

10. Metoda Największej Wiarygodności

Jedną z najpopularniejszych metod estymacji parametrów poza MNK jest **Metoda Największej Wiarygodności**. Metoda ta wykorzystuje do znalezienia estymatorów parametrów funkcję wiarygodności dla analizowanej próby. Estymatorem MNW są takie wielkości parametrów, dla których funkcja wiarygodności dla analizowanej próby przyjmuje maksimum. W tym rozdziale będziemy analizować własności estymatorów MNW oraz opiszemy kilka najpopularniejszych zastosowań tej metody w ekonometrii.

10.1. Własności metody największej wiarygodności

Żałómy, że mamy próbkę złożoną z n niezależnych obserwacji pochodzących z warunkowego rozkładu o gęstości $f_\theta(\mathbf{y}_i | \mathbf{x}_i)$ i założmy, że postać funkcji $f_\theta(\cdot)$ nie zależy od n a jedynie od $\theta \in \Theta \subset R^k$. Ponieważ \mathbf{y}_i jest losowa, więc wartość funkcji $f_\theta(\mathbf{y}_i | \mathbf{x}_i)$ jest także losowa. Przy tych założeniach łączna funkcja gęstości dla zaobserwowanej próby, nazywana dalej funkcją wiarygodności jest dana wzorem:

$$L(\theta, \mathbf{y} | \mathbf{x}) = f_\theta(\mathbf{y} | \mathbf{x}) = \prod_{i=1}^n f_\theta(\mathbf{y}_i | \mathbf{x}_i),$$

gdzie $\mathbf{y} = (\mathbf{y}_1, \dots, \mathbf{y}_n)$, $\mathbf{x} = (\mathbf{x}_1, \dots, \mathbf{x}_n)$ i $\mathbf{y} \in \mathbf{Y}$. Logarytm funkcji wiarygodności ma następującą postać

$$\ell(\theta, \mathbf{y} | \mathbf{x}) = \ln L(\theta, \mathbf{y} | \mathbf{x}) = \sum_{i=1}^n \ln f_\theta(\mathbf{y}_i | \mathbf{x}_i) = \sum_{i=1}^n \ell_i(\theta, \mathbf{y}_i | \mathbf{x}_i)$$

Najczęściej będziemy pomijać \mathbf{y} i \mathbf{x} oraz \mathbf{y}_i i \mathbf{x}_i i pisać $\ell(\theta)$, $L(\theta)$ i $\ell_i(\theta)$. Estymatorem MNW jest takie θ dla której funkcja wiarygodności ma największą wartość $\hat{\theta} = \arg \max \ell(\theta)$. W dal-

szych rozważaniach będziemy używać oznaczeń $\frac{\partial \ell_i(\tilde{\theta})}{\partial \theta'} = \frac{\partial \ell_i(\theta)}{\partial \theta'} \Big|_{\tilde{\theta}}$ i $\frac{\partial^2 \ell_i(\tilde{\theta})}{\partial \theta \partial \theta'} = \frac{\partial^2 \ell_i(\theta)}{\partial \theta \partial \theta'} \Big|_{\tilde{\theta}}$.

10.2. Identyfikowalność parametrów

Dla dowolnego wektora parametrów $\theta \in \Theta$, możemy oszacować na podstawie obserwacji zmiennej losowej \mathbf{y} tylko wtedy, gdy obserwacje te są generowane w taki sposób, że łączna funkcja gęstości \mathbf{y} , ma dla pewnego $\|\varepsilon\| > 0$ następującą własność

$$L(\theta, \mathbf{y} | \mathbf{X}) \neq L(\theta + \varepsilon, \mathbf{y} | \mathbf{X}), \quad (10.1)$$

dla pewnej próby (\mathbf{y}, \mathbf{X}) . Gdy (10.1) jest spełniona, dla $\|\varepsilon\| < \delta$ i pewnego $\delta > 0$ mówimy o lokalnej identyfikowalności. Jeśli jest spełniona dla wszystkich ε , to mówimy o globalnej identyfikowalności. Do wyprowadzenia własności metody największej wiarygodności wystarczy lokalna identyfikowalność parametrów.

Gdy warunek (10.1) nie jest spełniony dla pewnej $\hat{\theta} = \theta + \varepsilon$, wtedy nie ma możliwości rozróżnienia na podstawie obserwacji, czy prawdziwą funkcją wiarygodności \mathbf{y} jest $L(\theta, \mathbf{y} | \mathbf{x})$ czy też $L(\hat{\theta}, \mathbf{y} | \mathbf{x})$, ponieważ obie z nich generują identyczne prawdopodobieństwa \mathbf{y} . Regułą jest, że jedynym sposobem, dzięki któremu można zidentyfikować parametry, jest sięgnięcie po informacje spoza próby.

W kontekście *MNW* przyczyną, dla której identyfikacja parametrów jest, że w przypadku, kiedy 10.1 nie jest spełnione niemożliwe jest także znalezienie unikalnego $\tilde{\theta}$ maksymalizującego funkcję wiarygodności. Dla podwójnie różniczkowalnej funkcji $\ell(\theta)$ ujemna określoność macierzy drugich pochodnych jest wystarczająca do znalezienia takiego lokalnego maksimum i z tego powodu ten warunek jest wystarczający dla lokalnej identyfikacji parametrów.

Problem identyfikowalności parametrów będziemy jeszcze wielokrotnie rozpatrywać w kontekście konkretnych modeli statystycznych.

10.3. Zbieżność estymatorów MNW

Po to by udowodnić zbieżność estymatorów MNW posłużymy się następującym prostym lematem, który mówi, że wartość oczekiwana funkcji wiarygodności przyjmuje maksimum dla θ równej prawdziwemu parametrowi.

Lemat 10.1 *Jeśli θ jest identyfikowalna i dla każdego $\hat{\theta} \in \Theta$ z prawdopodobieństwem 1 zachodzi, że $L(\hat{\theta}) \neq 0$, to dla każdego ε*

$$E[\ell(\theta)] \geq E[\ell(\theta + \varepsilon)]$$

przy czym równość zachodzi jedynie dla $\varepsilon = 0$.

Dowód. Rozwińmy najpierw funkcję logarytmiczną dla $z \geq 1$ wokół 1

$$\ln(z) = \ln(1) + \frac{1}{1!}(z-1) - \frac{1}{2!\bar{z}^2}(z-1)^2 = z-1 - \frac{1}{2\bar{z}^2}(z-1)^2, \quad (10.2)$$

gdzie $\bar{z} \in [1, z]$. Zauważmy teraz, że

$$E[\ell(\theta + \varepsilon)] - E[\ell(\theta)] = \int_{\mathbf{Y}} \ln\left(\frac{L(\theta + \varepsilon)}{L(\theta)}\right) L(\theta) d\mathbf{y}$$

wykorzystując wzór (10.2) dla $z = \frac{L(\theta + \varepsilon)}{L(\theta)}$ otrzymujemy

$$E[\ell(\theta + \varepsilon)] - E[\ell(\theta)] = \int_{\mathbf{Y}} L(\theta + \varepsilon) d\mathbf{y} - \int_{\mathbf{Y}} L(\theta) d\mathbf{y} - \int_{\mathbf{Y}} \frac{1}{2\bar{z}^2} \left[\frac{L(\theta + \varepsilon) - L(\theta)}{L(\theta)} \right]^2 d\mathbf{y} \leq 0,$$

ponieważ $L(\theta + \varepsilon)$ i $L(\theta)$ są funkcjami gęstości a z definicji funkcji gęstości wynika, że jej całka po całej przestrzeni zdarzeń jest równa 1, więc

$$\int_{\mathbf{Y}} L(\theta + \varepsilon) d\mathbf{y} - \int_{\mathbf{Y}} L(\theta) d\mathbf{y} = 1 - 1 = 0$$

Równość będzie zachodzić jedynie dla $\varepsilon = \mathbf{0}$, ponieważ z identyfikowalności parametrów wnioskujemy, że z $\varepsilon \neq \mathbf{0}$ wynika, że istnieje \mathbf{y} dla którego $L(\boldsymbol{\theta} + \varepsilon) \neq L(\boldsymbol{\theta})$ i tym samym $\int_{\mathbf{Y}} [L(\boldsymbol{\theta} + \varepsilon) - L(\boldsymbol{\theta})]$

■

Lemat 10.2 *Jeśli $a_n \xrightarrow{p} a > 0$, to $\lim_{n \rightarrow \infty} \Pr(a_n > 0) = 0$.*

Dowód.

$$\begin{aligned} 0 &\leq \lim_{n \rightarrow \infty} \Pr(a_n < 0) \leq \lim_{n \rightarrow \infty} [\Pr(a_n < 0) + \Pr(a_n > 2a)] \\ &= \lim_{n \rightarrow \infty} [\Pr(|a_n - a| > a)] = 0 \end{aligned}$$

ponieważ z definicji zbieżności wynika, że $\lim_{n \rightarrow \infty} [\Pr(|a_n - a| > \delta)] = 0$ dla każdego $\delta > 0$, a więc i dla $\delta = a$. ■

Podane niżej twierdzenie podaje warunki zbieżności estymatora MNW

Twierdzenie 10.3 *Jeśli funkcja $\ell_i(\hat{\boldsymbol{\theta}})$ jest różniczkowalna dla każdej $\hat{\boldsymbol{\theta}} \in \Theta$ z prawdopodobieństwem 1 i $\boldsymbol{\theta}$ jest lokalnie asymptotycznie identyfikowalna, , oraz dla każdego ε*

$$n^{-1} \ell(\boldsymbol{\theta} + \varepsilon) = n^{-1} \sum_{i=1}^n \ell_i(\boldsymbol{\theta} + \varepsilon) \xrightarrow{p} E[n^{-1} \ell(\boldsymbol{\theta} + \varepsilon)] > -\infty$$

to

$$\tilde{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$$

Uwaga 10.4 *Zwykle zakłada się, że $E[\ell_1(\boldsymbol{\theta})] = \dots = E[\ell_n(\boldsymbol{\theta})]$, $\text{Var}[\ell_1(\boldsymbol{\theta})] = \dots = \text{Var}[\ell_n(\boldsymbol{\theta})]$ i wtedy to, że $n^{-1} \sum_{i=1}^n \ell_i(\boldsymbol{\theta} + \varepsilon) \xrightarrow{p} E[\ell_i(\boldsymbol{\theta} + \varepsilon)]$ jest konsekwencją słabego prawa wielkich liczb.*

Dowód. Z lematu (10.1) wynika, że $E[n^{-1} \ell(\boldsymbol{\theta})] \geq E[n^{-1} \ell(\boldsymbol{\theta} + \varepsilon)]$ dla każdego ε , a więc

$$n^{-1} [\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta} + \varepsilon)] \xrightarrow{p} \lim_{n \rightarrow \infty} E\{n^{-1} [\ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta} + \varepsilon)]\} > 0$$

Z lematu 10.2 wynika, że

$$\lim_{n \rightarrow \infty} \Pr \{ \ell(\boldsymbol{\theta}) - \ell(\boldsymbol{\theta} + \boldsymbol{\varepsilon}) < 0 \} = 0$$

Prawdopodobieństwo, że $\ell(\boldsymbol{\theta})$ nie jest maksimum $\ell(\cdot)$ dąży do zera. W konsekwencji $\tilde{\boldsymbol{\theta}} \xrightarrow{p} \boldsymbol{\theta}$, gdzie $\tilde{\boldsymbol{\theta}} = \arg \max_{\boldsymbol{\theta}} \ell(\boldsymbol{\theta})$. Dla $\ell(\boldsymbol{\theta})$ różniczkowalnej, $\tilde{\boldsymbol{\theta}}$ znajdujemy jako rozwiązanie równania

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = 0. \quad \blacksquare$$

10.4. Rozkład estymatorów MNW

W tym rozdziale dowiedzimy ważnego i uniwersalnego twierdzenia dotyczącego rozkładu estymatorów MNW. Ma ono fundamentalne znaczenie dla wnioskowania statystycznego na temat własności $\tilde{\boldsymbol{\theta}}$.

Twierdzenie 10.5 *Jeśli $\boldsymbol{\theta}$ należy do wnętrza Θ , $L(\boldsymbol{\theta})$ jest podtrójnie ciągle różniczkowalna względem każdego $\boldsymbol{\theta}$ z prawdopodobieństwem 1, $L(\boldsymbol{\theta}) \neq 0$ z prawdopodobieństwem 1, spełnione są warunki regularności*

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbf{Y}} L(\boldsymbol{\theta}) d\mathbf{y} = \int_{\mathbf{Y}} \frac{\partial}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}) d\mathbf{y} \quad (10.3)$$

$$\frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \int_{\mathbf{Y}} L(\boldsymbol{\theta}) d\mathbf{y} = \int_{\mathbf{Y}} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L(\boldsymbol{\theta}) d\mathbf{y} \quad (10.4)$$

warunek

$$\left\| \frac{\partial^2 (\partial \ell(\boldsymbol{\theta}) / \partial \theta_j)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right\| = O_p(1) \text{ dla } j = 1, \dots, k \quad (10.5)$$

i warunek lokalnej identyfikowalności parametrów

$$\mathbf{i}(\boldsymbol{\theta}) = \lim_{n \rightarrow \infty} \mathbf{i}_n(\boldsymbol{\theta}) > 0 \quad (10.6)$$

przy czym

$$\mathbf{i}_n(\boldsymbol{\theta}) = n^{-1} \text{Var} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = -n^{-1} \text{E} \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)$$

to rozkład estymatora MNW jest dany wzorem

$$\sqrt{n} \left(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \right) \xrightarrow{D} N \left(\mathbf{0}, \mathbf{i}^{-1}(\boldsymbol{\theta}) \right)$$

Dowód. Funkcja wiarygodności jest funkcją zmiennej losowej i tym samym może być sama traktowana jako zmienna losowa. Z ogólnych własności funkcji prawdopodobieństwa wiemy, że

$$\int_{\mathbf{Y}} L(\boldsymbol{\theta}) d\mathbf{y} = 1$$

gdzie \mathbf{Y} jest zbiorem możliwych wartości wektora losowego $\mathbf{y} = (y'_1, \dots, y'_n)'$. Zauważmy, że

$$\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = \frac{1}{L(\boldsymbol{\theta})} \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$$

Jeśli spełnione są warunki regularności (10.3) i (10.4), to wartość oczekiwana pochodnej funkcji wiarygodności jest równa zero:

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbf{Y}} L(\boldsymbol{\theta}) d\mathbf{y} = \int_{\mathbf{Y}} \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} d\mathbf{y} = \int_{\mathbf{Y}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}) d\mathbf{y} = \text{E} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{0}$$

ponieważ

$$\frac{\partial}{\partial \boldsymbol{\theta}} \int_{\mathbf{Y}} L(\boldsymbol{\theta}) d\mathbf{y} = \frac{\partial 1}{\partial \boldsymbol{\theta}} = 0$$

W ten sam sposób możemy pokazać, że wariancja $\ell(\boldsymbol{\theta})$ jest równa Hessianowi $-\ell(\boldsymbol{\theta})$:

$$\begin{aligned} \frac{\partial^2}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \int_{\mathbf{Y}} L(\boldsymbol{\theta}) d\mathbf{y} &= \int_{\mathbf{Y}} \frac{\partial}{\partial \boldsymbol{\theta}} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} L(\boldsymbol{\theta}) \right) d\mathbf{y} \\ &= \int_{\mathbf{Y}} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L(\boldsymbol{\theta}) d\mathbf{y} + \int_{\mathbf{Y}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} d\mathbf{y} = \mathbf{0} \end{aligned}$$

co implikuje, że

$$\begin{aligned} \int_{\mathbf{Y}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial L(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} d\mathbf{y} &= \int_{\mathbf{Y}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} L(\boldsymbol{\theta}) d\mathbf{y} = \text{Var} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \\ &= - \int_{\mathbf{Y}} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} L(\boldsymbol{\theta}) d\mathbf{y} = -\text{E} \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = \mathbf{I}_n(\boldsymbol{\theta}), \end{aligned}$$

gdzie $\mathbf{I}(\boldsymbol{\theta})$ nazywamy macierzą informacji Fishera. Ponieważ założyliśmy, że poszczególne obserwacje są niezależne, więc

$$\begin{aligned} \text{Var} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) &= \text{Var} \left(\sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \sum_{i=1}^n \text{Var} \left(\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = - \sum_{i=1}^n \text{E} \left(\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = n \mathbf{i}_n(\boldsymbol{\theta}) \\ \mathbf{I}_n(\boldsymbol{\theta}) &= n \mathbf{i}_n(\boldsymbol{\theta}) \end{aligned}$$

