Metoda Największej Wiarogodności

Jerzy Mycielski

WNE, UW, 2024

- Założenia

1. $\boldsymbol{y}_i$, $\boldsymbol{x}_i$ zaobserwowane dla $i = 1, \ldots, n$
2. poszczególne obserwacje niezależne
3. obserwacje $\boldsymbol{y}_i$ pochodzą z rozkładu warunkowego $f_\theta(\boldsymbol{y}_i | \boldsymbol{x}_i)$
4. θ prawdziwa wartość szacowanego wektora parametrów
5. postać funkcji $f_\theta(\cdot)$ nie zależy od n a jedynie od $\theta \in \Theta \subset R^k$

Funkcja wiarogodności

- Łączna funkcję gęstości wektora $\boldsymbol{y} = [y_1, \ldots, y_n]'$

$$f_\theta\left(\boldsymbol{y} | \boldsymbol{X}\right) = f_\theta\left(y_1, \ldots, y_n | \boldsymbol{x}_1, \ldots, \boldsymbol{x}_n\right)$$

  nazywamy funkcją wiarogodności i oznaczamy $L(\boldsymbol{y} | \boldsymbol{X}, \theta)$
- Z niezależności wynika, że funkcja wiarogodności jest równa:

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

  gdzie $\boldsymbol{y} = [\boldsymbol{y}_1, \ldots \boldsymbol{y}_n]'$, $\boldsymbol{X} = [\boldsymbol{x}'_1, \ldots, \boldsymbol{x}'_n]'$ i $\boldsymbol{y} \in \boldsymbol{Y}$

Logarytm funkcji wiarogodności

- Logarytm funkcji wiarogodności ma postać

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

- Będziemy pomijać $\boldsymbol{y}$ i $\boldsymbol{X}$ oraz y_i i $\boldsymbol{x}_i$ i pisać $\ell(\theta)$, $L(\theta)$ i $\ell_i(\theta)$

Definicja

Estymatorem *MNW* jest takie θ dla której funkcja wiarogodności ma największą wartość:

$$\widetilde{\theta} = \arg\max L(\theta) = \arg\max \ell(\theta)$$

Założenia

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

$$\varepsilon_i \sim N\left(0, \sigma^2\right)$$

i ε_i niezależne, w rezultacie

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

warunkowa funkcja gęstości

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

Funkcja wiarogodności

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

logarytm funkcji wiarogodności

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

- warunki pierwszego rzędu

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

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

- rozwiązanie

$$\widetilde{\beta} = \left(\mathbf{X}'\mathbf{X}\right)^{-1}\mathbf{X}'\mathbf{y}$$

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

- Jako estymator $\widetilde{\beta}$ uzyskaliśmy dokładnie estymator *MNK*!
- Estymator $\widetilde{\sigma}^2$ różni się od estymatora *MNK* ($s^2 = \frac{\mathbf{e}'\mathbf{e}}{n-k}$), więc jest obciążony ale różnica między $\widetilde{\sigma}^2$ i s^2 dąży do zera.

Własności estymatorów MNW (zgodność, normalność)

- Zgodność

$$\widetilde{\theta} \xrightarrow{p} \theta$$

- Macierz informacyjna Fishera

$$\boldsymbol{I}(\theta) = Var\left(\frac{\partial \ell(\theta)}{\partial \theta}\right) = -E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'}\right)$$

- Zgodnie z definicją macierz informacyjna równa jest wariancji gradientu - $Var\left(\frac{\partial \ell(\theta)}{\partial \theta}\right)$, lub minus wartość oczekiwanej Hessianu logarytmu funkcji wiarygodności.
- Rozkład asymptotyczny estymator *MNW* można przybliżyć za pomocą rozkładu normalnego:

$$\widetilde{\theta} - \theta \overset{a}{\sim} N\left(0, \boldsymbol{I}^{-1}(\theta)\right)$$

Własności estymatorów MNW (efektywność)

