

BAYESOWSKA ANALIZA MODELU NORMALNEJ REGRESJI LINIOWEJ
/BAYESOWSKI MODEL NORMALNEJ REGRESJI LINIOWEJ, BMNRL/

Łukasz Kwiatkowski

Katedra Ekonometrii i Badań Operacyjnych

Plan wykładu

- 1) Konstrukcja BMNRL
 - a. Rozkład próbkowy (funkcja wiarygodności) – wynika z KMNRL
 - b. Rozkład(y) *a priori*
- 2) Charakterystyka rozkładu(-ów) *a posteriori*
- 3) Testowanie pojedynczego parametru strukturalnego (w tym jego istotności)

Konstrukcja BMNRL – 1

➤ Zgrubne przypomnienie

- Równanie modelu regresji liniowej:

- Zapis skalarny – dla każdego $t = 1, 2, \dots, T$ (= liczba modelowanych obserwacji):

$$y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t = \underbrace{[x_{t1} \ x_{t2} \ \dots \ x_{tk}]_{(1 \times k)}}_{\tilde{x}_t} \underbrace{\begin{bmatrix} \beta_1 \\ \beta_2 \\ \vdots \\ \beta_k \end{bmatrix}}_{\beta_{(k \times 1)}} + \varepsilon_t = x_t \beta + \varepsilon_t$$

- Zapis macierzowy: $y = X\beta + \varepsilon$

$$y = \begin{bmatrix} y_1 \\ y_2 \\ \vdots \\ y_T \end{bmatrix}_{(T \times 1)} \quad X = \begin{bmatrix} x_{11} & x_{12} & \dots & x_{1k} \\ x_{21} & x_{22} & \dots & x_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ x_{T1} & x_{T2} & \dots & x_{Tk} \end{bmatrix}_{(T \times k)} \quad \varepsilon = \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \vdots \\ \varepsilon_T \end{bmatrix}_{(T \times 1)}$$

- Składniki losowe:

- Zapis skalarny ($t = 1, 2, \dots, T$): $E(\varepsilon_t) = 0$, $Var(\varepsilon_t) = \sigma^2$, $\forall_{t,s=1,\dots,T}^{t \neq s} Cov(\varepsilon_t, \varepsilon_s) = 0$

- Zapis macierzowy: $V(\varepsilon) = \underbrace{V(\varepsilon|\sigma^2)}_{\substack{\text{a dokładniej} \\ \text{rzecz ujmując}}} = \sigma^2 I_T = \begin{bmatrix} \sigma^2 & 0 & \dots & 0 \\ 0 & \sigma^2 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \sigma^2 \end{bmatrix} = \tau^{-1} I_T$,

gdzie $\tau = (\sigma^2)^{-1} = Prec(\varepsilon_t) = Prec(\varepsilon_t|\sigma^2)$, $\tau > 0$ (Prec = 1/Var, ale tylko w r. normalnym!)

Konstrukcja BMNRL – 2

➤ Wektor parametrów modelu:

$$\theta = \begin{bmatrix} \beta \\ \tau \end{bmatrix} \in \Theta = \mathbb{R}^k \times \mathbb{R}_+$$

➤ Model bayesowski:

$$p(y, \theta) = p(y|\theta)p(\theta)$$

- Rozkład próbkowy (funkcja wiarygodności): $p(y|\theta) = L_y(\theta)$??? (wynika z założeń KMNRL)

- Rozkład *a priori*: $p(\theta) = p(\beta, \tau) = \begin{cases} p(\beta)p(\tau) & \beta \text{ i } \tau - a \text{ priori niezależne} \\ p(\beta|\tau)p(\tau) & \beta \text{ i } \tau - a \text{ priori zależne} \\ \cancel{p(\tau|\beta)p(\beta)} & \beta \text{ i } \tau - a \text{ priori zależne} \end{cases}$