Z prawa wielkich liczb wynika, że jeśli prawdziwe jest założenie (10.6), to

$$n^{-1} \sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \xrightarrow{p} \lim_{n \rightarrow \infty} n^{-1} \text{Var} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{i}(\boldsymbol{\theta}) \quad (10.7)$$

i

$$n^{-1} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = n^{-1} \sum_{i=1}^n \frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \xrightarrow{p} \lim_{n \rightarrow \infty} n^{-1} \mathbb{E} \left(\frac{\partial^2 \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) = -\mathbf{i}(\boldsymbol{\theta}) \quad (10.8)$$

z czego wynika, że $\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = O_p(n)$. Dodatkowo

$$\mathbb{E} \left(\mathbf{i}^{-1}(\boldsymbol{\theta}) \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{i}^{-1}(\boldsymbol{\theta}) \mathbb{E} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{0}$$

a

$$\lim_{n \rightarrow \infty} \text{Var} \left(\mathbf{i}^{-1}(\boldsymbol{\theta}) \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) = \mathbf{i}^{-1}(\boldsymbol{\theta}) \lim_{n \rightarrow \infty} n^{-1} \text{Var} \left(\frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right) \mathbf{i}^{-1}(\boldsymbol{\theta}) = \mathbf{i}^{-1}(\boldsymbol{\theta}).$$

Na mocy (10.6) macierz $\mathbf{i}^{-1}(\boldsymbol{\theta})$ jest skończona, więc z centralnego twierdzenia granicznego wynika, że

$$n^{-\frac{1}{2}} \mathbf{i}^{-1}(\boldsymbol{\theta}) \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = n^{-\frac{1}{2}} \mathbf{i}^{-1}(\boldsymbol{\theta}) \sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \xrightarrow{D} N(\mathbf{0}, \mathbf{i}^{-1}(\boldsymbol{\theta})) \quad (10.9)$$

Rozważmy teraz estymator $\tilde{\boldsymbol{\theta}}$ policzony jako $\tilde{\boldsymbol{\theta}} = \arg \max \ell(\boldsymbol{\theta})$, czyli jako taka wartość $\boldsymbol{\theta}$, dla której wartość funkcji wiarygodności jest największa. Przy spełnionych założeniach (10.5) i (10.6) spełnione są założenia twierdzenia 10.3 i tym samym $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} = o_p(1)$. Możemy więc zastosować twierdzenie (17.39) do rozwinięcia kolejnych elementów wektora $\frac{\partial \ell(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}}$ wokół $\boldsymbol{\theta}$ tak, że

$$\frac{\partial \ell(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + o_p(1)$$

Ponieważ dla $f_{\boldsymbol{\theta}}(\cdot)$ ciągłych w maksimum $\frac{\partial \ell(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \mathbf{0}$, a $\left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} = O_p(n^{-1})$ więc

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} = - \left(\frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + o_p(n^{-1}) \quad (10.10)$$

Z wcześniej podanych wyprowadzeń wynika, że

$$n^{\frac{1}{2}} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) = \left(-n^{-1} \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} \right)^{-1} n^{-\frac{1}{2}} \sum_{i=1}^n \frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} + o_p \left(n^{-\frac{1}{2}} \right) \xrightarrow{D} N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta}))$$

stosując, inny mniej formalny¹ zapis

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \xrightarrow{D} N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta}))$$

■

Wniosek 10.6 (niezmienniczość) Estymator MNW dla $\boldsymbol{\omega} = \mathbf{g}(\boldsymbol{\theta})$, gdzie $\mathbf{g}(\boldsymbol{\theta})$ funkcją zdefiniowaną na Θ , różnowartościową i niezależną od n , jest dany wzorem $\tilde{\boldsymbol{\omega}} = \mathbf{g}(\tilde{\boldsymbol{\theta}})$. Jeśli $\ell(\boldsymbol{\theta})$ spełnia warunki twierdzenia (10.3), to

$$\tilde{\boldsymbol{\omega}} \xrightarrow{p} \boldsymbol{\omega}$$

Wniosek 10.7 Jeśli $\ell(\boldsymbol{\theta})$ spełniająca warunki twierdzenia (10.5) i $\mathbf{g}(\boldsymbol{\theta})$ jest podwójnie różniczkowalna oraz

$$\frac{\partial^2 \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'} = O_p(1),$$

to rozkład $\tilde{\boldsymbol{\omega}}$ jest dany wzorem

$$n^{\frac{1}{2}} (\tilde{\boldsymbol{\omega}} - \boldsymbol{\omega}) \xrightarrow{D} N\left(\mathbf{0}, \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}'} \mathbf{I}^{-1}(\boldsymbol{\theta}) \frac{\partial \mathbf{g}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}\right)$$

¹Zapis ten jest w zasadzie niepoprawny, skoro dla $n \rightarrow \infty$ macierz $\mathbf{I}^{-1}(\boldsymbol{\theta}) \rightarrow \mathbf{0}$. Niemniej jest on często stosowany, ponieważ w próbach skończonych zazwyczaj interesuje nas zazwyczaj wariancja $\tilde{\boldsymbol{\theta}}$ a nie $\sqrt{n}\tilde{\boldsymbol{\theta}}$.

Dowód.

$$\begin{aligned}\tilde{\omega} &= \arg \max_{\omega} \ell_{\omega}(\omega) = g \left(\arg \max_{\theta} \ell_{\theta}(g^{-1}(\omega)) \right) \\ &= g \left(\arg \max_{\theta} \ell_{\theta}(\theta) \right) = g(\tilde{\theta})\end{aligned}$$

Do dowodu drugiego wniosku możemy, przy przyjętych założeniach, wykorzystać twierdzenie (17.39)

$$g(\tilde{\theta}) = g(\theta) + \frac{\partial g(\theta)}{\partial \theta'} (\tilde{\theta} - \theta) + o_p(1)$$

Z twierdzenia (10.5) i twierdzenia Cramera (17.29) wynika, że

$$\sqrt{n} \left[g(\tilde{\theta}) - g(\theta) \right] \xrightarrow{D} \frac{\partial g(\theta)}{\partial \theta'} \eta,$$

gdzie $\eta \sim N(0, i^{-1}(\theta))$. ■

Estymator *MNW* funkcji parametrów jest równy funkcji estymatorów tych parametrów. Macierz informacji Fishera $\mathbf{I}(\theta)$ jest pewną funkcją θ .

Z powyższego twierdzenia wynika, że przy spełnionych założeniach twierdzenia 10.5

$$\begin{aligned}i(\tilde{\theta}) &= n^{-1} \sum_{i=1}^n \mathbb{E} \left(\frac{\partial \ell_i(\tilde{\theta})}{\partial \theta} \frac{\partial \ell_i(\tilde{\theta})}{\partial \theta'} \right) \\ &= -n^{-1} \sum_{i=1}^n \mathbb{E} \left(\frac{\partial^2 \ell_i(\tilde{\theta})}{\partial \theta \partial \theta'} \right) \xrightarrow{p} i(\theta)\end{aligned} \tag{10.11}$$

i jest tym samym $i(\tilde{\theta})$ jest zgodnym estymatorem $i(\theta)$. W praktyce często trudno jest analitycznie policzyć wariancję gradientu lub wartość oczekiwaną hessianu funkcji warygodności. Możemy jednak skorzystać z tego, że $\frac{\partial \ell_i(\tilde{\theta})}{\partial \theta}$ i $\frac{\partial^2 \ell_i(\tilde{\theta})}{\partial \theta \partial \theta'}$ są nieobciążonymi estymatorami $\frac{\partial \ell_i(\theta)}{\partial \theta}$ i $\frac{\partial^2 \ell_i(\theta)}{\partial \theta \partial \theta'}$. Z prawa wielkich liczb wynika, skoro $i^{-1}(\tilde{\theta})$ jest zbieżny, to we wzorze (10.11) możemy zrezygnować z operatorów wartości oczekiwanych.

Lemat 10.8 Przy spełnionych warunkach twierdzenia (10.5) estymatory macierzy $i^{-1}(\theta)$ postaci

$$\tilde{i}_n(\tilde{\theta}) = n^{-1} \sum_{i=1}^n \frac{\partial \ell_i(\tilde{\theta})}{\partial \theta} \frac{\partial \ell_i(\tilde{\theta})}{\partial \theta'} \xrightarrow{p} i(\theta) \quad (10.12)$$

i

$$\hat{i}_n(\tilde{\theta}) = -n^{-1} \sum_{i=1}^n \frac{\partial^2 \ell_i(\tilde{\theta})}{\partial \theta \partial \theta'} \xrightarrow{p} i(\theta) \quad (10.13)$$

są tym samym estymatorami zgodnymi. Estymator (10.12) nazywany jest estymatorem BHHH (Berndt, Hall, Hall, Hausman).

W konsekwencji estymatorem $\text{Var}(\tilde{\theta})$ będzie $\tilde{I}_n(\tilde{\theta}) = n\tilde{i}_n(\tilde{\theta})$ lub $\hat{I}_n(\tilde{\theta}) = n\hat{i}_n(\tilde{\theta})$. W próbach skończonych te dwa sposoby liczenia estymatory wariancji $\tilde{\theta}$ będą z reguły dawały różne wyniki. Bardziej popularny jest estymator $\tilde{I}^{-1}(\tilde{\theta})$, ponieważ wymaga policzenia jedynie pierwszych pochodnych $\ell_i(\theta)$.

Uwaga 10.9 W praktyce często definiuje się wektor (score)

$$\mathbf{G} = \begin{bmatrix} D\ell_i(\boldsymbol{\theta})' \\ \vdots \\ D\ell_i(\boldsymbol{\theta})' \end{bmatrix}, \quad \tilde{\mathbf{G}} = \begin{bmatrix} D\ell_i(\tilde{\boldsymbol{\theta}})' \\ \vdots \\ D\ell_i(\tilde{\boldsymbol{\theta}})' \end{bmatrix}$$

co pozwala zapisać estymator BHHH jako

$$D\ell(\boldsymbol{\theta}) = \mathbf{G}'\mathbf{l}$$

$$D\ell(\tilde{\boldsymbol{\theta}}) = \tilde{\mathbf{G}}'\mathbf{l}$$

i

$$\mathbf{i}(\boldsymbol{\theta}) = n^{-1}\mathbf{G}'\mathbf{G} = \sum_{i=1}^n D\ell_i(\boldsymbol{\theta}) D\ell_i(\boldsymbol{\theta})'$$

$$\tilde{\mathbf{i}}_n(\tilde{\boldsymbol{\theta}}) = n^{-1}\tilde{\mathbf{G}}'\tilde{\mathbf{G}}$$

Na żadnym z etapów dowodu nie zakładaliśmy, że macierz $\mathbf{i}(\boldsymbol{\theta}) < \infty$ a tym samym $\mathbf{i}^{-1}(\boldsymbol{\theta})$ jest nieosobliwe. Okazuje się, że przypadek, dla którego $\mathbf{i}^{-1}(\boldsymbol{\theta})$ jest osobliwe jest w ekonometrii bardzo ważny.

Wniosek 10.10 Estymatory największej wiarygodności są estymatorami zgodnymi $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \xrightarrow{p} \mathbf{0}$ a ich wariancja jest asymptotycznie równa dolnemu ograniczeniu Rao-Cramera $\mathbf{I}^{-1}(\boldsymbol{\theta})$. Estymatory te są więc asymptotycznie efektywne. W małych próbach estymatory MNW mogą być zarówno obciążone jak i nieefektywne.

10.5. MNW i liniowa i nieliniowa MNK

• Estymacja KMRL metodą największej wiarygodności

Założmy, że spełnione są założenia KMRL. Z założeń tych wynika, że obserwacje są po \mathbf{y}_i niezależne a rozkład y_i jest dany

$$y_i \sim N(\mathbf{x}_i\boldsymbol{\beta}, \sigma^2)$$

Funkcja gęstości dla pojedynczej obserwacji jest dana wzorem:

$$f_{\boldsymbol{\beta}, \sigma^2}(y_i | \mathbf{x}_i) = (2\pi\sigma^2)^{-\frac{1}{2}} \exp\left(-\frac{(y_i - \mathbf{x}_i\boldsymbol{\beta})^2}{2\sigma^2}\right)$$

Funkcja wiarygodności będzie więc równa

$$\begin{aligned} L(\boldsymbol{\beta}, \sigma^2) &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\sum_{i=1}^n \frac{(y_i - \mathbf{x}_i\boldsymbol{\beta})^2}{2\sigma^2}\right) \\ &= (2\pi\sigma^2)^{-\frac{n}{2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})\right) \end{aligned}$$

a logarytm funkcji wiarygodności będzie miał postać

$$\ell(\boldsymbol{\beta}, \sigma^2) = -\frac{n}{2} \ln(2\pi) - \frac{n}{2} \ln(\sigma^2) - \frac{1}{2\sigma^2} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})$$

Warunki pierwszego rzędu na maksimum są dane wzorami

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \boldsymbol{\beta}} = -\frac{1}{2\sigma^2} (2\mathbf{X}'\mathbf{X}\boldsymbol{\beta} - 2\mathbf{X}'\mathbf{y}) = \mathbf{0}$$

$$\frac{\partial \ell(\boldsymbol{\beta}, \sigma^2)}{\partial \sigma^2} = -\frac{n}{2} \frac{1}{\sigma^2} + \frac{1}{2\sigma^4} (\mathbf{y} - \mathbf{X}\boldsymbol{\beta})' (\mathbf{y} - \mathbf{X}\boldsymbol{\beta}) = \mathbf{0}$$

Wynika z tego, że estymatorami MNW w $KMRL$ parametrów β i σ^2 są

$$\tilde{\beta} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$$

$$\tilde{\sigma}^2 = \frac{(\mathbf{y} - \mathbf{X}\tilde{\beta})'(\mathbf{y} - \mathbf{X}\tilde{\beta})}{n} = \frac{\mathbf{e}'\mathbf{e}}{n}$$

gdzie $\mathbf{e} = \mathbf{y} - \mathbf{X}\mathbf{b}$

Estymator MNK parametru β ma taką samą postać jak estymator MNW tego parametru. Estymatory MNK i MNW σ^2 różnią się, przy czym estymator MNW jest zgodny ale obciążony w małych próbach

$$E(\tilde{\sigma}^2) = E\left(\frac{\mathbf{e}'\mathbf{e}}{n}\right) = \frac{n-K-1}{n} E(s^2) = \frac{n-K-1}{n} \sigma^2 \rightarrow \sigma^2$$

