30) ma szereg własności, które są pożądane przy szacowaniu wartości parametrów modelu. Własności te będą omówione w punkcie 3.7.

Warto zauważyć, że minimum funkcji $Q_n(\boldsymbol{\beta})$ zostaje osiągnięte przy spełnieniu warunku pierwszego rzędu:

$$\nabla_{\boldsymbol{\beta}} Q_n(\boldsymbol{\beta}) = -\frac{2}{n} \nabla_{\boldsymbol{\beta}} h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta}) [\mathbf{y} - h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta})] = \mathbf{0}_{k \times 1},$$

gdzie

$$\nabla_{\boldsymbol{\beta}} h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta}) = [\nabla_{\boldsymbol{\beta}} h(\mathbf{x}_1; \boldsymbol{\beta}) \quad \nabla_{\boldsymbol{\beta}} h(\mathbf{x}_2; \boldsymbol{\beta}) \quad \dots \quad \nabla_{\boldsymbol{\beta}} h(\mathbf{x}_n; \boldsymbol{\beta})]_{k \times n}$$

oraz

$$h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta}) = [h(\mathbf{x}_1; \boldsymbol{\beta}) \quad h(\mathbf{x}_2; \boldsymbol{\beta}) \quad \dots \quad h(\mathbf{x}_n; \boldsymbol{\beta})]'$$

Zwykle postać modelu nieliniowego jest na tyle skomplikowana, że nie da się przedstawić bezpośrednio wzoru na estymator parametrów $\boldsymbol{\beta}$ takiego modelu. Potrzebne jest wtedy zastosowanie metod numerycznego poszukiwania minimum funkcji $Q_n(\boldsymbol{\beta})$. Najpopularniejsze metody optymalizacyjne wykorzystywane w tym celu zostały omówione w rozdziale 4.

Przykład 3.16

W tym przykładzie przedstawiamy sposób wyliczenia i zapisania wyrażenia $h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta})$ i $\nabla_{\boldsymbol{\beta}} h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta})$ w postaci macierzowej w programie MATLAB. Przyjmujemy, że prawdziwy jest model zadany wzorem (3.28) z parametrami $\alpha = 1$, $\beta = 2$ i $\gamma = 3$, natomiast wartości x są losowane niezależnie z rozkładu $N(0,1)$.

Definiujemy funkcję funkcja_h, która zwraca wartość $h(x)$, i zapisujemy ją w oddzielnym pliku o nazwie funkcja_h.m.

```

function h = funkcja_h(x,param)
    alfa = param(1);           % przepisanie parametrów z wektora
    beta = param(2);
    gamma = param(3);
    h = alfa + beta*(exp(x).^gamma-1)/gamma; % wektor wartości funkcji h(x)
end

```

Następnie w wektorze h zapisujemy wartości $h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta})$ oraz w macierzy dh zapisujemy wartości $\nabla_{\boldsymbol{\beta}} h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \boldsymbol{\beta})$.

```

x = randn(100,1);           % losowane wartości x z N(0,1)
alfa = 1;                   % parametry modelu
beta = 2;
gamma = 3;
param = [ 1; 2; 3];         % przepisanie alfa, beta i gamma do wektora
h = zeros(100,1);          % startowe wartości wektora h
dh = zeros(3,100);         % startowe wartości macierzy dh
przyrost = 0.0001;         % parametr dokładności wyliczania gradientu
for i=1:100
    fh = @(param) funkcja_h(x(i),param); % funkcja anonimowa
    h(i,1) = fh(param);           % przypisuje elementy wektora h
    dh(:,i) = Gradient(fh,param,przyrost); % przypisuje elementy macierzy dh
end

```

Funkcja Gradient została zdefiniowana w przykładzie 3.12.

3.7. Założenia i własności nieliniowej metody najmniejszych kwadratów

Założenia i własności NMNK podajemy za Hayashim (2000, rozdz. 7). Podobny opis przedstawił między innymi Amemiya (1985, str. 127-135). Natomiast Greene (2012, str. 222-229) prezentuje uproszczony (i przez to bardziej intuicyjny) opis założeń NMNK, nawiązujący do założeń MNK dla modeli liniowych.