Konstrukcja BMNRL – rozkład próbkowy

➤ Założenia KMNRL – w zapisie macierzowym

1) $y = X\beta + \varepsilon$

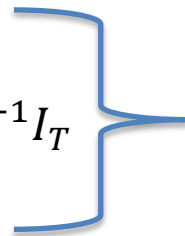
2) X – znana macierz nielosowa

3) $rz(X) = k \Leftrightarrow \det(X'X) \neq 0 \quad (\Rightarrow T \geq k)$

4) $E(\varepsilon) = \mathbf{0}_{(T \times 1)}$

5) $V(\varepsilon) = V(\varepsilon|\sigma^2) = V(\varepsilon|\theta) = \sigma^2 I_T = \tau^{-1} I_T$

6) $\varepsilon \sim N^{(T)}$



Założenia 4-6 $\Leftrightarrow \varepsilon \sim N^{(T)}(\mathbf{0}_{(T \times 1)}, \tau^{-1} I_T)$

➤ Pytanie o $p(y|\theta) = L_y(\theta)$:

• Skoro $\varepsilon \sim N^{(T)}(\mathbf{0}_{(T \times 1)}, \tau^{-1} I_T)$ i $y = X\beta + \varepsilon$, to $y|\theta \sim N^{(T)}(\mu, \Sigma)$,

gdzie $\mu \equiv E(y|\theta) = X\beta$, $\Sigma \equiv V(y|\theta) = V(\varepsilon|\sigma^2) = \tau^{-1} I_T$

• Zatem

$$\begin{aligned} p(y|\theta) &= f_N^T(y|X\beta, \tau^{-1} I_T) = (2\pi)^{-\frac{T}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp \left\{ -\frac{1}{2} (y - \mu)' \Sigma^{-1} (y - \mu) \right\} \\ &= (2\pi)^{-\frac{T}{2}} \tau^{\frac{T}{2}} \exp \left\{ -\frac{\tau}{2} (y - X\beta)' (y - X\beta) \right\} \end{aligned}$$

Konstrukcja BMNRL – rozkład(\mathbf{y}) *a priori*

$$\blacktriangleright p(\theta) = p(\beta, \tau) = \begin{cases} p(\beta)p(\tau) & \beta \text{ i } \tau - a \text{ priori niezależne} \\ p(\beta|\tau)p(\tau) & \beta \text{ i } \tau - a \text{ priori zależne} \\ p(\tau|\beta)p(\beta) & \beta \text{ i } \tau - a \text{ priori zależne} \end{cases}$$

Konstrukcja BMNRL – rozkład(y) *a priori* 1

➤ Cztery typowe specyfikacje $p(\theta)$:

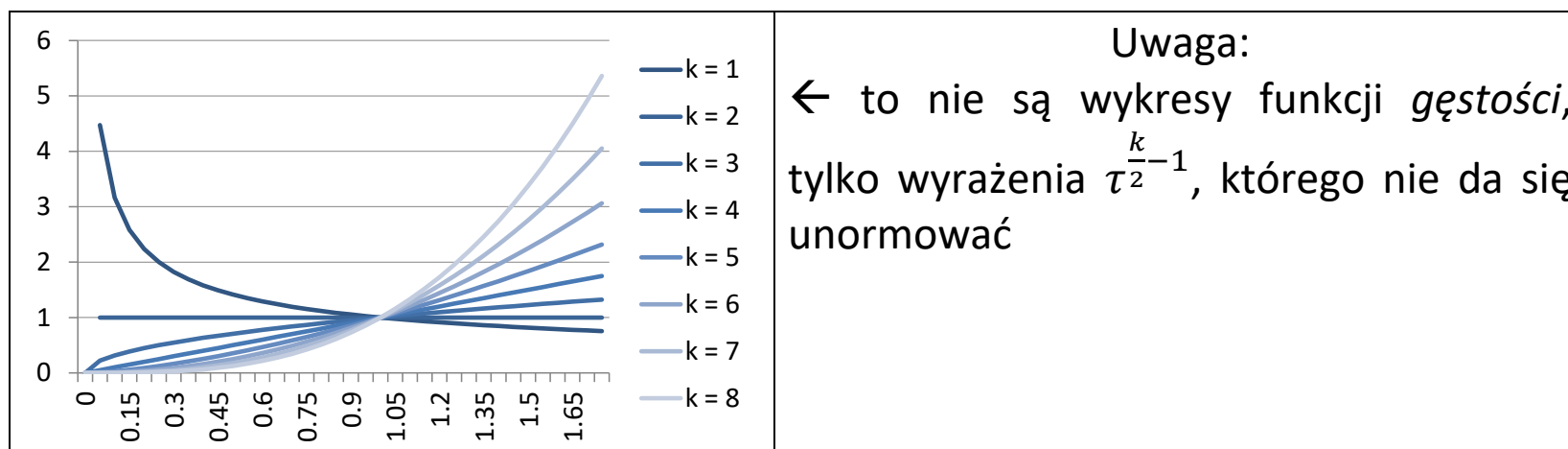
- [RJz] Reguła Jeffreysa (rozkład nieinformacyjny) – z zależnością pomiędzy β i τ
- [RJn] Reguła Jeffreysa (rozkład nieinformacyjny) – z niezależnością pomiędzy β i τ
- [GNz] Rozkład gamma-normalny – z zależnością pomiędzy β i τ
- [GNn] Rozkład gamma-normalny – z niezależnością pomiędzy β i τ