Macierz informacji Fishera dla estymatorów MNW liczymy znajdując macierz drugich pochodnych logarytmu funkcji wiarygodności

$$\frac{\partial^2 \ell(\beta, \sigma^2)}{\partial \beta \partial \beta'} = -\frac{\mathbf{X}'\mathbf{X}}{\sigma^2}$$

$$\frac{\partial^2 \ell(\beta, \sigma^2)}{\partial \sigma^2 \partial \sigma^2} = \frac{n}{2} \frac{1}{\sigma^4} - \frac{1}{\sigma^6} (\mathbf{y} - \mathbf{X}\beta)'(\mathbf{y} - \mathbf{X}\beta) = \frac{n}{2} \frac{1}{\sigma^4} - \frac{1}{\sigma^6} \mathbf{e}'\mathbf{e}$$

$$\frac{\partial^2 \ell(\beta, \sigma^2)}{\partial \beta \partial \sigma^2} = \frac{1}{\sigma^4} \mathbf{X}'(\mathbf{X}\beta - \mathbf{y}) = -\frac{1}{\sigma^4} \mathbf{X}'\mathbf{e}$$

Ponieważ $E(\mathbf{e}) = \mathbf{0}$ a $E(\mathbf{e}'\mathbf{e}) = n\sigma^2$ więc macierz informacji Fishera jest równa

$$\mathbf{I}(\beta, \sigma^2) = - \begin{bmatrix} E\left(-\frac{\mathbf{X}'\mathbf{X}}{\sigma^2}\right) & E\left(-\frac{1}{\sigma^4} \mathbf{X}'\mathbf{e}\right) \\ E\left(-\frac{1}{\sigma^4} \mathbf{X}'\mathbf{e}\right)' & E\left(\frac{n}{2} \frac{1}{\sigma^4} - \frac{1}{\sigma^6} \mathbf{e}'\mathbf{e}\right) \end{bmatrix} = \begin{bmatrix} \frac{\mathbf{X}'\mathbf{X}}{\sigma^2} & \mathbf{0} \\ \mathbf{0} & \frac{n}{2\sigma^4} \end{bmatrix} \quad (10.14)$$

W rezultacie asymptotyczna macierz wariancji-kowariancji dla estymatora MNW będzie równa

$$\text{Var} \left(\tilde{\beta}, \tilde{\sigma}^2 \right) = \mathbf{I}^{-1} \left(\beta, \sigma^2 \right) = \begin{bmatrix} \sigma^2 \left(\mathbf{X}' \mathbf{X} \right)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\sigma^4}{n} \end{bmatrix}$$

Zauważmy, że uzyskany wzór na wariancję estymatora $\tilde{\beta}$ jest analogiczny do wzoru uzyskanego dla estymatora \mathbf{b}_{MNK} i że podobnie jak w MNK estymatory MNW parametrów σ i β są niezależne. Podstawiając do wzoru na $\mathbf{I}^{-1} \left(\beta, \sigma^2 \right)$ estymatory $\tilde{\beta}, \tilde{\sigma}^2$ otrzymujemy estymatory wariancji $\hat{\beta}$ i $\hat{\sigma}^2$

$$\text{Var} \left(\tilde{\beta} \right) = \tilde{\sigma}^2 \left(\mathbf{X}' \mathbf{X} \right)^{-1} + o_p \left(1 \right)$$

$$\text{Var} \left(\tilde{\sigma}^2 \right) = \frac{2\tilde{\sigma}^4}{n} + o_p \left(1 \right)$$

Na koniec warto zauważyć, że wielkość logarytmu funkcji wiarygodności w maksimum jest równa

$$\begin{aligned} \ell \left(\tilde{\beta}, \tilde{\sigma}^2 \right) &= -\frac{n}{2} \ln \left(2\pi \right) - \frac{n}{2} \ln \left(\tilde{\sigma}^2 \right) - \frac{1}{\tilde{\sigma}^2} \left(\mathbf{y} - \mathbf{X}\tilde{\beta} \right)' \left(\mathbf{y} - \mathbf{X}\tilde{\beta} \right) \\ &= -\frac{n}{2} \ln \left(2\pi \right) - n - \frac{n}{2} \ln \left(\tilde{\sigma}^2 \right) = \text{const} - \frac{n}{2} \ln \left(\tilde{\sigma}^2 \right), \end{aligned}$$

gdzie jako const oznaczyliśmy część wartości tej funkcji $\ell \left(\tilde{\beta}, \tilde{\sigma}^2 \right)$, która nie zależy od wielkości wyestymowanych parametrów.

• Nieliniowa Metoda Najmniejszych Kwadratów

Wyobraźmy sobie, że mamy model

$$y_i = f \left(\mathbf{x}_i, \boldsymbol{\theta} \right) + \varepsilon_i$$

gdzie

$$\varepsilon \sim N \left(\mathbf{0}, \sigma^2 \mathbf{I} \right)$$

a funkcja $f(\mathbf{x}_i, \boldsymbol{\theta})$ jest nieliniowa względem wektora parametrów $\boldsymbol{\theta}$. Zlogarytmowana funkcja wiarygodności dla tego modelu ma postać

$$\ell(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta}) = -\left(\frac{n}{2}\right) \ln(2\pi) - \left(\frac{n}{2}\right) \ln(\sigma^2) - \frac{1}{2\sigma^2} \sum_{i=1}^n (y_i - f(\mathbf{x}_i, \boldsymbol{\theta}))^2$$

Parametr $\tilde{\boldsymbol{\theta}}$, maksymalizujący funkcję wiarygodności powinien minimalizować sumę kwadratów reszt

$$\min_{\tilde{\boldsymbol{\theta}}} \sum_{i=1}^n (y_i - f(\mathbf{x}_i, \tilde{\boldsymbol{\theta}}))^2 = \max_{\tilde{\boldsymbol{\theta}}} \sum_{i=1}^n e_i^2(\tilde{\boldsymbol{\theta}})$$

Z tego powodu metodę tę nazywamy nieliniową metodą najmniejszych kwadratów. Jak można wywnioskować z wyprowadzeń przeprowadzonych w kontekście Nieliniowej Metody Najmniejszych Kwadratów warunek pierwszego rzędu jest następujący

$$\frac{\partial \ell(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta})}{\partial \boldsymbol{\theta}} = -\frac{1}{2\sigma^2} \mathbf{X}^{0'} \mathbf{e} = \mathbf{0}$$

gdzie \mathbf{X}^0 jest macierzą pseudoregresorów. Rozwiązując warunki pierwszego rzędu dla σ^2 otrzymujemy, że estymator *MNW* parametru σ^2 w tym modelu ma postać

$$\tilde{\sigma}^2 = \frac{1}{n} \sum_{i=1}^n e_i^2(\hat{\boldsymbol{\theta}})$$

Z kolei wariancję estymatorów $\hat{\boldsymbol{\theta}}$ i $\tilde{\sigma}^2$ można policzyć za pomocą posługując się odwrotnością $\mathbf{I}^{-1}(\boldsymbol{\theta}, \sigma^2)$ policzoną dla $\hat{\boldsymbol{\theta}}$ i $\tilde{\sigma}^2$. Postać tej macierzy jest analogiczna do postaci tej macierzy dla *MNK* :

$$\text{Var}(\tilde{\boldsymbol{\beta}}, \tilde{\sigma}^2) = \mathbf{I}^{-1}(\hat{\boldsymbol{\beta}}, \hat{\sigma}^2) = \begin{bmatrix} \hat{\sigma}^2 (\mathbf{X}^{0'} \mathbf{X}^0)^{-1} & \mathbf{0} \\ \mathbf{0} & \frac{2\hat{\sigma}^4}{n} \end{bmatrix}$$

Wiele z testów, które zaprojektowano dla standardowej MNK można stosować po odpowiednich modyfikacjach także w nieliniowej MNK . Szacowanie modeli nieliniowych staje się coraz bardziej popularne w ekonometrii, ponieważ można dzięki nim używać elastyczniejszych form funkcyjnych. Problemy numeryczne związane z taką estymacją zostały znacznie złagodzone przez pojawienie się szybkich komputerów.

Przykład 10.11 *Model CES*

Uogólnieniem modelu Cobba-Douglasa jest model CES (Constant Elasticity of Substitution)

$$y = [a_1 x_1^\rho + a_2 x_2^\rho]^{\frac{1}{\rho}}$$

Funkcja ta sprowadza się do funkcji Cobba-Douglasa dla $\rho \rightarrow 0$ (ale jej szczególnymi przypadkami są także liniowa funkcja produkcji ($\rho = 1$) i Leontiefa ($\rho \rightarrow -\infty$)). Jeśli założymy, że funkcja produkcji ma postać funkcji CES to mamy otrzymamy model:

$$y_t = [a_1 x_{1t}^\rho + a_2 x_{2t}^\rho]^{\frac{1}{\rho}} \eta_t$$

po zlogarytmowaniu stronami mamy

$$\ln y_t = \frac{1}{\rho} \ln [a_1 x_{1t}^\rho + a_2 x_{2t}^\rho] + \varepsilon_t$$

i jeśli η_t miało rozkład lognormalny to $\varepsilon_t = \ln(\eta_t)$ ma rozkład normalny, to do estymacji tego modelu można użyć nieliniowej metody najmniejszych kwadratów.

Program PCGIVE:

```
actual = Q;
```

```
fitted = (1/&0)*log(&1*K^&0 + &2*L^&0);
```

10.6. Testowanie hipotez w MNW

Pożądaną cechą estymatorów MNW jest łatwość testowania ograniczeń narzuconych na wektor parametrów θ . Załóżmy, że testujemy ogólną nieliniową hipotezę zerową postaci $H_0 : \mathbf{h}(\theta) = \mathbf{0}$, przy czym zakładamy, że $\mathbf{h}(\theta)$ jest funkcją wektorową o g elementach i $g \leq k$. Hipotezę tą można przetestować za pomocą jednego z trzech testów: ilorazu wiarygodności (**Likelihood Ratio**) (LR), **Walda** (W) lub mnożników Lagrange'a (**Lagrange Multipliers**) (LM). Wszystkie testy opierają się na idei, że hipotezę zerową należy odrzucić jeśli jej przyjęcie wpływa istotnie na wynik maksymalizacji funkcji wiarygodności. W dalszych rozważaniach będziemy oznaczać jako $\mathbf{H}(\theta)$ wektor pierwszych pochodnych $\mathbf{h}(\theta)$ względem θ

$$\mathbf{H}(\theta) = \frac{\partial \mathbf{h}(\theta)}{\partial \theta'}$$

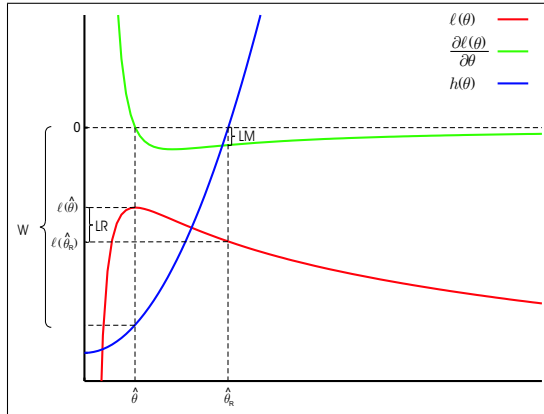
Okazuje się, że do wyprowadzenia rozkładów statystyk testowych konieczne jest przyjęcie dodatkowego założenia, że

$$\text{Rank} [\mathbf{H}(\theta) \mathbf{i}^{-1}(\theta) \mathbf{H}'(\theta)] = g \quad (10.15)$$

Założenie to gwarantuje, że $\mathbf{H}(\theta) \mathbf{i}^{-1}(\theta) \mathbf{H}'(\theta)$ jest odwracalna. Założenie to eliminuje z naszych rozważań przypadki, kiedy $\text{Rank} [\mathbf{H}(\theta)] < g$ oraz część przypadków, kiedy $\mathbf{i}^{-1}(\theta)$ jest osobliwa.

Jako $\tilde{\theta}$ oznaczymy estymator MNW wyliczony bez ograniczeń a jako $\tilde{\theta}_R$ estymator MNW z warunkami pobocznymi. Aby policzyć statystykę LM musimy znać zarówno $\tilde{\theta}$ jak i $\tilde{\theta}_R$. W przeciwieństwie do statystyki LR testując ograniczenia za pomocą testu LM , musimy policzyć jedynie estymator $\tilde{\theta}_R$, a w teście W jedynie estymator $\tilde{\theta}$. Poniższy rysunek ilustruje graficznie dla przypadku jednowymiarowej θ , jakie kryteria stosuje się w poszczególnych testach by określić, czy wprowadzenie ograniczeń silnie wpływa na wynik maksymalizacji funkcji wiarygodności.

Testy LM i W są bardzo użyteczne w sytuacjach, kiedy policzenie $\tilde{\theta}$ albo $\tilde{\theta}_R$ sprawiałoby duże problemy. Choć oparte na nieco innym rozumowaniu wszystkie te testy są asymptotycznie równoważne i mają rozkład asymptotyczny χ_g^2 . Na rysunku 1 można zauważyć, że statystyki LM ,



Rysunek 1: Przykładowe wielkości statystyk W , LM i LR dla $H_0 : h(\theta) = 0$.

LR , W mogą się znacznie różnić w małych próbach. Można udowodnić, że dla modeli liniowych zawsze zachodzi, że

$$LM \leq LR \leq W$$

W dalszej części tego rozdziału wyprowadzimy formalnie rozkład estymatora MNW z ograniczeniami a następnie postać testów W , LM i LR .

• Własności MNW dla estymatora z warunkami pobocznymi

Wyprowadzając postać estymatora MNW dla modelu z ograniczeniami możemy posłużyć się funkcją Lagrange'a

$$\mathcal{L}(\theta) = \ell(\theta) + \lambda' h(\theta)$$

Oznaczmy jako $D\ell(\tilde{\boldsymbol{\theta}}_R) = \left. \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\tilde{\boldsymbol{\theta}}_R}$ i $D^2\ell(\boldsymbol{\theta}) = \frac{\partial^2 \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}'}$. Warunki pierwszego rzędu dla tej maksymalizacji będą miały postać

$$\begin{aligned} \left. \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right|_{\tilde{\boldsymbol{\theta}}_R} &= D\ell(\tilde{\boldsymbol{\theta}}_R) + \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \tilde{\boldsymbol{\lambda}} = \mathbf{0} \\ \left. \frac{\partial \mathcal{L}(\boldsymbol{\theta})}{\partial \boldsymbol{\lambda}} \right|_{\tilde{\boldsymbol{\theta}}_R} &= \mathbf{h}(\tilde{\boldsymbol{\theta}}_R) = \mathbf{0} \end{aligned} \quad (10.16)$$

Jeśli ograniczenie jest prawdziwe, to z twierdzenia 10.3 wnioskujemy, że $\tilde{\boldsymbol{\theta}}_R \xrightarrow{p} \boldsymbol{\theta}$. Rozwijając $D\ell(\tilde{\boldsymbol{\theta}}_R)$ wokół $\tilde{\boldsymbol{\theta}}$ otrzymujemy z (17.39) i korzystając z tego, że $D\ell(\tilde{\boldsymbol{\theta}}) = \mathbf{0}$ otrzymujemy:

$$D\ell(\tilde{\boldsymbol{\theta}}_R) = D^2\ell(\tilde{\boldsymbol{\theta}}) (\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}}) + o_p(1)$$

Ponieważ $\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}} = O_p(n^{\frac{1}{2}})$ a $D^2\ell(\tilde{\boldsymbol{\theta}}) = O_p(n)$, więc $D\ell(\tilde{\boldsymbol{\theta}}_R) = O_p(n^{\frac{1}{2}})$. Mnożąc to równanie przez $[D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1}$ otrzymujemy

$$\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}} = [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} D\ell(\tilde{\boldsymbol{\theta}}_R) + o_p(n^{-1})$$

Ponieważ $[D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} = O_p(n^{-1})$, więc

$$\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}} = O_p(n^{-1}) O_p(n^{\frac{1}{2}}) + o_p(n^{-1}) = O_p(n^{-\frac{1}{2}}).$$