Założenie 3.10: Obserwacje $\{(y_1, \mathbf{x}_1), (y_2, \mathbf{x}_2), \dots, (y_n, \mathbf{x}_n)\}$ są losowe. Mogą być generowane przez procesy stochastyczne, a \mathbf{x}_i może zawierać między innymi opóźnione wartości y_i (np. y_{i-1}).

To założenie jest analogiczne do założenia 1.1 z rozdziału 1 dla modeli liniowych. Założenie o losowości zmiennych odpowiada przekonaniu ekonomistów, że zmienne ekonomiczne i finansowe (takie jak wzrost gospodarczy, czy ceny akcji na giełdzie) zwykle generowane są przynajmniej częściowo w sposób losowy. Amemiya (1985, str. 127-135) zakłada, że zmienne objaśniające są deterministyczne, co upraszcza wnioskowanie statystyczne. Newey i McFadden (1994) zakładają niezależność obserwacji (y_i, \mathbf{x}_i) generowanych z jednakowego rozkładu, podobnie jak w założeniu 3.1, dotyczącym własności MNW. Hayashi (2000, rozdz.

7) z kolei przedstawia własności NMNK i MNW przy założeniu, że istnieje proces stochastyczny $\{y_i, \mathbf{x}_i\}$ który generuje dane do modelu. Przy dowodzeniu zgodności estymatora NMNK zakłada on też, że proces ten jest stacjonarny i ergodyczny.

Założenie 3.11: Warunkowa wartość oczekiwana y_i względem \mathbf{x}_i , $E(y_i|\mathbf{x}_i)$, należy do parametrycznej klasy funkcji $\varphi(\mathbf{x}_i; \boldsymbol{\beta})$, $\boldsymbol{\beta} \in \mathbf{B}$. Dla prawdziwych wartości $\boldsymbol{\beta}_0$ zachodzi $E(y_i|\mathbf{x}_i) = \varphi(\mathbf{x}_i; \boldsymbol{\beta}_0)$. Dla błędów zdefiniowanych jako $\varepsilon_i \equiv y_i - E(y_i|\mathbf{x}_i)$ zachodzi równość $y_i = \varphi(\mathbf{x}_i; \boldsymbol{\beta}_0) + \varepsilon_i$ oraz $E(\varepsilon_i|\mathbf{x}_i) = 0$, $\boldsymbol{\beta}_0 \in \mathbf{B}$.

Podobnie jak w założeniu 1.2 przedstawionym w rozdziale 1 wartość oczekiwana składnika losowego jest równa zero niezależnie od wartości zmiennych objaśniających. To oznacza także, że $E(\varepsilon_i) = 0$ oraz $E(\varepsilon_i|h(\mathbf{x}_i; \boldsymbol{\beta})) = 0$. Nie jest natomiast wymagana niezależność składnika losowego od zmiennych objaśniających. W odróżnieniu do modeli liniowych, w modelach regresji nieliniowej liczba zmiennych objaśniających nie musi być równa liczbie parametrów w wektorze $\boldsymbol{\beta}$.

Założenie 3.12 (identyfikowalność parametrów): Dla wszystkich $\boldsymbol{\beta} \neq \boldsymbol{\beta}_0$ zachodzi $\varphi(\mathbf{x}_i; \boldsymbol{\beta}_0) \neq \varphi(\mathbf{x}_i; \boldsymbol{\beta})$ dla każdej obserwacji \mathbf{x}_i .

Założenie 3.12 jest warunkiem dla istnienia optymalnych wartości parametrów, $\boldsymbol{\beta}_0$, które minimalizują sumę kwadratów reszt w równaniu (3.29).

Założenie 3.13: Zbiór $\mathbf{B} \subset \mathbb{R}^p$ parametrów $\boldsymbol{\beta}_0$ modelu jest zwarty.

Założenie 3.14: Funkcja $m(y_i, \mathbf{x}_i; \boldsymbol{\beta}) = (y_i - h(\mathbf{x}_i; \boldsymbol{\beta}))^2$ jest ciągła dla wszystkich obserwacji (y_i, \mathbf{x}_i) ; $m(y_i, \mathbf{x}_i; \boldsymbol{\beta})$ jest mierzalna w obserwacji (y_i, \mathbf{x}_i) dla wszystkich $\boldsymbol{\beta} \in \mathbf{B}$.