Konstrukcja BMNRL – rozkład(y) *a priori* 2

I. [RJz] Reguła Jeffreysa – z zależnością pomiędzy β i τ

$$p(\beta, \tau) \neq p(\beta)p(\tau) \quad \text{i} \quad p(\theta) \propto \sqrt{\det I(\theta)} = \dots \quad \Rightarrow \quad p_J(\theta) = p_J(\beta, \tau) \propto \tau^{\frac{k}{2}-1}$$

→ W jaki sposób zależy od β ?



→ Wady i zalety:

+ rozkład nieinformacyjny

+ rozkład *a posteriori* należy do znanej rodziny rozkładów p-stwa \Rightarrow wzory analityczne dla r. *a posteriori*

– rozkład niewłaściwy, ale r. *a posteriori* jest właściwy (niemniej, niewłaściwy r. *a priori* utrudnia kwestię porównywania konkurencyjnych modeli – a tu będzie to ważne...)

Konstrukcja BMNRL – rozkład(y) *a priori* 3

II. [RJn] Reguła Jeffreysa – z niezależnością pomiędzy β i τ

$$p(\beta, \tau) = p(\beta)p(\tau) \quad \text{i}$$

$$p(\beta) \propto \sqrt{\det[I(\theta)]_{\beta\beta}} = \dots \propto c > 0 \quad (\text{flat prior}) \quad \text{i} \quad p(\tau) \propto \sqrt{\det[I(\theta)]_{\tau\tau}} = \dots \propto \tau^{-1}$$

$$\Rightarrow p_J(\theta) = p_J(\beta|\tau)p_J(\tau) \propto c \cdot \tau^{-1}$$

→ Wady i zalety (jak dla RJz):

+ rozkład nieinformacyjny

+ rozkład *a posteriori* należy do znanej rodziny rozkładów p-stwa \Rightarrow wzory analityczne dla r. a posteriori

– rozkład niewłaściwy, ale r. *a posteriori* jest właściwy (niemniej, niewłaściwy r. *a priori* utrudnia kwestię porównywania konkurencyjnych modeli – a tu będzie to ważne...)

→ Ciekawostka: przy tym rozkładzie *a priori*, wnioskowanie bayesowskie o parametrach strukturalnych „pokrywa się” z podejściem niebayesowskim = bayesowska reinterpretacja wyników „klasycznej” estymacji

- $E(\beta|y) = \hat{\beta} = (X'X)^{-1}X'y \quad \leftarrow$ Metoda Najmniejszych Kwadratów, MNK
- Przedziały *HPD* dla parametrów strukturalnych pokrywają się z „klasycznymi” przedziałami ufności (lecz mają inną – zdecydowanie bardziej naturalną – interpretację)

Konstrukcja BMNRL – rozkład(y) *a priori* 4

III. [GNz] Rozkład **gamma-normalny** – z zależnością pomiędzy β i τ (**hiperparametry**)

$$p(\beta, \tau) = p(\beta|\tau)p(\tau) \quad \text{i} \quad p(\tau) = f_G\left(\tau \mid \frac{n_0}{2}, \frac{s_0}{2}\right) \quad \text{i} \quad p(\beta|\tau) = f_N^k(\beta|a, \tau^{-1}C^{-1})$$