Wykorzystując ten fakt uzyskujemy z twierdzenia (17.39) lepsze przybliżenie $D\ell(\tilde{\boldsymbol{\theta}}_R)$:

$$D\ell(\tilde{\boldsymbol{\theta}}_R) = D^2\ell(\tilde{\boldsymbol{\theta}}) (\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}}) + O_p(n^{-1}) \quad (10.17)$$

Wstawiając to do (10.16) mamy

$$\mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \tilde{\boldsymbol{\lambda}} = -D^2\ell(\tilde{\boldsymbol{\theta}}) (\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}}) + O_p(n^{-1})$$

Mnożąc to równanie przez $\mathbf{H}(\tilde{\boldsymbol{\theta}}_R) [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} = O_p(n^{-1})$

$$\mathbf{H}(\tilde{\boldsymbol{\theta}}_R) [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \tilde{\boldsymbol{\lambda}} = -\mathbf{H}(\tilde{\boldsymbol{\theta}}_R) (\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}}) + O_p(n^{-2}).$$

Rozwiązując to równanie dla $\tilde{\boldsymbol{\lambda}}$ otrzymujemy:

$$\tilde{\boldsymbol{\lambda}} = - \left\{ \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \right\}^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) (\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}}) + O_p(n^{-1}) \quad (10.18)$$

Rozwinięcia $\mathbf{h}(\tilde{\boldsymbol{\theta}})$ wokół $\tilde{\boldsymbol{\theta}}_R$ i $\boldsymbol{\theta}$ mają postać

$$\mathbf{h}(\tilde{\boldsymbol{\theta}}) = \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) (\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}}) + O_p(n^{-1})$$

$$\mathbf{h}(\tilde{\boldsymbol{\theta}}) = \mathbf{H}(\boldsymbol{\theta}) (\tilde{\boldsymbol{\theta}}_R - \boldsymbol{\theta}) + O_p(n^{-1})$$

przy czym skorzystaliśmy z tego, że $\mathbf{h}(\boldsymbol{\theta}) = 0$ i $\mathbf{h}(\tilde{\boldsymbol{\theta}}_R) = 0$. W rezultacie

$$\mathbf{H}(\tilde{\boldsymbol{\theta}}_R) (\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}}) = \mathbf{H}(\boldsymbol{\theta}) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(n^{-1}) \quad (10.19)$$

z czego wynika, że

$$\tilde{\boldsymbol{\lambda}} = - \left\{ \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \right\}^{-1} \mathbf{H}(\boldsymbol{\theta}) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(1) \quad (10.20)$$

Wstawiając do pierwszego równania układu (10.16) mamy:

$$D\ell(\tilde{\boldsymbol{\theta}}_R) = \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \left\{ \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \right\}^{-1} \mathbf{H}(\boldsymbol{\theta}) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(1) \quad (10.21)$$

a podstawiając do lewej strony równania wzór (10.17) i mnożąc przez $[D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} = O_p(n^{-1})$ otrzymujemy

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}} &= [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \left\{ \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) [D^2\ell(\boldsymbol{\theta})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \right\}^{-1} \mathbf{H}(\boldsymbol{\theta}) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &\quad + O_p(n^{-1}) \end{aligned} \quad (10.22)$$

Odejmując od obu stron $\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}$ otrzymujemy z kolei

$$\begin{aligned} \tilde{\boldsymbol{\theta}}_R - \boldsymbol{\theta} &= \left\{ \mathbf{I} - [D^2\ell(\tilde{\boldsymbol{\theta}})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \left\{ \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) [D^2\ell(\boldsymbol{\theta})]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \right\}^{-1} \mathbf{H}(\boldsymbol{\theta}) \right\} (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) \\ &\quad + O_p(n^{-1}) \end{aligned}$$

Korzystając z kolei z twierdzenia 10.5, założenia (10.15) oraz wzoru (10.13) otrzymujemy:

$$\sqrt{n}(\tilde{\boldsymbol{\theta}}_R - \boldsymbol{\theta}) \xrightarrow{D} N(0, \mathbf{P}(\boldsymbol{\theta})),$$

gdzie

$$\mathbf{P}(\boldsymbol{\theta}) = \mathbf{i}^{-1}(\boldsymbol{\theta}) - \mathbf{i}^{-1}(\boldsymbol{\theta}) \mathbf{H}'(\boldsymbol{\theta}) [\mathbf{H}(\boldsymbol{\theta}) \mathbf{i}^{-1}(\boldsymbol{\theta}) \mathbf{H}'(\boldsymbol{\theta})]^{-1} \mathbf{H}(\boldsymbol{\theta}) \mathbf{i}^{-1}(\boldsymbol{\theta}).$$

Macierz $\mathbf{P}(\boldsymbol{\theta})$ jest macierzą wariancji kowariancji $\tilde{\boldsymbol{\theta}}_R$. Macierz ta jest w ogólnym przypadku macierzą osobliwą, ponieważ $\mathbf{P}(\boldsymbol{\theta}) \mathbf{H}'(\boldsymbol{\theta}) = \mathbf{0}$ a z założeń wiemy, że $\mathbf{H}'(\boldsymbol{\theta})$ jest macierzą

o pełnym rzędzie. Z jej formy można wywnioskować, że asymptotyczna wariancja estymatora $\sqrt{n}(\tilde{\theta}_R - \theta)$ jest mniejsza lub równa wariancji $\sqrt{n}(\tilde{\theta} - \theta)$.

• Statystyka Walda

Statystykę Walda wyprowadzamy wychodząc z rozkładu funkcji $h(\tilde{\theta})$. Z wniosku 10.6 wynika, że dla prawdziwej H_0 i estymatora $\tilde{\theta}$ policzonego dla modelu bez ograniczeń funkcja ta ma posiadać asymptotyczny rozkład

$$\sqrt{n}h(\tilde{\theta}) \xrightarrow{p} N(0, \mathbf{H}(\theta) \mathbf{I}^{-1}(\theta) \mathbf{H}'(\theta)) \quad (10.23)$$

Na podstawie tego samego wniosku możemy we wzorze na wariancję tej funkcji θ zastąpić estymatorem $MNW \tilde{\theta}$. Na podstawie twierdzenia 17.77 możemy sformułować następujące twierdzenie

Twierdzenie 10.12 *Przy spełnionych założeniach twierdzenia 10.5 i założeniu (10.15)*

$$W = h(\tilde{\theta})' [\mathbf{H}(\tilde{\theta}) \mathbf{I}^{-1}(\tilde{\theta}) \mathbf{H}'(\tilde{\theta})]^{-1} h(\tilde{\theta}) \xrightarrow{D} \chi_g^2, \quad (10.24)$$

gdzie $g = \text{Rank}(\mathbf{H}(\tilde{\theta}))$.

Do policzenia tej formy kwadratowej wystarczy znać wartość estymatora $\tilde{\theta}$, to jest estymatora dla modelu bez ograniczeń. Wadą statystyki Walda jest fakt, że wartość formy kwadratowej (10.24) może, w małych próbach, być różna się dla równoważnych ale analitycznie różnych sformułowań układu równań $h(\theta)$. Macierz $\mathbf{H}(\theta) \mathbf{I}^{-1}(\theta) \mathbf{H}'(\theta)$ zastępujemy zazwyczaj macierzą $\mathbf{H}(\tilde{\theta}) \tilde{\mathbf{I}}^{-1}(\tilde{\theta}) \mathbf{H}'(\tilde{\theta})$.

Przykład 10.13 *Testowanie hipotez nieliniowych w KMRL. Załóżmy, że w modelu*

$$y_t = \beta_0 + \beta_1 x_t + \varepsilon_t$$

chcemy przetestować hipotezę $H_0 : \beta_0\beta_1 = 1$. W przypadku KMRL estymator odwrotności macierz informacyjnej ma postać $\mathbf{I}^{-1}(\boldsymbol{\beta}) = \tilde{\sigma}^2 (\mathbf{X}'\mathbf{X})^{-1}$. Analizowaną hipotezę można sformułować jako $\mathbf{h}(\boldsymbol{\beta}) = \beta_0\beta_1 - 1$. Macierz $\mathbf{H}(\boldsymbol{\beta})$ ma postać

$$\mathbf{H}_{11}(\boldsymbol{\beta}) = \frac{\partial \mathbf{h}(\boldsymbol{\beta})}{\partial \beta_0} = \beta_1$$

$$\mathbf{H}_{12}(\boldsymbol{\beta}) = \frac{\partial \mathbf{h}(\boldsymbol{\beta})}{\partial \beta_1} = \beta_0$$

i w rezultacie $\mathbf{H}(\mathbf{b}) = [b_1, b_0]$. Postać testu Walda w tym przypadku będzie więc następująca

$$W = \frac{1}{\tilde{\sigma}^2} \frac{(b_0b_1 - 1)^2}{[b_1, b_0] (\mathbf{X}'\mathbf{X})^{-1} [b_1, b_0]'}$$

Sformułujmy teraz tę hipotezę w równoważny sposób jako $\beta_0 = \frac{1}{\beta_1}$. Hipoteza ma więc postać $\mathbf{h}(\boldsymbol{\beta}) = \beta_0 - \frac{1}{\beta_1}$. Macierz $\mathbf{H}(\boldsymbol{\beta})$ ma postać

$$\mathbf{H}_{11}(\boldsymbol{\beta}) = \frac{\partial \mathbf{h}(\boldsymbol{\beta})}{\partial \beta_0} = 1$$

$$\mathbf{H}_{11}(\boldsymbol{\beta}) = \frac{\partial \mathbf{h}(\boldsymbol{\beta})}{\partial \beta_1} = \frac{1}{\beta_1^2}$$

Postać testu Walda w tym przypadku będzie więc następująca

$$\begin{aligned} W &= \frac{1}{\tilde{\sigma}^2} \frac{\left(b_0 - \frac{1}{b_1}\right)^2}{\left[1, \frac{1}{b_1^2}\right] (\mathbf{X}'\mathbf{X})^{-1} \left[1, \frac{1}{b_1^2}\right]'} \\ &= \frac{1}{\tilde{\sigma}^2} \frac{(b_0 b_1 - 1)^2}{\left[b_1, \frac{1}{b_1}\right] (\mathbf{X}'\mathbf{X})^{-1} \left[b_1, \frac{1}{b_1}\right]'} \end{aligned}$$

Widać teraz, że dla dwóch różnych sposobów zapisu tej samej H_0 otrzymaliśmy dwa różne wzory na statystykę W . Przy założeniu prawdziwości H_0 obie te formy są asymptotycznie równoważne, bo $\text{plim}(b_0) = \text{plim}\left(\frac{1}{b_1}\right)$ dla $H_0 : \beta_0 = \frac{1}{\beta_1}$. W małych próbach ststystyki te mogą się jednak istotnie różnić.

• Statystyka mnożników Lagrange'a (LM)

Jeśli H_0 jest prawdziwa to ze wzorów (10.20) i (10.19) wnioskujemy, że

$$\tilde{\boldsymbol{\lambda}} = \left\{ \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) \left[D^2 \ell(\tilde{\boldsymbol{\theta}}) \right]^{-1} \mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \right\}^{-1} \mathbf{H}(\tilde{\boldsymbol{\theta}}_R) (\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}) + O_p(1)$$

wykorzystując twierdzenie 10.5 oraz wzoru (10.13) i wykorzystując założenie (10.15) uzyskujemy, że

$$n^{-\frac{1}{2}} \tilde{\boldsymbol{\lambda}} \xrightarrow{D} N\left(0, \left[\mathbf{H}(\boldsymbol{\theta}) \mathbf{i}^{-1}(\boldsymbol{\theta}) \mathbf{H}'(\boldsymbol{\theta}) \right]^{-1}\right) \quad (10.25)$$

z kolei zauważmy, że z (10.16) wynika, że

$$D\ell(\tilde{\boldsymbol{\theta}}_R) = -\mathbf{H}'(\tilde{\boldsymbol{\theta}}_R) \tilde{\boldsymbol{\lambda}} \quad (10.26)$$

Ze zbieżności $\tilde{\boldsymbol{\theta}}_R$ oraz twierdzenia (17.77) wynika z tego z kolei, że prawdziwe jest twierdzenie

Twierdzenie 10.14 Przy spełnionych założeniach twierdzenia 10.5 i założeniu (10.15)

$$\begin{aligned} LM &= \tilde{\lambda}' \left[\text{Var} \left(\tilde{\lambda} \right) \right]^{-1} \tilde{\lambda} = \tilde{\lambda}' \mathbf{H}(\boldsymbol{\theta}) \mathbf{I}^{-1}(\boldsymbol{\theta}) \mathbf{H}'(\boldsymbol{\theta}) \tilde{\lambda} \\ &= D\ell \left(\tilde{\boldsymbol{\theta}}_R \right)' \mathbf{I}^{-1}(\boldsymbol{\theta}) D\ell \left(\tilde{\boldsymbol{\theta}}_R \right) \xrightarrow{D} \chi_g^2, \end{aligned} \quad (10.27)$$

gdzie $g = \text{Rank}(\mathbf{H}(\boldsymbol{\theta}))$.

Wniosek 10.15 W Nieliniowej Metody Najmniejszych Kwadratów jest równe jest równe

$$LM = \frac{e_R'}{\hat{\sigma}} \mathbf{X}_R^0 \left(\mathbf{X}_R^{0'} \mathbf{X}_R^0 \right)^{-1} \mathbf{X}_R^{0'} \frac{e_R}{\hat{\sigma}}$$

gdzie $e_{Ri} = y_i - h(x_i, \tilde{\boldsymbol{\theta}}_R)$, \mathbf{X}_R^0 jest macierzą pseudoregresorów policzoną dla $\tilde{\boldsymbol{\theta}}_R$ a $\hat{\sigma}^2$ jest pewnym zgodnym estymatorem σ .

Dowód. W NMNK gradient równy jest $D\ell \left(\tilde{\boldsymbol{\theta}}_R \right) = -\frac{1}{\hat{\sigma}^2} \mathbf{X}_R^{0'} e_R$, odwrotność macierzy informacyjnej $\mathbf{I}^{-1}(\boldsymbol{\theta}) = \hat{\sigma}^2 \left(\mathbf{X}_R^{0'} \mathbf{X}_R^0 \right)^{-1}$. Statystyka LM ma więc postać

$$LM = \frac{e_R'}{\hat{\sigma}} \mathbf{X}_R^0 \left(\mathbf{X}_R^{0'} \mathbf{X}_R^0 \right)^{-1} \mathbf{X}_R^{0'} \frac{e_R}{\hat{\sigma}}$$

■

Wniosek 10.16 Jeśli wśród pseudoregresorów dla H_0 znajduje się wektor jedynek, to dla NMNK statystyka $LM = nR^2$, gdzie R^2 jest współczynnikiem determinacji w regresji e_R' na macierzy pseudoregresorów.