- Jeśli estymator $\widehat{\theta}$ jest zgodny, to asymptotycznie

$$Var\left(\sqrt{n}\left(\widehat{\theta}-\theta\right)\right) \geq n\boldsymbol{I}^{-1}(\theta)$$

- $\boldsymbol{I}^{-1}(\theta)$ równa najmniejszej możliwej macierzy wariancji estymatora asymptotycznie nieobciążonego
- Twierdzenie nazywane jest niekiedy twierdzeniem o dolnym ograniczeniu Cramera-Rao
- Łatwo się przekonać, że dla estymatorów nieobciążonych obowiązuje ono także w małych próbach (dolnym ograniczeniem jest wtedy macierz $\boldsymbol{I}^{-1}(\theta)$)
- Estymator *MNW* są asymptotycznie efektywne, ponieważ ich wariancja zbiega do dolnego ograniczenia Cramera-Rao

Estymator funkcji parametrów

- Niezmienniczość: Estymator *MNW* dla $\omega = \boldsymbol{g}(\theta)$, jest dany wzorem $\widetilde{\omega} = \boldsymbol{g}\left(\widetilde{\theta}\right)$. Jeśli $\widetilde{\theta}$ jest zgodne, to $\widetilde{\omega}$ jest też zgodne

$$\widetilde{\theta} \xrightarrow{p} \theta \Longrightarrow \widetilde{\omega} \xrightarrow{p} \omega$$

- Rozkład $\widetilde{\omega}$ można przybliżyć w następujący sposób (metoda delta)

$$\widetilde{\omega} - \omega \overset{a}{\sim} N\left(0, \frac{\partial \boldsymbol{g}\left(\widetilde{\theta}\right)}{\partial \theta'} \boldsymbol{I}^{-1}\left(\widetilde{\theta}\right) \frac{\partial \boldsymbol{g}\left(\widetilde{\theta}\right)}{\partial \theta}\right)$$

Estymator macierzy wariancji w MNW

- Z asymptotycznych własności estymatora *MNW* wynika, że estymatorem wariancji $\widetilde{\theta}$ jest odwrotność macierzy informacyjnej $\boldsymbol{I}^{-1}(\theta)$. Z kolei macierz informacyjną $\boldsymbol{I}(\theta)$ można oszacować za pomocą macierzy $\boldsymbol{I}\left(\widetilde{\theta}\right)$.

- gradient

$$
\begin{aligned}
\frac{\partial \ell(\beta, \sigma^2)}{\partial \beta} &= -\frac{1}{2\sigma^2}\left(2\boldsymbol{X}'\boldsymbol{X}\beta - 2\boldsymbol{X}'\boldsymbol{y}\right) \\
&= \frac{1}{\sigma^2}\boldsymbol{X}'\varepsilon \\
\frac{\partial \ell(\beta, \sigma^2)}{\partial \sigma^2} &= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}(\boldsymbol{y} - \boldsymbol{X}\beta)'(\boldsymbol{y} - \boldsymbol{X}\beta) \\
&= -\frac{n}{2}\frac{1}{\sigma^2} + \frac{1}{2\sigma^4}\varepsilon'\varepsilon
\end{aligned}
$$

- Wariancja pochodnej względem β

$$Var\left(\frac{\partial \ell\left(\beta,\sigma^2\right)}{\partial \beta}\right)$$
$$= \frac{1}{\sigma^4}\boldsymbol{X}'Var\left(\varepsilon\right)\boldsymbol{X} = \frac{1}{\sigma^2}\left(\boldsymbol{X}'\boldsymbol{X}\right)$$

- pochodnej względem σ^2

$$Var\left(-\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\varepsilon'\varepsilon\right)$$
$$= \frac{1}{4\sigma^8}\left(n2\sigma^4\right) = \frac{n}{2\sigma^4}$$

bo $Var\left(\varepsilon_i^2\right) = 2\sigma^4$.