→ Taki rozkład jest bardzo „wygodny”/pragmatyczny, albowiem:

$$\Rightarrow p(\beta) = \int_{\mathbb{R}_+} p(\beta|\tau)p(\tau)d\tau = \dots = f_{St}^k(\beta|a, P, n_0) \leftarrow \mathbf{k\text{-wymiarowy rozkład } t\text{-Studenta}}$$

(a rozkład *t*-Studenta jest dobrze „znany i lubiany”, więc łatwo z nim pracować)

n_0 – liczba stopni swobody; $a = Mo(\beta)$ – wektor niecentralności (także $E(\beta)$, gdy $n_0 > 1$)

$$P = Prec(\beta) = [p_{ij}]_{i,j=1,2,\dots,k} = \frac{n_0}{s_0} C \quad \rightarrow \quad P = Prec(\beta) \text{ – macierz precyzji}$$

$$\Rightarrow V(\beta) = \frac{n_0}{n_0-2} P^{-1} = \frac{s_0}{n_0-2} C^{-1}, \text{ gdy } n_0 > 2 \quad \text{Uwaga: } V(\beta) \neq V(\beta|\tau) = \tau^{-1}C^{-1}$$

⇒ Jeśli zadajemy $E(\tau)$, $Var(\tau)$ (alternatywnie $D(\tau)$) oraz $V(\beta)$, to hiperparametry:

$$n_0 = 2 \frac{E^2(\tau)}{Var(\tau)}, \quad s_0 = 2 \frac{E(\tau)}{Var(\tau)}, \quad C = \frac{s_0}{n_0-2} (V(\beta))^{-1}$$

$$\Rightarrow p(\beta_i) = \int_{\mathbb{R}^{k-1}} p(\beta) d\beta_{\setminus i} = f_{St}^1(\beta_i|a_i, P_i, n_0) \leftarrow \mathbf{1\text{-wymiarowy rozkład } t\text{-Studenta}}$$

$$\Rightarrow Var(\beta_i) = [V(\beta)]_{ii} \Rightarrow P_i = Prec(\beta_i) = \frac{n_0}{n_0-2} (Var(\beta_i))^{-1} = \frac{n_0}{s_0} ([C^{-1}]_{ii})^{-1}$$

Uwaga: $P_i \neq [P]_{ii} = p_{ii}$

Konstrukcja BMNRL – rozkład(y) *a priori* 5

III. [GNz] Rozkład gamma-normalny – z zależnością pomiędzy β i τ – C.D.

→ Wady i zalety:

+ rozkład właściwy

+ rozkład sprzężony \Rightarrow rozkład *a posteriori* jest tego samego typu (tj. gamma-normalny), więc należy do znanej rodziny rozkładów p-stwa \Rightarrow wzory analityczne dla r. *a posteriori*

+ może być wykorzystany do specyfikacji „w przybliżeniu” nieinformacyjnego rozkładu *a priori* – *a'la* [RJz]:

$$p(\beta|\tau) = f_N^k(\beta|\mathbf{a}, \tau^{-1}\mathbf{C}^{-1}) \propto c \text{ przy } \text{diag}(\mathbf{C}^{-1}) \rightarrow +\infty \quad (\text{vague prior, diffuse prior})$$

$$p(\tau) = f_G\left(\tau \mid \frac{n_0}{2}, \frac{s_0}{2}\right) \propto \tau^{\frac{k}{2}-1} \text{ przy } n_0 = k \text{ i } s_0 \rightarrow 0$$

– rozkład informacyjny (ALE i tak wraz z $T \rightarrow \infty$ rola rozkładu *a priori* maleje)

– określenie wstępnej wiedzy o zależności pomiędzy β i τ (poprzez \mathbf{C}) może być nieco problematyczne, ale zwykle przyjmuje się, że \mathbf{C} jest pewną macierzą diagonalną (zera poza przekątną)