Założenie 3.15: $E \left(\sup_{\boldsymbol{\beta} \in \mathbf{B}} |m(y_i, \mathbf{x}_i; \boldsymbol{\beta})| \right) < \infty$.

Założenia 3.12 – 3.15 są potrzebne do udowodnienia, że estymator NMNK jest zgodny. Ponieważ założenie 3.13 często nie jest spełnione (np. przestrzeń \mathbb{R} nie jest zwarta), to zarówno dla metody estymacji MNW jak i NMNK zaproponowano też alternatywny zestaw założeń (np. Hayashi, 2000, str. 460). Ten alternatywny zestaw założeń zamiast zwartości zbioru \mathbf{B} wymaga, by prawdziwy parametr $\boldsymbol{\beta}_0$ znajdował się we wnętrzu wypukłej przestrzeni

parametrów \mathbf{B} i funkcja $m(y_i, \mathbf{x}_i; \boldsymbol{\beta})$ była wypukła dla wszystkich obserwacji (y_i, \mathbf{x}_i) . Zamiast założenia 3.15 wymagany jest też wtedy prostszy warunek $E(|m(y_i, \mathbf{x}_i; \boldsymbol{\beta})|) < \infty$ – istnieje skończona wartość oczekiwana $m(y_i, \mathbf{x}_i; \boldsymbol{\beta})$ dla wszystkich $\boldsymbol{\beta} \in \mathbf{B}$. Jest to podobne założenie do założenia 1.3 z rozdziału 1, które nakłada ograniczenia na momenty zmiennych objaśniających modelu.

Założenie 3.16: Wektor $\boldsymbol{\beta}_0$ jest punktem wewnętrznym zbioru \mathbf{B} .

Założenie 3.17: Funkcja $m(y_i, \mathbf{x}_i; \boldsymbol{\beta})$ jest dwukrotnie różniczkowalna względem $\boldsymbol{\beta}$ i jej druga pochodna jest ciągła dla każdej obserwacji (y_i, \mathbf{x}_i) .

Założenie 3.18: $\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{s}(y_i, \mathbf{x}_i; \boldsymbol{\beta}_0) \xrightarrow{L} N(\mathbf{0}, \boldsymbol{\Sigma})$, gdzie $\boldsymbol{\Sigma}$ jest macierzą dodatnio określoną oraz $\mathbf{s}(y_i, \mathbf{x}_i; \boldsymbol{\beta}) = \frac{\partial m(y_i, \mathbf{x}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta}}$.

Założenie 3.19: W pewnym otoczeniu $N(\boldsymbol{\beta}_0)$ zachodzi

$E \left(\sup_{\boldsymbol{\beta} \in N(\boldsymbol{\beta}_0)} \|\mathbf{H}(y_i, \mathbf{x}_i; \boldsymbol{\beta})\| dy \right) < \infty$, gdzie $\|A\|$ oznacza pierwiastek sumy kwadratów wszystkich elementów A . Dzięki temu dla każdego zgodnego estymatora $\hat{\boldsymbol{\beta}}_n$ zachodzi $\frac{1}{n} \sum_{i=1}^n \mathbf{H}(y_i, \mathbf{x}_i; \hat{\boldsymbol{\beta}}_n) \xrightarrow{p} E(\mathbf{H}(y_i, \mathbf{x}_i; \boldsymbol{\beta}_0))$, gdzie $\mathbf{H}(y_i, \mathbf{x}_i; \boldsymbol{\beta}) = \frac{\partial^2 m(y_i, \mathbf{x}_i; \boldsymbol{\beta})}{\partial \boldsymbol{\beta} \partial \boldsymbol{\beta}'}$.

Założenie 3.20: Macierz $E(\mathbf{H}(y_i, \mathbf{x}_i; \boldsymbol{\beta}_0))$ jest nieosobliwa.

