

# Metoda Największej Wiarogodności

- Założenia

1.  $y_i, x_i$  zaobserwowane dla  $i = 1, \dots, n$
2. poszczególne obserwacje niezależne
3. obserwacje  $y_i$  pochodzą z rozkładu warunkowego  
 $f_{\theta}(y_i | x_i)$

4.  $\theta$  prawdziwa wartość szacowanego wektora parametrów
5. postać funkcji  $f_{\theta}(\cdot)$  nie zależy od  $n$  a jedynie od  $\theta \in \Theta \subset R^k$
- Łączna funkcję gęstości wektora  $\mathbf{y} = [y_1, \dots, y_n]'$

$$f_{\theta}(\mathbf{y} | \mathbf{X}) = f_{\theta}(y_1, \dots, y_n | \mathbf{x}_1, \dots, \mathbf{x}_n)$$

nazywamy funkcją wiarygodności i oznaczamy  $L(\mathbf{y} | \mathbf{X}, \theta)$

- Z niezależności wynika, że funkcja wiarygodności jest równa:

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

- 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 postać

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

- Będziemy pomijać  $\mathbf{y}$  i  $\mathbf{X}$  oraz  $y_i$  i  $x_i$  i pisać  $\ell(\boldsymbol{\theta})$ ,  $L(\boldsymbol{\theta})$  i  $\ell_i(\boldsymbol{\theta})$
- Estymatorem *MNW* jest takie  $\boldsymbol{\theta}$  dla której funkcja

wiarogodności ma największą wartość

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

## Przykład Przykład: *KMRL*

- Założenia

$$y_i = x_i \beta + \varepsilon_i$$

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

i  $\varepsilon_i$  niezależne

- w rezultacie

$$y_i \sim N(x_i \beta, \sigma^2)$$

- warunkowa funkcja gęstości

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

- funkcja wiarygodności

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



- logarytm funkcji wiarygodności

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

warunki pierwszego rzędu

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

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

- rozwiązanie

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

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

- Jako estymator  $\tilde{\beta}$  uzyskaliśmy dokładnie estymator *MNK*!
- Estymator  $\tilde{\sigma}^2$  różni się od estymatora *MNK* ( $s^2 = \frac{e'e}{n-k}$ ) a więc jest obciążony jednak w dużych próbach

jednak użyta w estymatorze  $s^2$  poprawka małopróbkowa w dużych próbach znika ( $\lim_{n \rightarrow \infty} \frac{n-k}{n} = 1$ ) a więc w dużych próbach estymatory te są równoważne.

# Własności estymatorów $MNW$

- Zgodność

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

- Macierz informacyjna Fishera

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

- Zgodnie z definicją macierz informacyjna równa jest

wariancji gradientu -  $\text{Var} \left( \frac{\partial \ell(\boldsymbol{\theta})}{\partial \boldsymbol{\theta}} \right)$ , lub minus wartość oczekiwanej Hessianu logarytmu funkcji wiarygodności.

- Rozkład asymptotyczny estymator  $MNW$  można przybliżyć za pomocą rozkładu normalnego:

$$\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta} \stackrel{a}{\sim} N(\mathbf{0}, \mathbf{I}^{-1}(\boldsymbol{\theta}))$$

# Estymator funkcji parametrów w *MNW*

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

$$\tilde{\theta} \xrightarrow{p} \theta \implies \tilde{\omega} \xrightarrow{p} \omega$$

- Rozkład  $\tilde{\omega}$  można przybliżyć w następujący sposób

$$\tilde{\omega} - \omega \stackrel{a}{\sim} N \left( \mathbf{0}, \frac{\partial g(\tilde{\theta})}{\partial \theta'} \mathbf{I}_n^{-1}(\tilde{\theta}) \frac{\partial g(\tilde{\theta})}{\partial \theta} \right)$$

## Test $W$ , $LM$ i $LR$

- Testowana ogólna hipoteza nieliniowa:  $H_0: h(\theta) = 0$ , gdzie  $h(\theta)$  jest funkcją wektorową  $h(\theta) = \begin{pmatrix} h_1(\theta) \\ \vdots \\ h_g(\theta) \end{pmatrix}$ .
- Trzy testy



- ilorazu wiarygodności (**Likelihood Ratio**) ( $LR$ ),

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

- **Walda** ( $W$ )

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

- mnożników Lagrange'a (**Lagrange Multipliers**)