Dowód. Wartości dopasowane e_R w takiej regresji mają postać $\hat{e}_R' = e_R' \mathbf{X}_R^0 \left(\mathbf{X}_R^{0'} \mathbf{X}_R^0 \right)^{-1} \mathbf{X}_R^{0'}$. Zauważmy, że w takiej regresji z warunków pierwszego rzędu wynika, że $l' \hat{e}_R' = \mathbf{0}$ oraz $TSS =$

$e'_R e_R$ a $ESS = \widehat{e}'_R \widehat{e}_R$. Jeśli H_0 jest prawdziwa, to zgodnym estymatorem σ^2 jest $\widehat{\sigma}^2 = \frac{e'_R e_R}{n}$. Wstawiając te wnioski do poprzedniego wzoru uzyskujemy

$$LM = \frac{\widehat{e}'_R \widehat{e}_R}{\widehat{\sigma}^2} = n \frac{ESS}{TSS} = nR^2$$

■

Przykład 10.17 Spróbujmy przetestować testem LM hipotezę, że $\gamma = 0$ w modelu

$$y_i = \alpha + \beta \ln(x_{2i} + \gamma) + \varepsilon_i$$

Zauważmy najpierw, że dla H_0 model ten sprowadza się do modelu liniowego postaci

$$y_i = \alpha + \beta \ln(x_{2i}) + \varepsilon_i$$

W pierwszym kroku liczymy estymatory $\widehat{\alpha}$ i $\widehat{\beta}$ dla modelu z ograniczeniami za pomocą MNK. Postać pseudoregresorów dla $h(\mathbf{x}_i, \alpha, \beta, \gamma) = \alpha + \beta \ln(x_{2i} + \gamma)$ jest następująca:

$$\frac{h(\mathbf{x}_i, \alpha, \beta, \gamma)}{\partial \alpha} = 1$$

$$\frac{h(\mathbf{x}_i, \alpha, \beta, \gamma)}{\partial \beta} = \ln(x_{2i} + \gamma)$$

$$\frac{h(\mathbf{x}_i, \alpha, \beta, \gamma)}{\partial \gamma} = \frac{\beta}{x_{2i} + \gamma}$$

dla $\gamma = 0$ i $\widehat{\beta}$ pseudoregresory przyjmą postać 1 , $\ln(x_{2i})$ i $\frac{\widehat{\beta}}{x_{2i}}$. Statystyka LM będzie więc równa nR^2 z regresji reszt uzyskanych w pierwszym kroku reszt na 1 , $\ln(x_{2i})$ i $\frac{\widehat{\beta}}{x_{2i}}$.

Przykład 10.18 Test Breuscha-Godfrey'a. Załóżmy, że autokorelacja ma następującą formę AR(1). Wtedy model ma postać

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + u_t$$

$$u_t = \rho u_{t-1} + \varepsilon_t$$

$$\varepsilon \sim N(0, \sigma^2 \mathbf{I})$$

Model ten można sprowadzić do modelu nieliniowego poprzez odjęcie od obu stron równania $y_{t-1} = \mathbf{x}_{t-1}\boldsymbol{\beta} + u_{t-1}$ pomnożonego przez ρ

$$y_t - \rho y_{t-1} = \mathbf{x}_t \boldsymbol{\beta} - \rho \mathbf{x}_{t-1} \boldsymbol{\beta} + u_t - \rho u_{t-1}$$

$$y_t = \mathbf{x}_t \boldsymbol{\beta} + \rho (y_{t-1} - \mathbf{x}_{t-1} \boldsymbol{\beta}) + \varepsilon_t$$

Przy założeniu, że $H_0 : \rho = 0$ jest prawdziwe model ten sprowadza się do modelu liniowego, z którego możemy uzyskać estymator \mathbf{b} i wektor reszt \mathbf{e}_R . Pseudoregresory w tym modelu będą miały postać

$$\frac{h(\mathbf{x}_t, \beta, \rho)}{\partial \beta} = \mathbf{x}_t - \rho \mathbf{x}_t$$

$$\frac{h(\mathbf{x}_t, \beta, \rho)}{\partial \rho} = y_{t-1} - \mathbf{x}_{t-1} \boldsymbol{\beta}$$

dla $\rho = 0$ i \mathbf{b} uzyskujemy, że pseudoregresory mają postać \mathbf{x}_t, e_{t-1} . Statystyka testowa jest więc równa TR^2 z regresji e_t na \mathbf{x}_t, e_{t-1} .

Jak już wspomnieliśmy wcześniej istnieje zwykle kilka opcji jeśli chodzi o sposób policzenia macierzy informacyjnej. Wybranie jednej z tych opcji generuje różne formy statystyki LM , równoważne asymptotycznie ale różne w małych próbach. Formy te mogą się też różnić stopniem odporności na odchylenia od modelu od założeń.

Używając notacji (10.9) możemy zapisać $Dl'(\tilde{\boldsymbol{\theta}}) = l' \tilde{\mathbf{G}}_R$ i w konsekwencji $LM = l' \tilde{\mathbf{G}}_R \mathbf{I}^{-1}(\boldsymbol{\theta}) \tilde{\mathbf{G}}$. W przykładzie poniżej pokazujemy jak różne estymatory macierzy $\mathbf{I}^{-1}(\boldsymbol{\theta})$ generują różne formy statystyki LM dla tej samej H_0 .

Przykład 10.19 *Test Breuscha-Pagana. Badamy, czy w modelu*

$$\mathbf{y} = \mathbf{x}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$$

występuje heteroskedastyczność postaci:

$$\sigma_i^2 = \sigma^2 f(\mathbf{z}_i\boldsymbol{\alpha})$$

przy założeniu, że $f(0) = 1$ i błąd losowy

$$\boldsymbol{\varepsilon} \sim N(0, \boldsymbol{\Omega}), \text{ gdzie } \boldsymbol{\Omega} = \begin{bmatrix} \sigma_1^2 & & 0 \\ & \ddots & \\ 0 & & \sigma_n^2 \end{bmatrix}$$

Logarytmy funkcji gęstości dla poszczególnych y_i pochodzą wtedy z rozkładu normalnego i ma postać:

$$\ell_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_i^2) = -\frac{1}{2} \ln(2\pi) - \frac{1}{2} \ln \sigma_i^2 - \frac{(y_i - \mathbf{x}_i\boldsymbol{\beta})^2}{2\sigma_i^2}$$

Najprostszym sposobem zweryfikowania tej hipotezy jest posłużenie się testem LM. Testujemy hipotezę o braku autokorelacji, która dla tego modelu sprowadza się do hipotezy $H_0: \boldsymbol{\alpha} = \mathbf{0}$. Aby znaleźć postać testu LM musimy znaleźć postać wyrażenia $D\ell(\tilde{\boldsymbol{\theta}}_R) = -\mathbf{H}'(\tilde{\boldsymbol{\theta}}_R)\tilde{\boldsymbol{\lambda}}$. Pierwsza pochodna funkcji wiarygodności $\ell(\mathbf{y}, \mathbf{X}, \boldsymbol{\theta})$ względem $\boldsymbol{\alpha}$ jest równa:

$$\begin{aligned} \frac{\partial \ell_i(\boldsymbol{\alpha}, \boldsymbol{\beta}, \sigma_i^2)}{\partial \boldsymbol{\alpha}} &= -\frac{\sigma^2}{2} \frac{1}{\sigma_i^2} \frac{\partial f(\mathbf{z}_i\boldsymbol{\alpha})}{\partial (\mathbf{z}_i\boldsymbol{\alpha})} \mathbf{z}_i' + \frac{\sigma^2}{2} \frac{(y_i - \mathbf{x}_i\boldsymbol{\beta})^2}{\sigma_i^4} \frac{\partial f(\mathbf{z}_i\boldsymbol{\alpha})}{\partial (\mathbf{z}_i\boldsymbol{\alpha})} \mathbf{z}_i' \\ &= -\frac{\sigma^2}{2} \left\{ \frac{\partial f(\mathbf{z}_i\boldsymbol{\alpha})}{\partial (\mathbf{z}_i\boldsymbol{\alpha})} \frac{1}{\sigma_i^4} [-\sigma_i^2 + \varepsilon_i^2] \mathbf{z}_i' \right\}, \end{aligned}$$

dla $H_0 : \sigma_i^2 = \sigma^2$ mamy $\boldsymbol{\alpha} = 0$ i tym samym, że $\left. \frac{\partial f(\mathbf{z}_i, \boldsymbol{\alpha})}{\partial (\mathbf{z}_i, \boldsymbol{\alpha})} \right|_{\boldsymbol{\alpha}=0} = d$, które nie zależy od i . Gradient policzony dla przy założeniu, że $\sigma_i^2 = \sigma^2$ dla $i = 1, 2, \dots, n$ jest więc równy

$$D_{\alpha} \ell_i(0, \boldsymbol{\beta}, \sigma^2) = -\frac{d}{2\sigma^2} (\varepsilon_i^2 - \sigma^2) \mathbf{z}'_i,$$

co można też zapisać jako

$$\mathbf{G}_R = \mathbf{G}(0, \boldsymbol{\beta}, \sigma^2) = -\frac{d}{2\sigma^2} \mathbf{W} \mathbf{Z},$$

gdzie $\mathbf{W} = \text{diag}(\varepsilon_1^2 - \sigma^2, \dots, \varepsilon_n^2 - \sigma^2)$. Jeśli zastąpimy teraz $\boldsymbol{\beta}$ i σ^2 ich estymatorami $\mathbf{b} = (\mathbf{X}'\mathbf{X})^{-1} \mathbf{X}'\mathbf{y}$ i $\tilde{\sigma}^2 = n^{-1} \sum_{i=1}^n e_i^2$, to

$$\tilde{\mathbf{G}}_R = \mathbf{G}(0, \mathbf{b}, \tilde{\sigma}^2) = -\frac{d}{2\tilde{\sigma}^2} \tilde{\mathbf{W}} \mathbf{Z}$$

gdzie $\tilde{\mathbf{W}} = \text{diag}(e_1^2 - \tilde{\sigma}^2, \dots, e_n^2 - \tilde{\sigma}^2)$, $e_i = y_i - \mathbf{x}_i \mathbf{b}$. Zdefiniujmy $\frac{\mathbf{l}' \tilde{\mathbf{W}}}{\tilde{\sigma}^2} = \left(\frac{e_1^2}{\tilde{\sigma}^2} - 1, \dots, \frac{e_n^2}{\tilde{\sigma}^2} - 1 \right) = \mathbf{w}'$. Wektor $\mathbf{l}' \tilde{\mathbf{G}}_R$ możemy zapisać

$$\mathbf{l}' \tilde{\mathbf{G}}_R = -\frac{d}{2} \mathbf{w}' \mathbf{Z}$$

Najpierw policzmy iloczyn gradientów dla poszczególnych obserwacji:

$$D_{\alpha} \ell_i(0, \boldsymbol{\beta}, \sigma^2) D'_{\alpha} \ell_i(0, \boldsymbol{\beta}, \sigma^2) = \frac{d^2}{4\sigma^4} (\varepsilon_i^2 - \sigma^2)^2 \mathbf{z}_i \mathbf{z}'_i$$

Możemy teraz znaleźć macierz informacyjną na dwa sposoby:

1. Wiemy, że dla rozkładu normalnego $E(\varepsilon_i^4) = 3\sigma^4$ a więc

$$E(\varepsilon_i^2 - \sigma^2)^2 = E(\varepsilon_i^4) - 2\sigma^2 E(\varepsilon_i^2) + \sigma^4 = 2\sigma^4$$

i w rezultacie

$$\mathbb{E} [D_{\alpha} \ell_i (0, \beta, \sigma^2) D'_{\alpha} \ell_i (0, \beta, \sigma^2)] = \frac{d^2}{2} \mathbf{z}_i \mathbf{z}'_i,$$

Policzona przy użyciu wzoru (10.11) macierz informacyjna $\mathbf{I}(\boldsymbol{\theta}_R)$ ma więc postać:

$$\mathbf{I}(\tilde{\boldsymbol{\theta}}_R) = \sum_{i=1}^n \frac{d^2}{2} \mathbf{z}_i \mathbf{z}'_i = \frac{d^2}{2} \mathbf{Z}' \mathbf{Z}$$

gdzie $\mathbf{Z} = (\mathbf{z}'_1, \dots, \mathbf{z}'_n)'$.

2. Możemy także policzyć macierz estymator macierzy informacyjnej bez odwoływania się do własności rozkładu normalnego. Jeśli hipoteza H_0 jest prawdziwa, to $\mathbb{E} [(\varepsilon_i^2 - \sigma^2)^2]$ jest stała i niezależna od \mathbf{z}_i . W rezultacie

$$\begin{aligned} \mathbf{I}(\boldsymbol{\theta}_R) &= \mathbb{E} \left[\sum_{i=1}^n \frac{d^2}{4\sigma^4} (\varepsilon_i^2 - \sigma^2)^2 \mathbf{z}_i \mathbf{z}'_i \right] \\ &= \frac{d^2}{4\sigma^4} \mathbb{E} [(\varepsilon_i^2 - \sigma^2)^2] \sum_{i=1}^n \mathbf{z}_i \mathbf{z}'_i \\ &= \frac{d^2}{4\sigma^4} \mathbb{E} [(\varepsilon_i^2 - \sigma^2)^2] \mathbf{Z}' \mathbf{Z} \end{aligned}$$

Wykorzystując fakt, że $E \left[(\varepsilon_i^2 - \sigma^2)^2 \right] = \text{plim} \left[n^{-1} \sum_{i=1}^n (e_i^2 - \tilde{\sigma}^2)^2 \right]$ otrzymujemy

$$\begin{aligned} \mathbf{I}(\tilde{\boldsymbol{\theta}}_R) &= \frac{d^2}{4\tilde{\sigma}^4} \mathbf{Z}' \mathbf{Z} \left[n^{-1} \sum_{i=1}^n (e_i^2 - \tilde{\sigma}^2)^2 \right] \\ &= \frac{d^2}{4n\tilde{\sigma}^4} \mathbf{Z}' \mathbf{Z} \sum_{i=1}^n (e_i^2 - \tilde{\sigma}^2)^2 = \frac{d^2}{4n} \mathbf{w}' \mathbf{w} \mathbf{Z}' \mathbf{Z} \end{aligned}$$

Przy wykorzystaniu pierwszego wzoru na macierz informacyjną statystykę testu LM ma postać:

$$\begin{aligned} LM &= \mathbf{l}' \tilde{\mathbf{G}}_R [\mathbf{I}(\boldsymbol{\theta})]^{-1} \tilde{\mathbf{G}}_R' \mathbf{l} = \frac{1}{2} \mathbf{w}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{w} \\ &= \frac{1}{2} \left[\mathbf{w}' \mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \right] \left[\mathbf{Z} (\mathbf{Z}' \mathbf{Z})^{-1} \mathbf{Z}' \mathbf{w} \right] \\ &= \frac{1}{2} \hat{\mathbf{w}}' \hat{\mathbf{w}} = \frac{1}{2} \sum_{i=1}^n \hat{w}_i^2 = \frac{1}{2} ESS \end{aligned}$$

gdzie \hat{w}_i są wartościami teoretycznymi z regresji $\frac{e_i^2}{\tilde{\sigma}^2} - 1$ na z_i . Ostatnia równość zachodzi ponieważ $\bar{\hat{w}} = 0$. Ponieważ ESS nie zmienia się jeśli dodamy stałą do zmiennej objaśnianej, więc identyczną wielkość statystyki odzyskamy przeprowadzając regresję $\frac{e_i^2}{\tilde{\sigma}^2}$ na samym z_i . Drugi wzór na statystykę otrzymamy wykorzystując drugi wzór na macierz informacyjną

$$LM = \mathbf{l}' \tilde{\mathbf{G}}_R [\mathbf{I}(\boldsymbol{\theta})]^{-1} \tilde{\mathbf{G}}_R' \mathbf{l} = n \frac{\mathbf{w}' \mathbf{Z} [\mathbf{Z}' \mathbf{Z}]^{-1} \mathbf{Z} \mathbf{w}}{\mathbf{w}' \mathbf{w}} = n \frac{\hat{\mathbf{w}}' \hat{\mathbf{w}}}{\mathbf{w}' \mathbf{w}} = nR^2,$$

gdzie R^2 pochodzi z regresji $\frac{e_i^2}{\tilde{\sigma}^2}$ na z_i . Identyczne R^2 uzyskamy też przeprowadzając regresję e_i^2 na z_i . Ponieważ przy wyprowadzeniu drugiego wzoru na statystykę testową nie korzystaliśmy z

właŝności rozkładu normalnego, więc jest ona bardziej odporna na odchylenia prawdziwego rozkładu ε_i od rozkładu normalnego.