Założenia 3.16 – 3.20 są wymagane, by zgodny estymator NMNK, $\hat{\boldsymbol{\beta}}_n$, był asymptotycznie normalny w tym sensie, że wyrażenie $\sqrt{n}\hat{\boldsymbol{\beta}}_n$ zbiegałoby z dystrybuantą do rozkładu normalnego przy $n \rightarrow \infty$.

Założenie 3.17 oznacza między innymi, że nieliniowa funkcja $h(\mathbf{x}_i; \boldsymbol{\beta})$ powinna być podwójnie różniczkowalna z ciągłą drugą pochodną względem $\boldsymbol{\beta}$ dla każdej obserwacji. Założenia 3.17 i 3.20 wspólnie odpowiadają założeniu 1.4 z rozdziału pierwszego, ponieważ dla modelu liniowego $E(\mathbf{H}(y_i, \mathbf{x}_i; \boldsymbol{\beta}_0)) = E(2\mathbf{x}_i' \mathbf{x}_i)$. Jeśli można policzyć drugie pochodne $h(\mathbf{x}_i; \boldsymbol{\beta})$ względem parametrów i dla danych obserwacji \mathbf{x}_i i y_i oraz macierz $\nabla_{\boldsymbol{\beta}}^2 Q_n(\boldsymbol{\beta})$ jest dodatnio określona, to oznacza, że można wyznaczyć minimum lokalne funkcji $Q_n(\boldsymbol{\beta})$, czyli minimalną sumę kwadratów reszt. Możliwe jest wiele minimów lokalnych tej funkcji i parametry niekoniecznie są jednoznacznie identyfikowalne (por. założenie 3.12). Tutaj także widać analogię do założenia 1.4 dla modeli liniowych, ponieważ:

$$\nabla_{\beta}^2 Q_N(\beta) = -\frac{2}{N} \nabla_{\beta}^2 h(\mathbf{X}, \beta) [y - h(\mathbf{X}, \beta)] + \frac{2}{N} [\nabla_{\beta} h(\mathbf{X}, \beta)] [\nabla_{\beta} h(\mathbf{X}, \beta)]'$$

oraz dla modelu liniowego $\nabla_{\beta}^2 Q_N(\beta) = \frac{2}{N} (\mathbf{X}'\mathbf{X})$, która jest odwracalna jeśli $\text{rz}(\mathbf{X}'\mathbf{X}) = k$.

Własność 3.7: Estymator NMNK jest zgodny, to znaczy $\text{plim}_{n \rightarrow \infty} \hat{\beta}_n = \beta$. Ciąg wartości estymatora $\hat{\theta}_n$ zbiega z prawdopodobieństwem do prawdziwej wartości parametru θ przy $n \rightarrow \infty$.

Własność 3.8: Estymator NMNK jest asymptotycznie normalny, to znaczy odpowiednio przeskalowany estymator NMNK ma asymptotyczny rozkład normalny, czyli

$$\sqrt{n}(\hat{\beta}_n - \beta) \xrightarrow{L} N(0, \mathbf{V}_{\hat{\beta}}),$$

$$\text{gdzie } \mathbf{V}_{\hat{\beta}} = \left(E(\mathbf{H}(y_i, \mathbf{x}_i; \beta_0)) \right)^{-1} \boldsymbol{\Sigma} \left(E(\mathbf{H}(y_i, \mathbf{x}_i; \beta_0)) \right)^{-1}.$$

W ogólnym przypadku wariancję $\hat{\beta}_n$ możemy oszacować wykorzystując zgodne oszacowanie $E(\mathbf{H}(y_i, \mathbf{x}_i; \beta_0))$, czyli $\frac{1}{n} \sum_{i=1}^n \mathbf{H}(y_i, \mathbf{x}_i; \hat{\beta}_n)$, oraz odpowiednie oszacowanie $\boldsymbol{\Sigma}$. Ponieważ z założenia 3.18 $\mathbf{s}(y_i, \mathbf{x}_i; \beta) = \frac{\partial m(y_i, \mathbf{x}_i; \beta)}{\partial \beta} = -2 \cdot \nabla_{\beta} h(\mathbf{x}_i; \beta) \cdot (y_i - h(\mathbf{x}_i; \beta))$ oraz $\text{avar} \left(\frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbf{s}(y_i, \mathbf{x}_i; \beta_0) \right) = \boldsymbol{\Sigma}$, to przy założeniu braku autokorelacji składnika losowego $\boldsymbol{\Sigma}$ można szacować wykorzystując wzór

$$\hat{\boldsymbol{\Sigma}} = \frac{4}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)] \cdot [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)]'. \quad (3.31)$$