Konstrukcja BMNRL – rozkład(y) *a priori* 6

IV. [GNn] Rozkład **gamma-normalny** – z **niezależnością** pomiędzy β i τ

$$p(\beta, \tau) = p(\beta)p(\tau) \quad \text{i} \quad p(\tau) = f_G\left(\tau \mid \frac{n_0}{2}, \frac{s_0}{2}\right) \quad \text{i} \quad p(\beta) = f_N^k(\beta \mid a, C^{-1})$$

$$\Rightarrow V(\beta) = C^{-1}$$

\Rightarrow Jeśli zadajemy $E(\tau)$, $Var(\tau)$ (alternatywnie $D(\tau)$), to hiperparametry (jak w GNz):

$$n_0 = 2 \frac{E^2(\tau)}{Var(\tau)}, \quad s_0 = 2 \frac{E(\tau)}{Var(\tau)}$$

\rightarrow Wady i zalety:

+ rozkład właściwy

+ nie ma potrzeby określania wstępnej wiedzy o zależności pomiędzy β i τ

+ może być wykorzystany do specyfikacji „w przybliżeniu” nieinformacyjnego rozkładu *a priori* – $a'la$ [RJn]:

$$p(\beta) = f_N^k(\beta \mid a, C^{-1}) \propto c \text{ przy } \text{diag}(C^{-1}) \rightarrow +\infty \quad (\text{vague prior, diffuse prior})$$

$$p(\tau) = f_G\left(\tau \mid \frac{n_0}{2}, \frac{s_0}{2}\right) \propto \tau^{-1} \text{ przy } n_0 \rightarrow 0 \text{ i } s_0 \rightarrow 0$$

– rozkład informacyjny, ALE wraz z $T \rightarrow \infty$ rola rozkładu *a priori* maleje

– nie jest rozkładem sprzężonym i, co więcej, rozkład *a posteriori* NIE należy do żadnej znanej rodziny rozkładów p-stwa \Rightarrow BRAK wzorów analitycznych na charakterystyki r. *a posteriori* \Rightarrow konieczność sięgania po zaawansowane metody symulacyjne w celu uzyskania próby (pseudo-)losowej z r. *a posteriori* – tzw. metody Monte Carlo oparte na łańcuchach Markowa; ang. *Markov Chain Monte Carlo*, **MCMC**

Charakterystyka rozkładu *a posteriori* – rozkład łączny

➤ Poniżej omówimy **tylko przypadki II (RJn) i III (GNz)**, gdyż:

- I (RJz) – rzadko wykorzystywany w praktyce
- IV (GNn) – brak formuł analitycznych → konieczność stosowania zaawansowanych metod numerycznych (MCMC)

➤ Rozkład *a posteriori* – zarówno w II (RJn), jak i III (GNz) – jest rozkładem **gamma-normalnym**:

$$p(\beta, \tau | y) = p(\beta | \tau, y) p(\tau | y) = f_N^k(\beta | \bar{a}, \tau^{-1} \bar{C}^{-1}) f_G\left(\tau | \frac{\bar{n}}{2}, \frac{\bar{s}}{2}\right)$$

| II (RJn) | III (GNz) |
|--|---|
| $\bar{a} = \hat{\beta} = (X'X)^{-1}X'y$ | $\bar{a} = \bar{C}^{-1}(Ca + X'X\hat{\beta})$ |
| $\bar{C} = X'X$ | $\bar{C} = C + X'X$ |
| $\bar{n} = T - k$ | $\bar{n} = n_0 + T$ |
| $\bar{s} = S(\hat{\beta}) = (y - X\hat{\beta})'(y - X\hat{\beta})$ | $\bar{s} = s_0 + y'y + a'Ca - \bar{a}'\bar{C}\bar{a}$ |

→ Najbardziej nurtujące nas PYTANIE: $p(\tau | y), p(\beta | y), p(\beta_i | y) = ???$

Charakterystyka rozkładu *a posteriori* – rozkłady brzegowe 1

➤ Brzegowy rozkład *a posteriori* dla τ – już mamy:

$$p(\tau|y) = f_G\left(\tau \mid \frac{\bar{n}}{2}, \frac{\bar{s}}{2}\right)$$

- $E(\tau|y) = \frac{\bar{n}}{\bar{s}}$
- $D(\tau|y) = \frac{\sqrt{2\bar{n}}}{\bar{s}}$
- Pozostałe charakterystyki – numerycznie (w pakietach lub symulacyjnie)