- kowariancja między pochodnymi

$$E\left[\frac{1}{\sigma^2}\boldsymbol{X}'\varepsilon\left(-\frac{n}{2}\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\varepsilon'\varepsilon\right)\right]=0$$

bo $E\left(\varepsilon_j^3\right)=0,\ E\left(\varepsilon_i\varepsilon_j^2\right)=E\left(\varepsilon_i\right)E\left(\varepsilon_j^2\right)=0$

- Elementy Hessianu

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

$$\frac{\partial^2\ell\left(\beta,\sigma^2\right)}{\partial\sigma^2\partial\sigma^2}=\frac{n}{2}\frac{1}{\sigma^4}-\frac{1}{\sigma^6}\varepsilon'\varepsilon$$

$$\frac{\partial^2\ell\left(\beta,\sigma^2\right)}{\partial\beta\partial\sigma^2}=-\frac{1}{\sigma^4}\boldsymbol{X}'\varepsilon$$

- Macierz informacyjna Fishera

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

- Wariancja $\widetilde{\theta} = \left[\widetilde{\beta}', \widetilde{\sigma}^2\right]'$

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

- Estymator wariancji

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

Estymator wariancji w MNW

- Analityczne policzenie wariancji gradientu $Var\left(\frac{\partial \ell(\theta)}{\partial \theta}\right)$ bądź minus wartość oczekiwanej Hessianu $-E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'}\right)$ jest w większości przypadków bardzo skomplikowane bądź wręcz niemożliwe.
- Jeśli jednak spełniony jest warunek niezależności i tych samych rozkładów poszczególnych elementów próby, to jak wcześniej wspomniano:

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

Estymator wariancji w MNW c.d.

- Można pokazać, że

$$E\left(\frac{\partial \ell_i(\theta)}{\partial \theta}\right) = 0 \text{ dla każdego } i$$

a więc w związku z równością rozkładów

$$Var\left(\frac{\partial \ell_i(\theta)}{\partial \theta}\right) = -E\left(\frac{\partial \ell_i(\theta)}{\partial \theta}\frac{\partial \ell_i(\theta)}{\partial \theta'}\right) \text{ dla każdego } i$$

- Z prawa wielkich liczb

$$\frac{1}{n}\sum_{i=1}^{n}\frac{\partial \ell_i(\theta)}{\partial \theta}\frac{\partial \ell_i(\theta)}{\partial \theta'} \xrightarrow{p} E\left(\frac{\partial \ell_i(\theta)}{\partial \theta}\frac{\partial \ell_i(\theta)}{\partial \theta'}\right)$$

- Z niezależności

$$
\begin{aligned}
\boldsymbol{I}\left(\beta,\sigma^2\right) &= Var\left(\frac{\partial\ell(\theta)}{\partial\theta}\right) = Var\left(\sum_{i=1}^{n}\frac{\partial\ell_i(\theta)}{\partial\theta}\right) \\
&= \sum_{i=1}^{n} Var\left(\frac{\partial\ell_i(\boldsymbol{y}_i|\boldsymbol{x}_i,\theta)}{\partial\theta}\right) \\
&= nE\left(\frac{\partial\ell_i(\theta)}{\partial\theta}\frac{\partial\ell_i(\theta)}{\partial\theta'}\right) \overset{a}{\approx} \sum_{i=1}^{n}\frac{\partial\ell_i(\theta)}{\partial\theta}\frac{\partial\ell_i(\theta)}{\partial\theta'}
\end{aligned}
$$

- Estymator macierzy wariancji można więc wyznaczyć jako odwrotność sumy iloczynów gradientów logarytmów funkcji gęstości dla poszczególnych zdarzeń: ten estymator określany jest niekiedy mianem *OPG* (**O**uter **P**roduct of **G**radients)

Estymator wariancji w MNW c.d.