Jako estymatora $E(\mathbf{H}(y_i, \mathbf{x}_i; \beta_0))$ wykorzystuje się wtedy często

$$\hat{\mathbf{H}} = \frac{2}{n} \sum_{i=1}^n [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)] \cdot [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)]', \quad (3.32)$$

a przy możliwej błędnej specyfikacji warunkowej wartości oczekiwanej $E(y_i | \mathbf{x}_i)$

$$\hat{\mathbf{H}}^* = \frac{2}{n} \sum_{i=1}^n [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)] \cdot [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)]' + \frac{2}{n} \sum_{i=1}^n \mathbf{H}(y_i, \mathbf{x}_i; \hat{\beta}_n) (y_i - h(\mathbf{x}_i; \hat{\beta}_n)),$$

Estymator wariancji asymptotycznej $\hat{\beta}_n$ ma następujący wzór

$$\widehat{\text{avar}}(\hat{\beta}_n) = \hat{\mathbf{H}}^{-1} \hat{\boldsymbol{\Sigma}} \hat{\mathbf{H}}^{-1} = \left(\frac{2}{n} \sum_{i=1}^n [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)] \cdot [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)]' \right)^{-1} \cdot \left(\frac{4}{n} \sum_{i=1}^n \hat{\varepsilon}_i^2 [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)] \cdot [\nabla_{\beta} h(\mathbf{x}_i; \hat{\beta}_n)]' \right).$$

$$\left(\frac{2}{n}\sum_{i=1}^n[\nabla_{\beta}h(x_i;\hat{\beta}_n)] \cdot [\nabla_{\beta}h(x_i;\hat{\beta}_n)]'\right)^{-1}. \quad (3.33)$$

W idealnym przypadku, w którym model nieliniowy jest homoskedastyczny, to znaczy $E(\varepsilon_i^2 | \mathbf{x}_i) = \sigma^2 < \infty$, otrzymujemy wzór analogiczny jak dla homoskedastycznych modeli liniowych

$$\widehat{avar}(\hat{\beta}_n) = \hat{\sigma}^2 \left(\frac{1}{n}\sum_{i=1}^n[\nabla_{\beta}h(x_i;\hat{\beta}_n)] \cdot [\nabla_{\beta}h(x_i;\hat{\beta}_n)]'\right)^{-1}, \quad (3.34)$$

gdzie $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^n \hat{\varepsilon}_i^2$.

Jak już zauważyliśmy przy omawianiu własności asymptotycznych estymatorów MNK i MNW (por. dyskusję własności 1.7 w punkcie 1.2 oraz własności 3.3 w punkcie 3.4) wariancja asymptotyczna dla estymatora $\hat{\beta}_n$ jest w rzeczywistości graniczną wariancją dla wyrażenia $\sqrt{n}\hat{\beta}_n$ przy $n \rightarrow \infty$. Dlatego przybliżeniem wariancji estymatora $\hat{\beta}_n$ w skończonej próbie jest

$$\hat{V}_{\hat{\beta}} = \hat{H}^{-1}\hat{\Sigma}\hat{H}^{-1}/n \quad (3.35)$$

oraz dla modeli homoskedastycznych

$$\hat{V}_{\hat{\beta}} = \hat{\sigma}^2 \left(\sum_{i=1}^n[\nabla_{\beta}h(x_i;\hat{\beta}_n)] \cdot [\nabla_{\beta}h(x_i;\hat{\beta}_n)]'\right)^{-1}. \quad (3.36)$$

Czasami to właśnie to przybliżenie jest nazywane wariancją asymptotyczną estymatora $\hat{\beta}_n$.