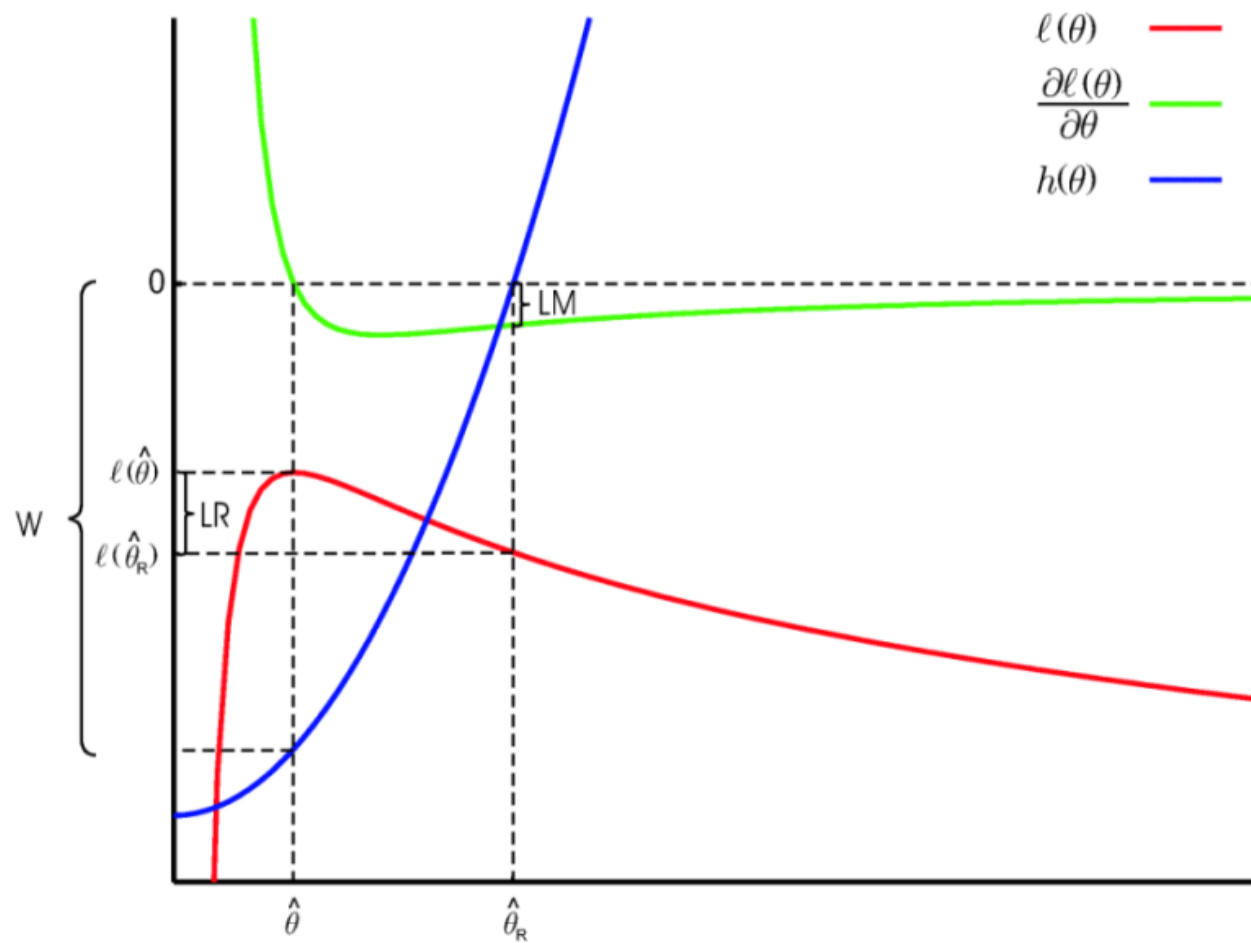
$(LM)$ .

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

gdzie  $g$  jest liczbą nałożonych ograniczeń

- Intuicja: jeśli warunki poboczne są prawdziwe, to
  - $LR$  wartość  $\ell(\tilde{\boldsymbol{\theta}})$  nie powinna istotnie różnić się od  $\ell(\tilde{\boldsymbol{\theta}}_R)$
  - $W$  wartość  $\tilde{\boldsymbol{\theta}}$  powinna być zbliżona do  $\tilde{\boldsymbol{\theta}}_R$

- $LM$  wartość  $\frac{\partial \ell_i(\tilde{\theta}_R)}{\partial \theta}$  powinna być bliska 0 (wartość gradientu powinna być bliska zeru)



# Zalety i wady poszczególnych testów

- Statystyka  $W$ 
  - wymaga policzenia jedynie  $\tilde{\theta}$
  - wymaga policzenia odpowiedniej formy kwadrato-  
wej
  - nie jest niezmiennicza ze względu na sposób zde-  
finiowania  $h(\theta)$
- Statystyka  $LM$

- wymaga policzenia jedynie  $\tilde{\theta}_R$
- wymaga policzenia gradientu lub  $\tilde{\lambda}$  i odpowiedniej formy kwadratowej
- Statystyka  $LR$ 
  - wymaga policzenia  $\tilde{\theta}$  i  $\tilde{\theta}_R$
  - wymaga policzenia różnicy funkcji wiarygodności (naogół łatwe)
- Dla modeli liniowych

$$LM \leq LR \leq W$$

# Efektywność estymatorów *MNW*

**Twierdzenie** Jeśli estymator  $\hat{\theta}$  jest zgodny, to

$$\text{Var} \left( \hat{\theta} - \theta \right) \geq \mathbf{I}_n^{-1} (\theta)$$

- $\mathbf{I}_n^{-1} (\theta)$  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  $I^{-1}(\theta)$ )

**Wniosek:** Estymator  $MNW$  są asymptotycznie efektywne, ponieważ ich wariancja zbiega do dolnego ograniczenia Cramera-Rao



# Modele nieliniowe - interpretacja współczynników

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

$$\frac{\Delta E(y_i | \mathbf{x}_i)}{\Delta x_k} = \frac{\partial E(y_i | \mathbf{x}_i)}{\partial x_k}$$

- Wielkość efektu częściowego zależy od punktu w jakim jest liczony

- Zazwyczaj liczymy go dla średnich wartości zmiennych egzogenicznych.
- Dla zero-jedynkowych zmiennych objaśniających efekt krańcowy definiuje się niekiedy jako  $\Delta E(y_i | x_i) = E(y_i | \bar{x}_{k1}) - E(y_i | \bar{x}_{k0})$  - różnice między wartościami oczekiwanymi  $y_i$  dla danej zmiennej zerojedynekowej równej zero 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. Można porównywać efekty cząstkowe.