- Podobnie pokazujemy, że

$$\boldsymbol{I}\left(\beta,\sigma^2\right) = -E\left(\frac{\partial^2 \ell(\theta)}{\partial\theta\partial\theta'}\right) \overset{a}{\approx} -\sum_{i=1}^{n} \frac{\partial^2 \ell_i\left(\widetilde{\theta}\right)}{\partial\theta\partial\theta'}$$

- Estymator macierzy wariancji możemy więc też uzyskać jako odwrotność minusa sumy Hessianów logarytmów funkcji wiarygodności dla poszczególnych zdarzeń. Czasami estymator ten jest nazywany estymatorem empirycznego Hessianu.

Estymator wariancji w MNW c.d.

- Wady i zalety poszczególnych estymatorów:
  - Dla estymatorów opartych na wzorach analitycznych $Var\left(\frac{\partial\ell(\theta)}{\partial\theta}\right) = -E\left(\frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta'}\right)$. Wybieramy wersję, którą łatwiej wyprowadzić. Uzyskany estymator jest zawsze nieujemnie określony i dawaje najlepsze oszacowanie wariancji. Wyprowadzenie postaci tych estymatorów w przypadku funkcji nieliniowych jest jednak zazwyczaj niemożliwe.
  - *OPG* - wyprowadzenie tego estymatora jest względnie łatwa, wymaga on policzenia jedynie wektora pierwszych pochodnych, uzyskana macierz jako suma macierzy nieujemnie określonych jest zawsze nieujemnie określona.
  - Empiryczny Hessian - wyprowadzenie go może być nieco trudniejsze ze względu na konieczność policzenia macierzy drugich pochodnych. Estymator ten jest nieujmnie określony jeśli udało nam się znaleźć minimum funkcji wiarygodności.
- Uwaga: niektóre współczesne programy ekonometryczne są w stanie policzyć numerycznie pochodne $\frac{\partial\ell(\theta)}{\partial\theta}$ i $\frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta'}$.

KNMRL c.d.

- Logarytm funkcji gęstości dla poszczególnego zdarzenia

$$\begin{aligned}\ell_i\left(\beta,\sigma^2\right) &= -\frac{1}{2}\ln\left(2\pi\right)-\frac{1}{2}\ln\left(\sigma^2\right)\\ &\quad -\frac{1}{2\sigma^2}\left(y_i-\boldsymbol{x}_i\beta\right)'\left(y_i-\boldsymbol{x}_i\beta\right)\end{aligned}$$

- Pierwsza pochodna

$$\frac{\partial\ell_i\left(\beta,\sigma^2\right)}{\partial\theta}=\frac{1}{\sigma^2}\left(\boldsymbol{x}_i'y_i-\boldsymbol{x}_i'\boldsymbol{x}_i\beta\right)=\boldsymbol{x}_i'\left(y_i-\boldsymbol{x}_i\beta\right)$$

$$\frac{\partial\ell_i\left(\beta,\sigma^2\right)}{\partial\sigma^2}=-\frac{1}{2}\frac{1}{\sigma^2}+\frac{1}{2\sigma^4}\left(y_i-\boldsymbol{x}_i\beta\right)^2$$

- Z powyższych wzorów wynika, że pierwsza pochodna policzona dla wcześniej uzyskanych oszacowań $\widetilde{\beta}, \widehat{\sigma}^2$ ma postać:

$$\begin{aligned} \frac{\partial \ell_i\left(\widetilde{\beta}, \widetilde{\sigma}^2\right)}{\partial \beta} &= \frac{1}{\widetilde{\sigma}^2}\boldsymbol{x}_i e_i \\ \frac{\partial \ell_i\left(\widetilde{\beta}, \widetilde{\sigma}^2\right)}{\partial \sigma^2} &= \frac{1}{2\widetilde{\sigma}^2}\left(\frac{e_i^2}{\widetilde{\sigma}^2} - 1\right) \end{aligned}$$

gdzie reszta $e_i = y_i - \boldsymbol{x}_i\widetilde{\beta}$

- Estymator *OPG* macierzy informacyjnej

$$
\sum_{i=1}^{n} \frac{\partial \ell_i\left(\widetilde{\theta}\right)}{\partial \theta} \frac{\partial \ell_i\left(\widetilde{\theta}\right)}{\partial \theta'}
$$