Zamiast nieznanego $\mathbf{I}(\boldsymbol{\theta})$ wykorzystujemy zazwyczaj estymator $\tilde{\mathbf{I}}(\tilde{\boldsymbol{\theta}}_R) = n\tilde{\mathbf{i}}(\tilde{\boldsymbol{\theta}}_R)$, gdzie $\tilde{\mathbf{i}}(\tilde{\boldsymbol{\theta}}_R)$ liczymy z wariantu wzoru (10.12):

$$\tilde{\mathbf{i}}(\tilde{\boldsymbol{\theta}}_R) = n^{-1} \sum_{i=1}^n \frac{\partial \ell_i(\tilde{\boldsymbol{\theta}}_R)}{\partial \boldsymbol{\theta}} \frac{\partial \ell_i(\tilde{\boldsymbol{\theta}}_R)}{\partial \boldsymbol{\theta}'} \xrightarrow{p} \mathbf{i}(\boldsymbol{\theta}) \quad (10.28)$$

Do policzenia statystyki LM wystarczy policzyć analitycznie $\frac{\partial \ell_i(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}}$, numerycznie estymator $\tilde{\boldsymbol{\theta}}_R$ zastosować je do policzenia (10.28) i formy kwadratowej z twierdzenia 10.27.

Uwaga 10.20 Definiując podobnie jak w uwadze (10.9) macierz

$$\tilde{\mathbf{G}}_R = \left(D\ell_1(\tilde{\boldsymbol{\theta}}_R), \dots, D\ell_n(\tilde{\boldsymbol{\theta}}_R) \right)'$$

można zapisać estymator BHHH macierzy informacyjnej jako

$$\tilde{\mathbf{I}}(\tilde{\boldsymbol{\theta}}_R) = \tilde{\mathbf{G}}_R' \tilde{\mathbf{G}}_R$$

a statystykę testu LM jako

$$LM = \mathbf{l}' \tilde{\mathbf{G}}_R (\mathbf{G}'_R \mathbf{G}_R)^{-1} \tilde{\mathbf{G}}_R' \mathbf{l} = nR^2$$

gdzie $R^2 = \frac{\hat{y}'\hat{y}}{y'y}$ jest niecentrowanym współczynnikiem determinacji w regresji wektora jedynek na macierzy \mathbf{G}_R . Można pokazać, że w tej regresji $nR^2 = n - RSS$

Dowód.

$$\begin{aligned} \mathbf{l}' \mathbf{G}'_R (\mathbf{G}'_R \mathbf{G}_R)^{-1} \mathbf{G}_R \mathbf{l} &= \mathbf{l}' \mathbf{G}_R (\mathbf{G}'_R \mathbf{G}_R)^{-1} \mathbf{G}'_R \mathbf{G}_R (\mathbf{G}'_R \mathbf{G}_R)^{-1} \mathbf{G}'_R \mathbf{l} \\ &= n \frac{\widehat{\mathbf{l}}' \widehat{\mathbf{l}}}{\widehat{\mathbf{l}}' \widehat{\mathbf{l}}} = nR^2 \end{aligned}$$

gdzie $\widehat{\mathbf{l}} = \mathbf{G}_R (\mathbf{G}'_R \mathbf{G}_R)^{-1} \mathbf{G}'_R \mathbf{l}$ są wartościami teoretycznymi \mathbf{l} z regresji na \mathbf{G}_R . Zauważmy, że w tej regresji warunek konieczny minimalizacji sumy kwadratów reszt $\mathbf{G}'_R \mathbf{e} = \mathbf{G}'_R (\mathbf{l} - \widehat{\mathbf{l}}) = \mathbf{0}$ można sprowadza się do warunku $\mathbf{G}'_R \mathbf{l} = \mathbf{G}'_R \widehat{\mathbf{l}}$. Suma kwadratów reszt ma w tym przypadku postać:

$$\begin{aligned} RSS &= (\mathbf{l} - \widehat{\mathbf{l}})' (\mathbf{l} - \widehat{\mathbf{l}}) = \mathbf{l}' \mathbf{l} - 2\mathbf{l}' \widehat{\mathbf{l}} + \widehat{\mathbf{l}}' \widehat{\mathbf{l}} \\ &= n - 2\mathbf{l}' \mathbf{G}_R (\mathbf{G}'_R \mathbf{G}_R)^{-1} \mathbf{G}'_R \mathbf{l} + \widehat{\mathbf{l}}' \widehat{\mathbf{l}} \\ &= n - 2\widehat{\mathbf{l}}' \mathbf{G}_R (\mathbf{G}'_R \mathbf{G}_R)^{-1} \mathbf{G}'_R \mathbf{l} + \widehat{\mathbf{l}}' \widehat{\mathbf{l}} \\ &= n - \widehat{\mathbf{l}}' \widehat{\mathbf{l}} \end{aligned}$$

a więc rzeczywiście $nR^2 = \widehat{\mathbf{l}}' \widehat{\mathbf{l}} = n - RSS$ ■

Poniżej w wielu pakietach ekonometrycznych nie ma możliwości uzyskania wielkości niecentrowanego R^2 , statystykę LM w tym przypadku można uzyskać zauważając, że w rozważanej regresji $nR^2 = n - RSS$

- **Statystyka ilorazu wiarygodności (LR)**

Najłatwiejszą metodą testowania jest zastosowanie testu ilorazu wiarygodności, którego rozkład dany jest następującym twierdzeniem

Twierdzenie 10.21 Przy spełnionych założeniach twierdzenia 10.5 i założeniu (10.15)

$$LR = 2 \left(\ell \left(\tilde{\boldsymbol{\theta}} \right) - \ell \left(\tilde{\boldsymbol{\theta}}_R \right) \right) \xrightarrow{D} \chi_g^2, \quad (10.29)$$

gdzie $g = \text{Rank} \left(\mathbf{H} \left(\boldsymbol{\theta} \right) \right)$.

Dowód. Na podstawie wzoru (10.22) oraz wzoru (10.23) na dystybuantę $\mathbf{h} \left(\tilde{\boldsymbol{\theta}} \right)$ można stwierdzić, że

$$\sqrt{n} \left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_R \right) \xrightarrow{D} N \left(0, \mathbf{Q} \left(\boldsymbol{\theta} \right) \right)$$

gdzie

$$\mathbf{Q} \left(\boldsymbol{\theta} \right) = \mathbf{i}^{-1} \left(\boldsymbol{\theta} \right) \mathbf{H}' \left(\boldsymbol{\theta} \right) \left[\mathbf{H} \left(\boldsymbol{\theta} \right) \mathbf{i}^{-1} \left(\boldsymbol{\theta} \right) \mathbf{H}' \left(\boldsymbol{\theta} \right) \right]^{-1} \mathbf{H} \left(\boldsymbol{\theta} \right) \mathbf{i}^{-1} \left(\boldsymbol{\theta} \right) = \mathbf{i}^{-1} \left(\boldsymbol{\theta} \right) - \mathbf{P} \left(\boldsymbol{\theta} \right)$$

Dodatkowo

$$\mathbf{Q} \left(\boldsymbol{\theta} \right) \mathbf{H}' \left(\boldsymbol{\theta} \right) \left[\mathbf{H} \left(\boldsymbol{\theta} \right) \mathbf{i}^{-1} \left(\boldsymbol{\theta} \right) \mathbf{H}' \left(\boldsymbol{\theta} \right) \right]^{-1} \mathbf{H} \left(\boldsymbol{\theta} \right) \mathbf{Q} \left(\boldsymbol{\theta} \right) = \mathbf{Q} \left(\boldsymbol{\theta} \right),$$

a więc $\mathbf{H}' \left(\boldsymbol{\theta} \right) \left[\mathbf{H} \left(\boldsymbol{\theta} \right) \mathbf{i}^{-1} \left(\boldsymbol{\theta} \right) \mathbf{H}' \left(\boldsymbol{\theta} \right) \right]^{-1} \mathbf{H} \left(\boldsymbol{\theta} \right)$ jest uogólnoną odwrotnością $\mathbf{Q} \left(\boldsymbol{\theta} \right)$. Z twierdzenia 17.80 wynika, że

$$\left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_R \right)' \mathbf{H}' \left(\boldsymbol{\theta} \right) \left[\mathbf{H} \left(\boldsymbol{\theta} \right) \mathbf{I}^{-1} \left(\boldsymbol{\theta} \right) \mathbf{H}' \left(\boldsymbol{\theta} \right) \right]^{-1} \mathbf{H} \left(\boldsymbol{\theta} \right) \left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_R \right) \xrightarrow{D} \chi_g^2 \quad (10.30)$$

Rozwinięcie Taylora $\ell \left(\tilde{\boldsymbol{\theta}} \right)$ wokół $\tilde{\boldsymbol{\theta}}_R$ ma postać

$$\ell \left(\tilde{\boldsymbol{\theta}} \right) = \ell \left(\tilde{\boldsymbol{\theta}}_R \right) + \frac{1}{2} \left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_R \right)' \left[D^2 \ell \left(\tilde{\boldsymbol{\theta}}_R \right) \right] \left(\tilde{\boldsymbol{\theta}} - \tilde{\boldsymbol{\theta}}_R \right) + o_p \left(1 \right)$$

a więc aproksymację ilorazu wiarygodności można zapisać jako

$$LR = \left(\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}} \right)' \left[D^2 \ell \left(\tilde{\boldsymbol{\theta}}_R \right) \right] \left(\tilde{\boldsymbol{\theta}}_R - \tilde{\boldsymbol{\theta}} \right) + o_p \left(1 \right) \quad (10.31)$$

Na podstawie wzorów (10.26) i (10.25), możemy stwierdzić, że

$$n \left[D^2 \ell \left(\tilde{\theta}_R \right) \right] = n \left[D \ell \left(\tilde{\theta}_R \right) D \ell \left(\tilde{\theta}_R \right)' \right] \xrightarrow{p} \mathbf{H}'(\theta) \left[\mathbf{H}(\theta) \mathbf{i}^{-1}(\theta) \mathbf{H}'(\theta) \right]^{-1} \mathbf{H}(\theta)$$

a na podstawie wzoru (10.30) i wzoru (10.13) wywnioskować, że forma kwadratowa (10.31) dąży do χ_g^2 . ■

W celu policzenia statystyki LR wystarczy policzyć wartość funkcji wiarygodności dla $\tilde{\theta}$ i $\tilde{\theta}_R$, nie ma potrzeby liczenia macierzy estymatora macierzy informacyjnej $\mathbf{I}(\theta)$.

10.7. Równoczesność i egzogeniczność w MNW

Teoretyczne sformułowanie warunków egzogeniczności jest możliwe jeśli zdefiniujemy Proces Generujący Dane (PGD), jako proces, który generuje obserwowane przez nas dane. Definicję słabej egzogeniczności \mathbf{X}_2 względem parametru θ_1 można sformułować w kategoriach własności PGD.

Definicja 10.22 *W procesie losowym $\mathbf{X}_t = (\mathbf{X}_{1t}, \mathbf{X}_{2t})$ generowanym przez PGD, dla którego łączna dystrybucja \mathbf{X}_t jest równa $F_X(\mathbf{X}_t | \mathbf{X}_{-1}, \theta)$, warunkowa dystrybucja \mathbf{X}_{1t} jest równa $F_{\mathbf{X}_1 | \mathbf{X}_2}(\mathbf{X}_{1t} | \mathbf{X}_{-1}, \theta)$, a dystrybucja brzegowa \mathbf{X}_{2t} $F_{\mathbf{X}_2}(\mathbf{X}_{2t} | \mathbf{X}_{-1}, \theta_2)$, gdzie $\theta = (\theta_1, \theta_2)$, $\mathbf{X}_{-1} = (\mathbf{X}_0, \dots, \mathbf{X}_{t-1})$ o m wymiarowym wektorze zmiennych \mathbf{X}_2 mówimy, że jest słabo egzogeniczny względem parametru θ_1 jeśli:*

1. PGD tego procesu można zdekomponować w taki sposób, że

$$F_X(\mathbf{X}_t | \mathbf{X}_{-1}, \theta) = F_{\mathbf{X}_1 | \mathbf{X}_2}(\mathbf{X}_{1t} | \mathbf{X}_{-1}, \theta_1) F_{\mathbf{X}_2}(\mathbf{X}_{2t} | \mathbf{X}_{-1}, \theta_2), \text{ dla } t = 1, \dots, T$$

2. Zbiory dopuszczalnych wartości parametrów $\theta_1 \in \Theta_1$, $\theta_2 \in \Theta_2$ są wzajemnie niezależne (niezmiennicze).

Warunki te implikują, że dystrybucja brzegowa \mathbf{X}_{2t} nie zależy od θ_1 i tym samym posługiwanie się w procesie estymacji dystrybucją warunkową \mathbf{X}_{1t} względem \mathbf{X}_{2t} nie prowadzi do utraty

informacji na temat θ_1 . Estymacja wektora parametrów θ_1 na podstawie procesu warunkowego jest więc równie efektywna jak estymacja na podstawie procesu bezwarunkowego. Definicja słabej egzogeniczności ma sens jedynie, gdy mówimy o egzogeniczności X_{2t} względem pewnego wektora parametrów θ_1 . Definicja słabej egzogeniczności jest ogólniejsza od definicji zmiennych z góry ustalonych i można ją stosować dla szerszej klasy modeli.

Zauważmy, że jeśli X_2 jest słabo egzogeniczne względem θ_1 , to w myśl definicji 10.22 logarytm funkcji wiarygodności przyjmuje postać

$$\begin{aligned} \ell(\theta) &= \ln \left(\prod_{t=1}^T F_X(\mathbf{X}_t | \mathbf{X}_{-1}, \theta) \right) \\ &= \sum_{t=1}^T \ln F_{X_1|X_2}(\mathbf{X}_{1t} | \mathbf{X}_{-1}, \mathbf{X}_{2t}, \theta_1) + \sum_{t=1}^T \ln F_{X_2}(\mathbf{X}_{2t} | \mathbf{X}_{-1}, \theta_2) \end{aligned}$$

Warunki pierwszego rzędu na maksymalizację logarytmu funkcji wiarygodności względem θ_1 przyjmują postać

$$\frac{\partial \ell(\theta)}{\partial \theta_1} = \sum_{t=1}^T \frac{f_{X_1|X_2}(\mathbf{X}_{1t} | \mathbf{X}_{-1}, \mathbf{X}_{2t}, \theta_1)}{F_{X_1|X_2}(\mathbf{X}_{1t} | \mathbf{X}_{-1}, \mathbf{X}_{2t}, \theta_1)} = 0$$

Postać tych równań zależy jedynie od postaci funkcji $F_{X_1|X_2}(\mathbf{X}_{1t} | \mathbf{X}_{-1}, \mathbf{X}_{2t}, \theta_1)$. Wynika z tego, że jeśli zmienna X_{2t} jest słabo egzogeniczna względem parametru θ_1 , to rzeczywiście postać asymptotycznie efektywnego estymatora *MNW* tego parametru zależęć będzie jedynie od postaci procesu warunkowego

10.8. Kryteria informacyjne

W przypadku modeli szacowanych metodą *MNW* takich jak model probitowy niemożliwe jest zdefiniowanie \bar{R}^2 . Pojawia się pytanie, czy możliwe jest zdefiniowanie statystyki, która mogłaby

służyć badaniu jakości dopasowania, która uwzględniałaby ilość traconych w procesie estymacji stopni swobody. Statystyki takie istnieją i noszą ogólną nazwę kryteriów informacyjnych. Najbardziej popularnymi z nich jest kryterium informacyjne Akaike *AIC* (**A**kaike **I**nformation **C**riterion) i Bayesowskie kryterium informacyjne Schwartza *BIC* (**B**ayes **I**nformation **C**riterion) dane wzorami

$$BIC = -2n^{-1}\ell(\hat{\theta}) + 2n^{-1}K$$

$$AIC = -2n^{-1}\ell(\hat{\theta}) + n^{-1}K \ln n$$

Za najlepszy uznaje się ten model, dla którego kryterium informacyjne uzyskuje najniższą wartość. Pokazano, że dla pewnych ogólnych założeń, dla $n \rightarrow \infty$ wybrany na podstawie *BIC* model zawierać będzie poprawny zbiór zmiennych objaśniających. W przypadku *AIC* okazuje się, że zbiór ten nawet dla $n \rightarrow \infty$ może być zbyt duży.