Przykład 3.17

W tym przykładzie przedstawiamy sposób oszacowania modelu typu Cobba-Douglasa z addytywnym składnikiem losowym, $y = x_1^\alpha x_2^\beta e^\gamma + \varepsilon_i$, przy pomocy NMNK.

Definiujemy funkcję funkcja_CD, która zwraca wartość $h(\mathbf{X}, \boldsymbol{\beta})$, i zapisujemy ją w oddzielnym pliku o nazwie funkcja_CD.m.

```
function h = funkcja_CD(x,param)
    alfa = param(1);           % przepisanie parametrów z wektora
    beta = param(2);
    gamma = param(3);
    h = x(:,1).^alfa.*x(:,2).^beta.*exp(gamma); % wektor wartości funkcji h(x)
end
```

Definiujemy także ogólną funkcję reszty, która zwraca reszty z dowolnego modelu opisanego pewną funkcją Funkcja (na przykład funkcją funkcja_CD). Zapisujemy kod funkcji w oddzielnym pliku o nazwie reszty.m.

```
function res = reszty(Funkcja,y,X,param)
    res = y - Funkcja(X,param); % liczy reszty z modelu
end
```

W głównym kodzie programu przygotowujemy losowe dane do obliczeń i przypisujemy wartości parametrom $\alpha = 0,6$, $\beta = 0,4$ i $\gamma = 1$.

```
X = 3*abs(randn(100,2));           % wartości bezwzględne x1 i x2 z N(0,3)
alfa = 0.6;                       % parametry modelu
beta = 0.4;
gamma = 1;
param0 = [ 0.6; 0.4; 1];          % przepisanie alfa, beta i gamma do wektora
epsilon = randn(100,1);
y = X(:,1).^alfa.*X(:,2).^beta*exp(gamma) + epsilon; % generowanie y
```

W kolejnym kroku udajemy, że nie znamy wartości parametrów alfa, beta i gamma i je szacujemy przy pomocy NMNK. Za minimalizację sumy kwadratów reszt odpowiada w MATLABie funkcja `lsqnonlin`.

```
m = @(param) reszty(@funkcja_CD,y,X,param); % funkcja zwracająca reszty
[param_est,Q,e] = lsqnonlin(m,param0);    % szuka min. sumy kwadratów reszt
param_est
```

Interesują nas także błędy szacunku parametrów. Dlatego wyliczamy $\nabla_{\beta} h(\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_n; \hat{\beta}_n)$ w postaci macierzowej w programie MATLAB podobnie jak w przykładzie 3.16.

```
h = zeros(100,1);                 % startowe wartości wektora h
dh = zeros(3,100);                % startowe wartości macierzy dh
przyrost = 0.0001;                 % parametr dokładności wyliczania gradientu
for i=1:100
    fh = @(param) funkcja_CD(X(i,:),param); % funkcja anonimowa
    dh(:,i) = Gradient(fh,param_est,przyrost); % przypisuje elem. macierzy dh
end
```

Funkcja `Gradient` została zdefiniowana w przykładzie 3.12. Wykorzystujemy obliczoną macierz `dh` do wyznaczenia macierzy $\hat{V}_{\hat{\beta}}$ na podstawie wzoru (3.36). Następnie wyliczamy średnie błędy szacunku dla trzech oszacowanych parametrów.

```
sigma2 = e'*e/100;                 % szacunek wariancji składnika losowego
V = sigma2*inv(dh*dh');            % szacunek wariancji est. parametrów
S_param = sqrt(diag(V))            % średnie błędy szacunku parametrów
```

Literatura

Amemiya, T. (1985) *Advanced Econometrics*, Harvard University Press.

Gourieroux, C., Monfort, A., Trognon A. (1984) *Pseudo Maximum Likelihood Methods: Theory*, *Econometrica*, 52, 681-700.

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

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

Heijmans, R. D., & Magnus, J. R. (1986). Consistent maximum-likelihood estimation with dependent observations: the general (non-normal) case and the normal case, *Journal of Econometrics*, 32(2), 253-285.

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

Newey, W., McFadden, D. (1994) *Large Sample Estimation and Hypothesis Testing*, rozdz. 36, w: red. R. Engle, D. McFadden, *Handbook of Econometrics*, 4, North Holland.