$$
= \frac{1}{\widetilde{\sigma}^4} \begin{bmatrix} \sum_{i=1}^n \boldsymbol{x}_i' \boldsymbol{x}_i e_i^2 & \frac{1}{2} \sum_{i=1}^n \boldsymbol{x}_i e_i \left(\frac{e_i^2}{\widetilde{\sigma}^2} - 1\right) \\ \frac{1}{2} \sum_{i=1}^n \boldsymbol{x}_i e_i \left(\frac{e_i^2}{\widetilde{\sigma}^2} - 1\right) & \frac{1}{4} \sum_{i=1}^n \left(\frac{1}{\widetilde{\sigma}^2} e_i^2 - 1\right)^2 \end{bmatrix}
$$

$$
= \frac{1}{\widetilde{\sigma}^4} \begin{bmatrix} \sum_{i=1}^n \boldsymbol{x}_i' \boldsymbol{x}_i e_i^2 & \frac{1}{2} \sum_{i=1}^n \boldsymbol{x}_i e_i^3 \\ \frac{1}{2} \sum_{i=1}^n \boldsymbol{x}_i e_i^3 & \frac{1}{4} \left(\sum_{i=1}^n \frac{e_i^4}{\widetilde{\sigma}^4} - n\right) \end{bmatrix}
$$

przy czym skorzystaliśmy z tego, że $\sum_{i=1}^n \boldsymbol{x}_i e_i = 0$ (z własności hiperpłaszczyzny regresji) a $n\widetilde{\sigma}^2 = \sum_{i=1}^n e_i^2$. Estymator ten ma wyraźnie różną postać od estymatora macierzy $\boldsymbol{I}\left(\widetilde{\theta}\right)$, który policzyliśmy wcześniej licząc analitycznie wariancję gradientu. Odwrotność tej macierzy da nam estymator *OPG* wariancji.

KNMRL c.d.

- Estymator empirycznego Hessianu

$$\frac{\partial^2 \ell_i\left(\beta,\sigma^2\right)}{\partial\beta\partial\beta'} = -\frac{1}{\widehat{\sigma}^2}\boldsymbol{x}_i'\boldsymbol{x}_i$$

$$\frac{\partial^2 \ell_i\left(\beta,\sigma^2\right)}{\partial\sigma^2\partial\sigma^2} = \frac{1}{\widehat{\sigma}^4}\left(\frac{1}{2} - \frac{e_i^2}{\widetilde{\sigma}^2}\right)$$

$$\frac{\partial^2 \ell_i\left(\beta,\sigma^2\right)}{\partial\beta\partial\sigma^2} = -\frac{1}{\sigma^4}\boldsymbol{x}_i' e_i$$

$$\sum_{i=1}^{n}\frac{\partial^2 \ell_i\left(\widetilde{\theta}\right)}{\partial\theta\partial\theta'} = \frac{1}{\widetilde{\sigma}^2}\begin{bmatrix} \sum_{i=1}^n \boldsymbol{x}_i'\boldsymbol{x}_i & 0 \\ 0 & \frac{n}{2\widetilde{\sigma}^2}\end{bmatrix}$$

a więc estymator wariancji ma postać

$$\begin{bmatrix} \widetilde{\sigma}^2\left(\boldsymbol{X}'\boldsymbol{X}\right)^{-1} & 0 \\ 0 & \frac{n}{2}\widetilde{\sigma}^4 \end{bmatrix}$$

równą wcześniej policzonej jako wariancja gradientu i wartość oczekiwana minus Hesianu. Jest to jednak wyjątkowy wynik.

Testowanie hipotez w MNW

- Testowana ogólna hipoteza nieliniowa: H_0:$\boldsymbol{h}(\theta) = 0$, gdzie $\boldsymbol{h}(\theta)$ jest funkcją wektorową $\boldsymbol{h}(\theta) = \left\{ \begin{array}{c} h_1(\theta) \\ \vdots \\ h_g(\theta) \end{array} \right.$.
- Oznaczmy przez $\boldsymbol{H}(\theta)$ macierz pochodnych wektora $\boldsymbol{h}(\theta)$ względem θ (Jakobian)
- Do weryfikacji tak sformułowanej hipotezy najczęściej używa się jednego z trzech testów.