➤ Brzegowy rozkład *a posteriori* dla β (czyli łączny dla wszystkich parametrów strukturalnych)

$$p(\beta|y) = \int_{\mathbb{R}_+} p(\beta|\tau, y)p(\tau|y)d\tau = \dots = f_{St}^k(\beta|\bar{a}, \bar{P}, \bar{n}),$$

gdzie $\bar{P} = \frac{\bar{n}}{\bar{s}}\bar{C}$

- $E(\beta|y) = \bar{a}$, dla $\bar{n} > 1$
- $Mo(\beta|y) = \bar{a}$ (\bar{n} – dowolne)
- mediana w przypadku rozkładów wielowymiarowych – za trudne...)
- $V(\beta|y) = \frac{\bar{n}}{\bar{n}-2}\bar{P}^{-1} = \frac{\bar{s}}{\bar{n}-2}\bar{C}^{-1}$, dla $\bar{n} > 2$ → $Var(\beta_i|y) = [V(\beta|y)]_{ii}$

→ Zauważmy, że przy $\bar{n} \rightarrow \infty$: $p(\beta|y) = f_{St}^k(\beta|\bar{a}, \bar{P}, \bar{n} \rightarrow \infty) = f_N^k(\beta|\bar{a}, V(\beta|y))$

Charakterystyka rozkładu *a posteriori* – rozkłady brzegowe 2

➤ Brzegowe rozkłady *a posteriori* dla pojedynczych β_i ($i = 1, 2, \dots, k$):

$$p(\beta_i|y) = \int_{\mathbb{R}^{k-1}} p(\beta|y) d\beta_{\setminus i} = f_{St}^1(\beta_i|\bar{a}_i, \bar{P}_i, \bar{n})$$

gdzie \bar{a}_i jest i -tą współrzędną wektora \bar{a} , natomiast

$$\bar{P}_i = \text{Prec}(\beta_i|y) = \frac{\bar{n}}{\bar{n} - 2} (\text{Var}(\beta_i|y))^{-1} = \frac{\bar{n}}{\bar{s}} ([\bar{C}^{-1}]_{ii})^{-1}$$

Uwaga: $\bar{P}_i \neq [\bar{P}]_{ii}$

- $E(\beta_i|y) = \bar{a}_i$ dla $\bar{n} > 1$
- $Mo(\beta_i|y) = Me(\beta_i|y) = \bar{a}_i$ (\bar{n} – dowolne)
- $Var(\beta_i|y) = [V(\beta|y)]_{ii}$ dla $\bar{n} > 2$ → $D(\beta_i|y) = \sqrt{Var(\beta_i|y)}$
- $HPD_{1-\alpha}(\beta_i) = \left(\bar{a}_i - t_{\frac{\alpha}{2}} \sqrt{\bar{P}_i^{-1}}, \bar{a}_i + t_{\frac{\alpha}{2}} \sqrt{\bar{P}_i^{-1}} \right)$, $t_{\frac{\alpha}{2}}$ – dwustronna wartość kryt. w $St(0, 1, \bar{n})$
 - $\Pr(|x| \geq t_{\frac{\alpha}{2}}) = 1 - \alpha$, gdzie $x \sim St(0, 1, \bar{n})$;
 - $t_{\frac{\alpha}{2}}$ w Excelu: → =ROZKŁAD.T.ODW(prawdopodobieństwo = α ; stopnie_swobody = \bar{n}) (starsza wersja)
→ =ROZKŁ.T.ODWR.DS(prawdopodobieństwo = α ; stopnie_swobody = \bar{n}) (nowsza wersja)

Charakterystyka rozkładu *a posteriori* – [RJn] = bayesowska reinterpretacja MNK

➤ Zauważmy, że w przypadku [RJn] otrzymujemy konkretnie:

- $E(\beta|y) = \hat{\beta} = (X'X)^{-1}X'y$, $E(\beta_i|y) = \hat{\beta}_i$, dla $\bar{n} > 1$
- $Mo(\beta|y) = \hat{\beta}$, $Mo(\beta_i|y) = Mo(\beta_i|y) = \hat{\beta}_i$ (\bar{n} – dowolne)
- $V(\beta|y) = \frac{\bar{s}}{\bar{n}-2} \bar{C}^{-1} = \frac{s(\hat{\beta})}{T-k-2} (X'X)^{-1} = \frac{T-k}{T-k-2} \cdot \underbrace{\frac{s(\hat{\beta})}{T-k} (X'X)^{-1}}_{\hat{V}(\hat{\beta})=s^2(X'X)^{-1}} = \frac{T-k}{T-k-2} \hat{V}(\hat{\beta})$, dla $\bar{n} > 2$

- $D(\beta_i|y) = \sqrt{Var(\beta_i|y)} = \sqrt{[V(\beta|y)]_{ii}} = \sqrt{\frac{T-k}{T-k-2}} \cdot \sqrt{[\hat{V}(\hat{\beta})]_{ii}} = \sqrt{\frac{T-k}{T-k-2}} \cdot \underbrace{D(\hat{\beta}_i)}_{\substack{\text{"klasyczny"} \\ \text{błąd średni} \\ \text{szacunku } \beta_i}}$

→ Zauważmy, że $\frac{T-k}{T-k-2} > 1 \Rightarrow D(\beta_i|y) > D(\hat{\beta}_i)$, czyli bayesowska ocena niepewności wnioskowania o parametrze β_i jest zawsze większa aniżeli „klasyczna” (choć zbiegają w $T \rightarrow \infty$)

→ Ale dlaczego? → I czy to „dobrze”, czy „niedobrze”?

→ I czy w ogóle można porównywać $D(\beta_i|y)$ z $D(\hat{\beta}_i)$?

- $HPD_{1-\alpha}(\beta_i) = \left(\bar{a}_i - t_{\frac{\alpha}{2}} \sqrt{\bar{P}_i^{-1}}, \bar{a}_i + t_{\frac{\alpha}{2}} \sqrt{\bar{P}_i^{-1}} \right) = \left(\hat{\beta}_i - t_{\frac{\alpha}{2}} D(\hat{\beta}_i), \hat{\beta}_i + t_{\frac{\alpha}{2}} D(\hat{\beta}_i) \right) = \text{„klasyczny”}$

$$\Rightarrow \sqrt{\bar{P}_i^{-1}} = \sqrt{\frac{\bar{s}}{\bar{n}} [\bar{C}^{-1}]_{ii}} = \sqrt{\frac{s(\hat{\beta})}{T-k} [(X'X)^{-1}]_{ii}} = \sqrt{[\hat{V}(\hat{\beta})]_{ii}} = D(\hat{\beta}_i)$$

Testowanie pojedynczego parametru strukturalnego

➤ **Bayesowskie testowanie hipotez statystycznych** = „temat rzeka”...

➤ **TU: rozważmy układ hipotez** – typowy na gruncie „klasycznym” ($i \in \{1, 2, \dots, k\}$):

$$H_0: \beta_i = \beta_i^* \quad \text{vs.} \quad H_1: \beta_i \neq \beta_i^*$$

→ **Test Lindley’a**: Czy $\beta_i^* \in \text{HPD}_{1-\alpha}(\beta_i)$ (przy zadanym poziomie p-stwa α *a posteriori* $1 - \alpha$):

- TAK: Testowana wartość β_i^* parametru β_i jest wiarygodna (wspierana przez dane), co wskazuje na zasadność H_0
- NIE: Testowana wartość β_i^* parametru β_i NIE jest wiarygodna (nie jest wspierana przez dane), co wskazuje na zasadność H_1

→ W szczególności – **testowanie istotności pojedynczego parametru**: $\beta_i^* = 0$

$$H_0: \beta_i = 0 \quad \text{vs.} \quad H_1: \beta_i \neq 0$$

Testowanie istotności parametrów strukturalnych 1

➤ W szczególności ~~testowanie istotności pojedynczego parametru: $\beta_t^* = 0$~~