Często nie jest możliwe znalezienie analitycznych rozwiązań dla warunków pierwszego rzędu. Na szczęście obecnie istnieje wiele programów ekonometrycznych, w których oprogramowane są numeryczne procedury maksymalizacji funkcji wiarygodności. Przykładowo w programie PcGive wystarczy podać postać logarytmu funkcji wiarygodności, aby dla konkretnego zbioru obserwacji uzyskać oszacowania szukanych parametrów.

Literatura: ? str. 121-135, ? str. 129-140, ? str. 43-46, ? str. 170-184 i 236-237, ? str. 109-111 i 118-120, ? str. 196-217.

10.9. Modele dla dyskretnych zmiennych zależnych i prób nielosowych

W wielu przypadkach analizowanych w ramach modeli ekonometrycznych zmienna zależna nie jest ciągła ale przyjmuje wartości z pewnego przeliczalnego i skończonego zbioru wartości. Przykładowo w ankietach dotyczących ekonomicznej aktywności ludności pojawia się pytanie o wykształcenie. Poziom wykształcenia kodowany jest jako jedna z siedmiu możliwości i tym samym jest

zmienną dyskretną. W takim przypadku nie można założyć, że $y_i = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$, gdzie ε_i jest pewną ciągłą zmienną losową, ponieważ model taki byłby niezgodny z obserwacjami na y_i .

Najprostsza sytuacja ma miejsce, gdy interesujące nas zjawisko może przyjąć dwa stany (np. respondent może mieć pracę lub być bezrobotny). Prawdopodobieństwo, że zjawisko znajdzie się w określonym stanie zależy od pewnych zmiennych endogenicznych (np. charakterystyk społeczno-demograficznych respondentów). W takim przypadku interesuje nas oszacowanie wpływu charakterystyk respondenta na prawdopodobieństwo, że ma on pracę. Zazwyczaj zakładamy, że prawdopodobieństwo zajścia tych zdarzeń ma następującą formę

$$\Pr(y_i) = \begin{cases} 1 - F(\mathbf{x}_i \boldsymbol{\beta}) & \text{dla } y_i = 0 \\ F(\mathbf{x}_i \boldsymbol{\beta}) & \text{dla } y_i = 1 \end{cases}$$

gdzie $F(\cdot)$ jest dystrybuantą pewnego rozkładu prawdopodobieństwa. W zależności od tego jaką dystrybuantę przyjmiemy za $F(\cdot)$ otrzymamy model probitowy i logitowy. Do liczenia estymatorów parametru $\boldsymbol{\beta}$ używa się *MNW*. Ponieważ analityczne rozwiązanie równań wynikających z równań pierwszego rzędu jest zazwyczaj zbyt trudne lub niemożliwe, więc w praktyce stosuje się numeryczne metody maksymalizacji.

• Model probitowy

W przypadku modelu probitowego, że dystrybuanta $F(\mathbf{x}_i \boldsymbol{\beta}) = \Phi(\mathbf{x}_i \boldsymbol{\beta})$, czyli jest równa rozkładowi normalnemu. Funkcja prawdopodobieństwa zajścia pojedynczego zdarzenia jest dana wzorem

$$\Pr(y_i) = \begin{cases} 1 - \Phi(\mathbf{x}_i \boldsymbol{\beta}) & \text{dla } y_i = 0 \\ \Phi(\mathbf{x}_i \boldsymbol{\beta}) & \text{dla } y_i = 1 \end{cases} = (1 - \Phi(\mathbf{x}_i \boldsymbol{\beta}))^{1-y_i} \Phi(\mathbf{x}_i \boldsymbol{\beta})^{y_i}$$

Dla takiej funkcji prawdopodobieństwa pojedynczego zdarzenia funkcja wiarygodności ma postać

$$L(\mathbf{y}, \mathbf{x}, \boldsymbol{\beta}) = \prod_{i=1}^n (1 - \Phi(\mathbf{x}_i \boldsymbol{\beta}))^{1-y_i} \Phi(\mathbf{x}_i \boldsymbol{\beta})^{y_i}$$

zaś logarytm funkcji wiarygodności będzie miał postać

$$\ell(\mathbf{y}, \mathbf{x}, \boldsymbol{\beta}) = \sum_{i=1}^n [(1 - y_i) \ln(1 - \Phi(\mathbf{x}_i \boldsymbol{\beta})) + y_i \ln \Phi(\mathbf{x}_i \boldsymbol{\beta})]$$

W przypadku wielu programów ekonometrycznych do policzenia metodami numerycznymi wartości estymatorów dla konkretnego zbioru obserwacji wystarczy podać formę analityczną logarytmu funkcji gęstości dla poszczególnej obserwacji, który równy jest w tym przypadku $\ell(\mathbf{y}_i, \mathbf{x}_i, \boldsymbol{\beta}) = (1 - y_i) \ln(1 -$

Program w PCGIVE:

```
actual = vaso;
```

```
xbeta = &0 + &1 * Lrate + &2 * Lvolume;
```

```
fitted = 1 - tailnormal(xbeta);
```

```
loglik = actual * log(fitted) + (1 - actual) * log(1 - fitted);
```

• Model logitowy

Model logitowy otrzymujemy jeśli przyjmiemy, że $F(\mathbf{x}_i \boldsymbol{\beta}) = \frac{e^{\mathbf{x}_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}}}$, a więc:

$$\Pr(y_i) = \begin{cases} \frac{1}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}}} & \text{dla } y_i = 0 \\ \frac{e^{\mathbf{x}_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}}} & \text{dla } y_i = 1 \end{cases}$$

w tym przypadku postać logarytmu funkcji wiarygodności jest następująca

$$\ell(\mathbf{y}, \mathbf{x}, \boldsymbol{\beta}) = \sum_{i=1}^n \left[(1 - y_i) \ln \left(\frac{1}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}}} \right) + y_i \ln \left(\frac{e^{\mathbf{x}_i \boldsymbol{\beta}}}{1 + e^{\mathbf{x}_i \boldsymbol{\beta}}} \right) \right]$$

Program w PCGIVE:

```
actual = vaso;
```

```
xbeta = &0 + &1 * Lrate + &2 * Lvolume;
```

```
fitted = 1 / (1 + exp(xbeta));
loglik = (1 - actual) * log(fitted) + actual * log(1 - fitted);
```

• Wielomianowy logit

W wielu przypadkach respondenci mają więcej niż dwa możliwe warianty odpowiedzi. W takim przypadku możemy zastosować wielomianowego logita. Dla przypadku, kiedy jest $K+1$ wariantów odpowiedzi model ten dany jest następującą funkcją prawdopodobieństwa:

$$\Pr(y_i = k) = \frac{e^{\mathbf{x}_i \beta_k}}{\sum_{k=0}^K e^{\mathbf{x}_i \beta_k}} \quad (10.32)$$

Zakładamy, że warianty nie mogą być w żaden sposób uporządkowane². Tak sformułowany model nie daje jednoznacznego rozwiązania dla maksimum funkcji wiarygodności, ponieważ dzieląc licznik i mianownik przez dowolne $e^{\mathbf{x}_i \beta}$ otrzymujemy model

$$\Pr(y_i = k) = \frac{\frac{e^{\mathbf{x}_i \beta_k}}{e^{\mathbf{x}_i \beta}}}{\frac{\sum_{k=0}^K e^{\mathbf{x}_i \beta_k}}{e^{\mathbf{x}_i \beta}}} = \frac{e^{\mathbf{x}_i (\beta_k - \beta)}}{\sum_{k=0}^K e^{\mathbf{x}_i (\beta_k - \beta)}}$$

o tej samej funkcji wiarygodności ale parametrach $\hat{\beta}_k = \beta_k - \beta \neq \beta_k$. Tak sformułowany model nie jest więc zidentyfikowany. Próba zmaksymalizacji funkcji wiarygodności (10.32) najprawdopodobniej spowoduje problemy ze zbieżnością procedur numerycznych. Po to, by uzyskać jednoznaczne rozwiązanie konieczne jest narzucenie ograniczeń na parametru modelu (10.32). Problem z identyfikacją bierze się w z tego, że nie mamy ustalonego punktu odniesienia, takiego jakim było $y_i = 0$ (zdarzenie nie zaszło) dla przypadku wyboru binarnego. Jeśli jednak ustalimy, że interesują nas zmiany stosunku między prawdopodobieństwem wariantu $y_i = 0$ a prawdopodobieństwami

²Z tego powodu model ten nazywa się także modelem dla wyborów nieuporządkowanych (*unordered choice set*).

pozostałych wariantów, to możemy przyjąć, że $\beta_0 = \mathbf{0}$ i osiągnąć dzięki temu jednoznaczny identyfikację parametrów.

Estymowany model będzie miał teraz postać

$$\Pr(y_i = k) = \begin{cases} \frac{1}{1 + \sum_{j=1}^K e^{x_i \beta_j}} & \text{dla } y_i = 0 \\ \frac{e^{x_i \beta_k}}{1 + \sum_{j=1}^K e^{x_i \beta_j}} & \text{dla } 1 \leq y_i \leq K \end{cases}$$

Dodatni s -ty element wektora β_k będzie w tej sytuacji oznaczał, że prawdopodobieństwo k -tego wyboru w stosunku do prawdopodobieństwa wyboru zerowego w przypadku wzrośnie, gdy zwiększymy element s -ty wektora x .

• Logit i probit dla uporządkowanych zmiennych dyskretnych

W wielu przypadkach zmienne dyskretne mają pewne logiczne uporządkowanie. Na przykład w przypadku wykształcenia logiczne jest uporządkowanie odpowiedzi od wykształcenia wyższego aż do niepełnego podstawowego. Czy takie logiczne uporządkowanie może być wykorzystane w trakcie esymacji modelu? Zauważmy, że w przypadku wielomianowego logita nie miało ono najmniejszego znaczenia. Jest jednak możliwe zbudowanie funkcji prawdopodobieństwa, dla których uporządkowanie takie będzie miało znaczenie. Do modelowania prawdopodobieństwa zajścia jednego z $K + 1$ możliwości, używamy w takim wypadku funkcji prawdopodobieństwa

$$\Pr(y_i = k) = \begin{cases} F(-x_i \beta) & \text{dla } y_i = 0 \\ F(\mu_1 - x_i \beta) - F(-x_i \beta) & \text{dla } y_i = 1 \\ F(\mu_2 - x_i \beta) - F(\mu_1 - x_i \beta) & \text{dla } y_i = 2 \\ \vdots & \vdots \\ 1 - F(\mu_{K-1} - x_i \beta) & \text{dla } y_i = K \end{cases}$$

Jeśli zastosujemy w tym modelu jako $F(\cdot)$ rozkład normalny wtedy otrzymamy probit dla uporządkowanych zmiennych zależnych. W przypadku zastosowania rozkładu wykładniczego otrzymamy

w analogiczny model logitowy.

• Modele dla liczebności

Często dane, którymi dysponujemy pochodzą ze zbioru liczb naturalnych przy czym najbardziej prawdopodobne są niskie wartości zmiennej zależnej. Takimi zmiennymi może być na przykład odpowiedź na pytanie jak często dana osoba zmieniała pracę w danym roku. Przybliżenie tego rodzaju zmiennych za pomocą *KMRL* wydaje się nieprawidłowe a z drugiej strony zastosowanie modeli dla uporządkowanych zmiennych zależnych też nie jest w pełni uzasadnione ponieważ, przynajmniej teoretycznie, ilość możliwych odpowiedzi na to pytanie jest nieograniczona. W takim przypadku można zastosować model regresji Poissona:

$$\Pr(y_i = k) = \frac{e^{-\lambda_i} \lambda_i^{y_i}}{k!}$$

gdzie najczęściej przyjmuje się, że λ_i pochodzi z modelu log-liniowego

$$\ln \lambda_i = \beta \mathbf{x}_i$$

Ponieważ dla rozkładu Poissona $E[y_i | x_i] = \text{Var}[y_i | x_i] = \lambda_i = e^{\beta \mathbf{x}_i}$ więc $\frac{\partial E[y_i | x_i]}{\partial x_i} = \lambda_i \beta$. Podstawowym zastrzeżeniem jakie formułowane jest w stosunku do tego modelu są implikacje założenia, wynikające z własności rozkładu Poissona, że wartość oczekiwana jest równa wariancji.

• Próby ucięte

W niektórych badaniach ekonomicznych próba obejmuje jedynie respondentów spełniających pewne kryteria. Na przykład przeprowadza się niekiedy badania jedynie najbiedniejszych gospodarstw domowych. W tym przypadku, jeśli modelujemy dochód, to musimy wziąć pod uwagę, że zmienna zależna nie może przyjmować wartości większych niż ta, która została przyjęta za granicę, powyżej której gospodarstwa nie wchodzi do badania. Próbę taką nazywamy uciętą (*truncated*), ponieważ

obserwacje, dla których zmienna ta przyjmuje wartość wyższą niż pewna graniczna wartość zostają usunięte z próby. W praktyce można spotkać zarówno zmienne obcięte z dołu jak i z góry.

Dla *KMRL* przyjmujemy, że $y_i = \mathbf{x}_i\boldsymbol{\beta} + \varepsilon_i$ i $\varepsilon_i \sim N(0, \sigma^2)$ co implikuje, że rozkład $\Pr(Y_i < y_i) = \Phi\left(\frac{\mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)$. Jeśli zmienna zależna została obcięta z dołu w punkcie y_* , wtedy rozkład warunkowy

$$\Pr(Y_i < y_i | y_i > y_*) = \frac{\Phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(\frac{y_* - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}$$

Funkcja gęstości dla takiego rozkładu prawdopodobieństwa ma postać

$$f(y_i | y_i > y_*) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{1 - \Phi\left(\frac{y_* - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}$$

Podobnie w przypadku zmiennej uciętej z góry mamy

$$f(y_i | y_i < y_*) = \frac{\frac{1}{\sigma}\phi\left(\frac{y_i - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}{\Phi\left(\frac{y_* - \mathbf{x}_i\boldsymbol{\beta}}{\sigma}\right)}$$

Oczywiste uogólnienie tego rodzaju modelu można zbudować dla zmiennych obciętych z góry i z dołu oraz dla innych niż normalny rozkładów prawdopodobieństwa na przykład rozkładu jednostajnego.

● **Ocenzurowane zmienne zależne (tobit)**

Problem ocenzurowanych (*censored*) zmiennych zależnych pojawia się, gdy wartości tych zmiennych muszą być większe lub równe zero, a w próbie znajduje się wiele obserwacji równych zero.

Na przykład w badaniach gospodarstw domowych wydatki na poszczególne kategorie dóbr są zawsze większe lub równe zero. W praktyce jednak okazuje się, że w danym roku wiele gospodarstw ma wydatki na określoną kategorię dóbr (np. sprzęt RTV) równe zero. Wyraźnie widać, że tego rodzaju obserwacje nie mogą pochodzić z *KMRL*, ponieważ dla tego modelu nieprawdopodobne jest uzyskanie wielu obserwacji równych 0.