Test ilorazu wiarogodności

- Test ilorazu wiarogodności (**L**ikelihood **R**atio) (LR),

$$LR = 2\left(\ell\left(\widetilde{\theta}\right) - \ell\left(\widetilde{\theta}_R\right)\right) \xrightarrow{D} \chi^2_g$$

gdzie g jest liczbą nałożonych ograniczeń

- wymaga policzenia $\widetilde{\theta}$ i $\widetilde{\theta}_R$
- wymaga policzenia różnicy funkcji wiarogodności (naogół łatwe)

- Test **W**alda (W)

$$W = \boldsymbol{h}\left(\widetilde{\theta}\right)' \left[\boldsymbol{H}\left(\widetilde{\theta}\right) \boldsymbol{I}^{-1}\left(\widetilde{\theta}\right) \boldsymbol{H}'\left(\widetilde{\theta}\right)\right]^{-1} \boldsymbol{h}\left(\widetilde{\theta}\right) \xrightarrow{D} \chi^2_g$$

gdzie g jest liczbą nałożonych ograniczeń

- wymaga policzenia jedynie $\widetilde{\theta}$
- wymaga policzenia odpowiedniej formy kwadratowej
- nie jest niezmiennicza ze względu na sposób zdefiniowania $\boldsymbol{h}(\theta)$

Test mnożników Lagrange'a

- mnożników Lagrange'a (**L**agrange **M**ultipliers) (LM).

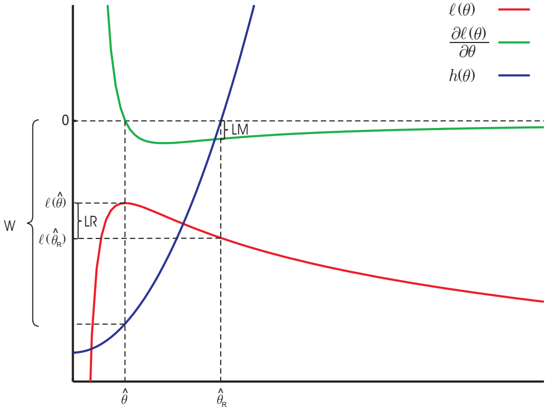
$$LM = \left.\frac{\partial \ell(\theta)}{\partial \theta'}\right|_{\theta=\widetilde{\theta}_R} \boldsymbol{I}^{-1}\left(\widetilde{\theta}_R\right) \left.\frac{\partial \ell(\theta)}{\partial \theta}\right|_{\theta=\widetilde{\theta}_R} \xrightarrow{D} \chi^2_g$$

  gdzie g jest liczbą nałożonych ograniczeń
- wymaga policzenia jedynie $\widetilde{\theta}_R$
- wymaga policzenia gradientu lub $\widetilde{\lambda}$ i odpowiedniej formy kwadratowej

Testowanie hipotez w MNW

- Intuicja: jeśli warunki poboczne są prawdziwe, to
  - LR wartość $\ell(\widetilde{\theta})$ nie powinna istotnie różnić się od $\ell(\widetilde{\theta}_R)$
  - W wartość $\widetilde{\theta}$ powinna być zbliżona do $\widetilde{\theta}_R$
  - LM wartość $\frac{\partial \ell_i(\widetilde{\theta}_R)}{\partial \theta}$ powinna być bliska 0 (wartość gradientu powinna być bliska zeru)
- Dla modeli liniowych

$$LM \leqslant LR \leqslant W$$



Rysunek: Testy LM, W i LR

- *Efekty cząstkowe (*partial effects*)* dla zmiennej ciągłej:

$$\frac{\Delta E(y|\boldsymbol{x})}{\Delta x_k} = \frac{\partial E(y|\boldsymbol{x})}{\partial x_k}$$

- Wielkość efektu cząstkowego zależy od wielkości $\boldsymbol{x}$
- Zazwyczaj liczymy go dla średnich wartości zmiennych egzogenicznych albo jako średnią wyliczoną dla obserwacji w próbie (APE - Average Partial Effect)

$$APE = \frac{1}{n}\sum_{i=1}^{n}\frac{\partial E(y_i|\boldsymbol{x}_i)}{\partial x_k}$$

- Dla zero-jedynkowej zmiennej objaśniającej x_k efekt krańcowy definiuje się jako:

$$\Delta E(y|\boldsymbol{x}) = E(y|\overline{\boldsymbol{x}}_{k1}) - E(y|\overline{\boldsymbol{x}}_{k0}),$$

gdzie $\boldsymbol{x}_{k1} = [x_1, \ldots, x_{k-1}, 1, x_{k+1}, \ldots, x_K]$, $\boldsymbol{x}_{k0} = [x_1, \ldots, x_{k-1}, 0, x_{k+1}, \ldots, x_K]$ lub jako średni efekt cząstkowy:

$$APE = \frac{1}{n}\sum_{i=1}^{n}[E(y_i|\boldsymbol{x}_{i,k1}) - E(y_i|\boldsymbol{x}_{i,k0})]$$

a więc jako różnice między wartościami oczekiwanymi y_i dla danej zerojedynkowej zmiennej objaśniającej równej zeru i równej 1.

- W modelach nieliniowych dobrze interpretowalne są zazwyczaj jedynie wielkości efektów cząstkowych
- Oszacowanych współczynników w różnych modelach nieliniowych o tej samej zmiennej zależnej i zmiennych niezależnych nie można zazwyczaj porównywać.
- Porównywać można efekty cząstkowe.

- W wielu zastosowaniach ekonomicznych forma liniowa modelu nie jest zadowalająca.
- W tych przypadkach możemy posłużyć się *NMNK*.
- Zależność y_i od $\boldsymbol{x}_i$ w *NMNK* dana jest równaniem:

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

  gdzie

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

  a funkcja $f\left(\boldsymbol{x}_i, \theta\right)$ jest nieliniowa względem wektora parametrów θ.
- Zauważmy, że błąd losowy jest addytywny.

- Zlogarytmowana funkcja wiarygodności dla tego modelu ma postać

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

- Parametr $\widetilde{\theta}$, maksymalizujący funkcję wiarygodności powinien minimalizować sumę kwadratów reszt

$$\min_{\widetilde{\theta}}\sum_{i=1}^{n}\left(y_i-f\left(\boldsymbol{x}_i,\widetilde{\theta}\right)\right)^2=\min_{\widetilde{\theta}}\sum_{i=1}^{n}e_i^2\left(\widetilde{\theta}\right)$$

gdzie $e_i^2\left(\widetilde{\theta}\right)=y_i-f\left(\boldsymbol{x}_i,\widetilde{\theta}\right)$

Przykład: Model CES

- Uogólnieniem modelu Cobba-Douglasa jest model *CES* (**C**onstant **E**lasticity of **S**ubstitution)

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

- Funkcja ta sprowadza się do funkcji Cobba-Douglasa dla $\rho \longrightarrow 0$ (ale jej szczególnymi przypadkami są także liniowa funkcja produkcji ($\rho = 1$) i Leontiefa($\rho \longrightarrow -\infty$)).

- Jeśli założymy, że funkcja produkcji ma postać funkcji *CES* to mamy otrzymamy model:

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

po zlogarytmowaniu stronami mamy

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

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.