Z tego powodu przyjmuje się, że w takim przypadku istnieje zmienna ukryta (*latent variable*) y_i^* , która pochodzi z *KMRL*, ale następnie podlega obcięciu:

$$y_i^* = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$$

$$y_i = \begin{cases} 0 & \text{dla } y_i^* \leq 0 \\ y_i^* & \text{dla } y_i^* > 0 \end{cases}$$

Obserwujemy jedynie y_i , zaś na temat charakterystyk y_i^* możemy wnioskować jedynie pośrednio. Najczęściej przyjmujemy, że ε_i ma rozkład normalny. W takim przypadku funkcja wiarygodności będzie miała postać

$$L = \prod_{y_i > 0} \left[\phi \left(\frac{y_i - \mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) \right] \prod_{y_i = 0} \left[1 - \Phi \left(\frac{\mathbf{x}_i \boldsymbol{\beta}}{\sigma} \right) \right]$$

Wprowadzając niewielkie modyfikacje, można zdefiniować tobita, który jest obcięty w innej wartości niż zero, jest obcięty z góry, lub też obcięty z góry i z dołu.

• Modele przełącznikowe

Modele przełącznikowe są to modele, w których postać równania regresji zależy od dodatkowego równania, które "przełącza" model między różnymi dwoma lub więcej stanami. Typowa postać modelu przełącznikowego ma postać

$$\begin{cases} y_t = \mathbf{X}_{1t} \boldsymbol{\beta}_1 + u_{1t} & \text{model w stanie 1} \\ y_t = \mathbf{X}_{2t} \boldsymbol{\beta}_2 + u_{2t} & \text{model w stanie 2} \end{cases} \quad (10.33)$$

Wybór stanu dokonuje się na podstawie wielkości zmiennej indykatowej:

$$I_t = \begin{cases} 1 & \text{model w stanie 1} \\ 0 & \text{model w stanie 2} \end{cases}$$

W wielu przypadkach wartości zmiennej indykatowej są znane i można wtedy wyliczyć estymatory β_1, β_2 stosując *MNK* dla osobno dla obserwacji z pierwszego i drugiego reżimu. Możliwe jest także, że nieznane są wartości zmiennej indykatowej ale wiemy od jakich zmiennych ona zależy. Definiujemy wtedy równanie zmiennej ukrytej (*latent variable*) o nieznanach wartościach

$$z_t = \mathbf{X}_{3t}\beta_3 + u_{3t} \quad (10.34)$$

i przyjmujemy, że zmienna indykatowa przyjmuje wartości zgodnie z równaniem

$$I_t = \begin{cases} 1 & \text{jeśli } z_t > 0 \\ 0 & \text{jeśli } z_t \leq 0 \end{cases} \quad (10.35)$$

Ponieważ w tym przypadku interesuje nas jedynie znak zmiennej z_t , więc normalizujemy ją zwykle tak, że $z'z = 1$. Postać macierzy wariancji u_1, u_2, u_3 będzie następująca

$$\Sigma = \begin{bmatrix} \sigma_{11} & \sigma_{12} & \sigma_{13} \\ \sigma_{21} & \sigma_{22} & \sigma_{23} \\ \sigma_{31} & \sigma_{32} & \sigma_{33} \end{bmatrix}$$

Jeśli $\sigma_{31} = \sigma_{32} = 0$ wtedy mówimy o modelu z przełączaniem egzogenicznym, zaś jeśli któraś z tych kowariancji nie jest równa zero mówimy o modelu z przełączaniem endogenicznym. Za-uważmy najpierw, że w przypadku, kiedy prawdziwy jest model z przełączaniem endogenicznym niemożliwe jest uzyskanie zgodnych estymatorów za pomocą *MNK* nawet wtedy, gdy znane są wartości zmiennej I_t . Model przełącznikowy, ma też tą interesującą cechę, że nie można go oszacować za pomocą *MNW*.

• Selekcja próby (heckit)

Kolejną problemem, z którym niekiedy się spotykamy jest problem nieprzypadkowej selekcji próby. Zastanówmy się nad przypadkiem, kiedy badamy zależność zarobków z od charakterystyk społeczno-ekonomicznych respondentów. Jest oczywiste, że zależność taką można badać jedynie dla osób posiadających jakieś zatrudnienie. Oczywiście możemy zbudować osobny model tłumaczący dlaczego dana osoba jest bezrobotna. Pojawia się jednak pytanie, czy nieobserwowalne charakterystyki, które stanowią jedną z determinant bezrobocia nie wpływają także na poziom zarobków. Model takich danych składa się zwykle z dwóch równań. Pierwsze równanie, to równanie selekcji, które determinuje czy dana obserwacja znajdzie się w próbie, czy też poza nią. Drugie równanie jest równaniem regresji, które determinuje sposób w jaki kształtowane są obserwowalne obserwacje na zmiennej zależnej.

$$\text{równanie selekcji} \quad w_i = \mathbf{z}_i \boldsymbol{\gamma} + \eta_i$$

$$\text{równanie regresji} \quad y_i | (w_i > 0) = \mathbf{x}_i \boldsymbol{\beta} + \varepsilon_i$$

Estymator *MNK* równania wektora $\boldsymbol{\beta}$ w równaniu regresji będzie asymptotycznie obciążony, jeśli $E(\varepsilon_i, \eta_i) \neq 0$, ponieważ w tym przypadku np. $E(\varepsilon_i | \eta_i > 0) \neq 0$ i w konsekwencji $E(\varepsilon_i | \mathbf{z}_i \boldsymbol{\gamma}) \neq 0$. Jeśli $n^{-1}(\mathbf{X}'\mathbf{Z}) \xrightarrow{p} 0$, wtedy problem obciążenia dotyczy jedynie stałej, jednak dla najczęstszego przypadku, kiedy \mathbf{X} i \mathbf{Z} są skorelowane, korelacja ε_i i \mathbf{z}_i implikuje występowanie korelacji między ε_i i \mathbf{x}_i i powoduje obciążenie całego estymatora \mathbf{b} na skutek występowania równoczesności. Do estymacji takiego modelu można użyć *MNW* zauważając, że dla pojedynczej obserwacji funkcja gęstości jest dana wzorem

$$f(y_i | w_i > 0) = \frac{f(y_i \cup w_i > 0)}{\Pr(w_i > 0)}$$

W przypadku rozkładu normalnego, $\Pr(w_i > 0) = \Phi(\mathbf{z}_i \boldsymbol{\gamma})$ a

$$f(y_i \cup w_i > 0) = \int_{\mathbf{z}_i \boldsymbol{\gamma}}^{\infty} \phi(w, y_i - \mathbf{x}_i \boldsymbol{\beta}) dw$$

gdzie przez $\phi(\cdot, \cdot)$ oznaczamy łączną funkcję gęstość rozkładu normalnego dla dwóch zmiennych, dla którego $E(\varepsilon_i) = E(\eta_i) = 0$ i $\text{Var}(\eta_i) = 1$.³ Prostsza, dwustopniową procedurę zaproponował Heckman. Można udowodnić, że dla rozkładu normalnego

$$E(y_i | w_i > 0) = \mathbf{x}_i \boldsymbol{\beta} + \beta_w \left(\frac{\phi(\mathbf{z}_i \boldsymbol{\gamma})}{\Phi(\mathbf{z}_i \boldsymbol{\gamma})} \right) \quad (10.36)$$

1. Najpierw estymujemy więc za pomocą probitu równanie selekcji i tworzymy wartości teoretyczne $\frac{\phi(\mathbf{z}_i \hat{\boldsymbol{\gamma}})}{\Phi(\mathbf{z}_i \hat{\boldsymbol{\gamma}})}$.
2. Następnie estymujemy równanie (10.36) przy czym za $\frac{\phi(\mathbf{z}_i \boldsymbol{\gamma})}{\Phi(\mathbf{z}_i \boldsymbol{\gamma})}$ podstawiamy wartości uzyskane w pierwszym kroku.

Procedura zaproponowana przez Heckmana daje zbieżny estymator parametru $\boldsymbol{\beta}$, ponieważ estymator uzyskany w pierwszym kroku jest estymatorem zbieżnym.

• Modele dla długości trwania (duration models)

Zainteresowanie modelami długości trwania wzięło się pierwotnie z zastosowań przemysłowych i aktuarialnych. W zastosowaniach inżynierskich chodziło o oszacowanie trwałości urządzeń przemysłowych. Z kolei w ubezpieczeniach życiowych, jednym z najważniejszych czynników determinujących koszt polisy jest oczekiwana długość życia.

Zastosowania ekonomiczne analizy długości trwania dotyczą szacowania długości bezrobocia, długości czasu, który upływa między zakupami dóbr trwałego użytku, długości trwania strajków, itd.

Dane dotyczące długości trwania t_1, \dots, t_n dotyczą długości czasu jaki upłynął między początkiem jakiegoś zdarzenia a jego końcem lub momentem badania. Powoduje to, że problem ocenzurowania

³Jeśli w równaniach selekcji i regresji są stałe to założenie dotyczące wartości oczekiwanych ε_i i η_i są pewną normalizacją. Ponieważ interesuje nas jedynie, czy $w_i > 0$ więc założenie, że $\text{Var}(\eta_i) = 1$ także nie jest związane z ograniczenia ogólności modelu.

danych może istotnie wpływać na wynik badania. Na przykład, od T lat prowadzone są badania nad przeżywalnością ofiar Czarnobyla to oczywiste jest, że dla t_i oznaczające czas, który dana osoba przeżyła od czasu katastrofy będzie zawsze mniejsza od T .

Prawdopodobieństwo, że czas trwania będzie krótszy niż t , przy pewnej funkcji gęstości $f(t)$ jest dany dystrybuantą

$$F(t) = \int_0^t f(s) ds = \Pr(T \leq t)$$

Zazwyczaj analizuje się funkcję przeżywalności (*survival function*), która opisuje prawdopodobieństwo, że czas trwania będzie conajmniej równy t

$$S(t) = 1 - F(t) = \Pr(T > t)$$

Często interesuje nas także pytanie, jakie jest prawdopodobieństwo, że po przetrwaniu t , zjawisko ulegnie zakończeniu w trakcie okresu $t + h$.

$$l(t, \Delta) = \Pr(t < T \leq t + h | T > t)$$

Aby opisać tę cechę dystrybuanty dla granicznie małych Δ posługujemy się funkcją hazardu (*hazard rate*)

$$\lambda(t) = \lim_{\Delta \rightarrow 0} \frac{\Pr(t < T \leq t + h | T > t)}{h} = \lim_{\Delta \rightarrow 0} \frac{F(t + h) - F(t)}{hS(t)} = \frac{f(t)}{S(t)}$$

Funkcja hazardu nie może być interpretowana jako funkcja prawdopodobieństwa, ponieważ może przyjmować wartości spoza przedziału $[0, 1]$. Niemniej jest ona miarą ryzyka, na które wystawiony jest dany obiekt. Na przykład rosnąca funkcja hazardu w przypadku osób starszych odbija empiryczną prawidłowość, że są oni bardziej wystawieni na ryzyko śmierci niż osoby młode. Mamy

następujące zależności między funkcją hazardu, funkcją gęstości i funkcją przeżywalności

$$\lambda(t) = -\frac{\partial \ln S(t)}{\partial t}$$

$$f(t) = \lambda(t) S(t)$$

Definiujemy także zcałkowaną funkcję hazardu

$$\Lambda(t) = \int_0^t \lambda(s) ds$$

wtedy

$$S(t) = e^{-\Lambda(t)}$$

a więc

$$\Lambda(t) = -\ln S(t)$$

Funkcja wiarygodności dla modeli długości trwania może więc być zapisana jako

$$\ell(\boldsymbol{\theta}) = \ln \left[\prod_{i=1}^n f_{\theta}(t_i) \right] = \sum_{i=1}^n \ln [\lambda_{\theta}(t_i) S_{\theta}(t_i)] = \sum_{i=1}^n \ln \lambda_{\theta}(t) + \sum_{i=1}^n \ln S_{\theta}(t)$$

Jak już wcześniej wspomnieliśmy dane dotyczące długości trwania mogą być ocenzone. W tym przypadku dla t_i nieocenzurowanych wstawiamy w funkcji wiarygodności $f(t_i)$ a dla t_i ocenzonego $S(t_i)$. Funkcja wiarygodności przybierze w tym przypadku następującą postać

$$\begin{aligned} \ell(\boldsymbol{\theta}) &= \ln \left[\prod_{\text{nieocenzurowane}} f_{\theta}(t_i) \right] \ln \left[\prod_{\text{ocenzurowane}} S_{\theta}(t_i) \right] \\ &= \sum_{\text{nieocenzurowane}} \ln \lambda_{\theta}(t) + \sum_{\text{wszystkie}} \ln S_{\theta}(t) \end{aligned}$$

Najpopularniejsze modele dla liczebności to model wykładniczy, Weibulla, lognormalny i loglogistyczny. Różnią się one między sobą założonymi dystrybuantami $S(t)$. W zależności od przyjętej formy funkcyjnej i wielkości parametrów funkcja hazardu może monotonicznie rosnać, maleć lub zmieniać się w sposób niemonotoniczny. Dla monotonicznych funkcji hazardu najczęściej stosuje się warianty rozkładu wykładniczego

| Hazard | Funkcja hazardu $\lambda(t)$ | Funkcja przeżywalności $S(t)$ |
|-------------|---------------------------------|---|
| Stały | λ | $S(t) = e^{-\lambda t}$ |
| Liniowy | $\lambda(t) = \alpha + \beta t$ | $S(t) = e^{-(\alpha t + \frac{1}{2}\beta t^2)}$ |
| Wykładniczy | $\lambda(t) = e^{a+\beta t}$ | $S(t) = e^{-\frac{1}{\beta} \exp(a+\beta t)}$ |

Tablica 1: Warianty rozkładu wykładniczego

$$\int_{-\infty}^{\infty} \frac{\partial \left(\frac{1}{1+e^{-t}} \right)}{\partial t} dt$$

Dla niemonotonicznych funkcji hazardu dopasowujemy zwykle rozkłady inne niż wykładniczy. Najczęściej spotykane w tym kontekście formy funkcyjne dla funkcji hazardu i związanej z nią funkcji $S(t)$ znajdują się w tabeli (2).

Warto przy tym zwrócić uwagę, że dla $p = 1$ rozkład Weibulla sprowadza się do rozkładu wykładniczego. Wartość oczekiwana t w rozkładzie Weibulla wynosi $E(t) = \frac{1}{\lambda} \Gamma\left(\frac{1}{p} + 1\right)$. Z kolei w przypadku rozkładu Lognormalnego, $\ln(t)$ powinno mieć rozkład normalny $N(-p \ln \lambda, p^{-2})$ a dla rozkładu loglogistycznego $\ln(t)$ ma rozkład logistyczny o wartości oczekiwanej $E(\ln(t)) = -p \ln \lambda$.

| Rozkład | Funkcja hazardu $\lambda(t)$ | Funkcja przeżywalności $S(t)$ |
|----------------|--|--------------------------------------|
| Weibull | $\lambda(t) = \lambda p (\lambda t)^{p-1}$ | $S(t) = e^{-(\lambda t)^p}$ |
| Lognormalny | $f(t) = \left(\frac{p}{t}\right) \phi [p \ln(\lambda t)]$ | $S(t) = \Phi [-p \ln(\lambda t)]$ |
| Loglogistyczny | $\lambda(t) = \frac{\lambda p (\lambda t)^{p-1}}{1 + (\lambda t)^p}$ | $S(t) = \frac{1}{1 + (\lambda t)^p}$ |

Tablica 2: Funkcje hazardu i przeżywalności

Podobnie jak w przypadku modeli dla liczebności aby wprowadzić zmienne objaśniające do modelu zakładamy, że dla parametr λ może się zmieniać w czasie pod wpływem zmian zmiennych egzogenicznych

$$\ln \lambda_t = -\mathbf{x}_t \boldsymbol{\beta}$$

Interpretacja wyestymowanego wektora parametrów $\boldsymbol{\beta}$ wynika z faktu, że dla każdej z analizowanych funkcji

Literatura: ? str. 289-299, ? str. 948-984, ? str. 635-642.