- Z ogólnych własności prawdopodobieństwa

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

gdzie $\boldsymbol{Y}$ zbiórem możliwych wartości $\boldsymbol{y}$

- Zauważmy, że

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

- Jeśli spełnione są warunki regularności

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

ponieważ

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

- Wariancja pochodnej $\ell(\theta)$ jest równa wartości oczekiwanej Hessianowi $-\ell(\theta)$:

$$\frac{\partial^2}{\partial\theta\partial\theta'}\int_{\mathbf{Y}} L(\theta)\,d\mathbf{y} = \int_{\mathbf{Y}} \frac{\partial}{\partial\theta}\left(\frac{\partial\ell(\theta)}{\partial\theta}L(\theta)\right)d\mathbf{y}$$

$$= \int_{\mathbf{Y}} \frac{\partial^2\ell(\theta)}{\partial\theta\partial\theta'}L(\theta)\,d\mathbf{y} + \int_{\mathbf{Y}} \frac{\partial\ell(\theta)}{\partial\theta}\frac{\partial L(\theta)}{\partial\theta'}d\mathbf{y} = 0$$

Własności gradientu i Hessianu $\ell()$ c.d.

- Wynika z tego, że:

$$\int\limits_{\boldsymbol{Y}} \frac{\partial \ell(\theta)}{\partial \theta} \frac{\partial L(\theta)}{\partial \theta'} d\boldsymbol{y} = \int\limits_{\boldsymbol{Y}} \frac{\partial \ell(\theta)}{\partial \theta} \frac{\partial \ell(\theta)}{\partial \theta'} L(\theta)\, d\boldsymbol{y}$$

$$= Var\left(\frac{\partial \ell(\theta)}{\partial \theta}\right)$$

$$= -\int\limits_{\boldsymbol{Y}} \frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'} L(\theta)\, d\boldsymbol{y}$$

$$= -E\left(\frac{\partial^2 \ell(\theta)}{\partial \theta \partial \theta'}\right) = \boldsymbol{I}(\theta)$$

Optymalizacja numeryczna (metoda Newtona-Raphsona)

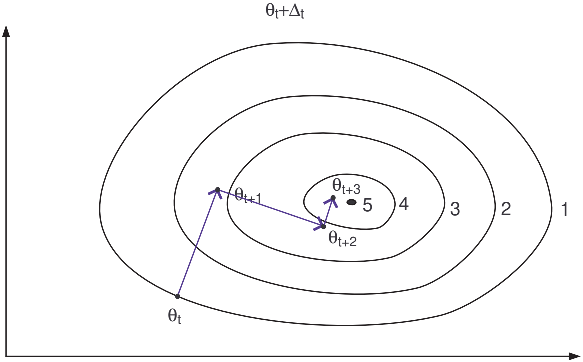
- Rozwinięcie warunków pierwszego rzędu dla θ_{t+1} wokół θ_t ma postać
$$\frac{\partial F(\theta_{t+1})}{\partial \theta_{t+1}} \approx \boldsymbol{g}_t + \boldsymbol{H}_t(\theta_{t+1} - \theta_t) = 0$$
gdzie $\boldsymbol{H}_t$ macierz drugich pochodnych $F(\theta)$ policzona dla θ_t
- Rozwiązując dla θ_{t+1}
$$\theta_{t+1} \approx \theta_t - \boldsymbol{H}_t^{-1}\boldsymbol{g}_t$$

Metoda Newtona-Raphsona c.d.

- Niech $\boldsymbol{\Delta}_t = \theta_{t+1} - \theta_t = -\boldsymbol{H}_t^{-1}\boldsymbol{g}_t$
- Z rozwinięcia Taylora:

$$
\begin{aligned}
F(\theta_{t+1}) &= F(\theta_t + \boldsymbol{\Delta}_t) \approx F(\theta_t) + \boldsymbol{\Delta}'_t \boldsymbol{g}_t \\
&= F(\theta_t) - \boldsymbol{g}'_t \boldsymbol{H}_t^{-1} \boldsymbol{g}_t
\end{aligned}
$$

- Dla ujemnie określonego $\boldsymbol{H}_t$ (a taki powinien być w otoczeniu maksimum) funkcja celu powinna rosnąć w każdym kroku!



Rysunek: Kolejne kroki optymalizacji