➤ Alternatywnie, do tego zagadnienia można podejść jako do zagadnienia ~~porównywania modeli:~~

$$H_0: \beta_t = 0 \text{ vs. } H_1: \beta_t \neq 0$$

\Leftrightarrow

~~H_0 : Model zredukowany – bez β_t vs. H_1 : Model pełny – z β_t~~

➤ Do zagadnienia porównywania modeli można również sprowadzić ~~testowanie łącznej istotności grupy parametrów~~, których wektor możemy oznaczyć jako $\beta_{(0)}$, przy następującym podziale części deterministycznej modelu:

$$y = X\beta + \varepsilon = X_{(0)} \cdot \underbrace{\beta_{(0)}}_{\substack{\text{podlegają} \\ \text{testowaniu} \\ k_0 \times 1}} + X_{(1)} \cdot \underbrace{\beta_{(1)}}_{\substack{\text{nie podlegają} \\ \text{testowaniu} \\ (k - k_0) \times 1}} + \varepsilon$$

Wtedy

$$H_0: \beta_{(0)} = 0 \text{ vs. } H_1: \beta_{(0)} \neq 0$$

\Leftrightarrow

~~H_0 : Model zredukowany – bez $\beta_{(0)}$ vs. H_1 : Model pełny – z $\beta_{(0)}$~~

➤ ~~Wymóg: w porównywanych modelach musi się dać obliczyć wartość $p(y|Model)$ → [RIn] „odpada”~~

Testowanie istotności parametrów strukturalnych 2

➤ ~~Rozważamy tu tylko przypadek [GNz]~~

➤ ~~Do porównywania modeli potrzebujemy $p(y|M_{\pm})$, które tu możemy oznaczyć jako $p(y|H_0)$ i $p(y|H_{\pm})$, lub równoważnie $p(y|M_0)$ i $p(y|M_{\pm})$, czyli wartości brzegowej gęstości obserwacji w modelach, odpowiednio: zredukowanym (M_0 , odpowiadającym H_0) i pełnym (M_{\pm} , odpowiadającym H_{\pm})~~

➤ ~~Niech M oznacza model regresji z danym zestawem regresorów. Ogólnie~~

$$p(y|M) = \int_{\Theta} p(y, \theta|M) d\theta = \int_{\Theta} p(y|\theta, M) p(\theta|M) d\theta$$

➔ ~~W BMNRL z [GNz] da się ją analitycznie obliczyć, ALE mamy też inny sposób: $p(\theta|y, M) = \frac{p(y, \theta|M)}{p(y|M)}$,~~

wobec czego

$$p(y|M) = \frac{p(y, \theta|M)}{p(\theta|y, M)}$$

➤ ~~W przypadku [GNz] zarówno $p(\theta|y, M)$, jak i $p(y, \theta|M)$ dane są dokładnymi wzorami analitycznymi, więc powyższe działanie da się wykonać:~~

$$p(y, \theta|M) = p(\theta)p(y|\theta) = p(\tau)p(\beta|\tau)p(y|\theta) = f_G\left(\tau \mid \frac{n_{\theta}}{2}, \frac{s_{\theta}}{2}\right) f_N^k(\beta|a, \tau^{-1}C^{-1}) f_N^T(y|X\beta, \tau^{-1}I_T)$$

$$p(\theta|y, M) = p(\beta|\tau, y)p(\tau|y) = f_G\left(\tau \mid \frac{\bar{n}}{2}, \frac{\bar{s}}{2}\right) f_N^k(\beta|\bar{a}, \tau^{-1}\bar{C}^{-1})$$

➔ ~~W $p(y|M)$ parametry nie są obecne, co oznacza, że ich wartości w powyższych zapisach możemy ustalić na dowolnym poziomie, a i tak się skrócą. Proponuje się przyjąć np. $\theta = E(\theta|y)$ lub $\theta = Mo(\theta|y)$, co powinno nas uchronić przed otrzymaniem numerycznego zero dla $f_N^T(y|X\beta, \tau^{-1}I_T)$~~

Testowanie istotności parametrów strukturalnych 3

➔ ~~W MS Excel – brak implementacji wielowymiarowego rozkładu normalnego, ALE:~~

- Przy typowym założeniu niezależności a priori poszczególnych parametrów strukturalnych:

$$\forall_{\substack{i,j=1,\dots,k \\ i \neq j}} \beta_i \perp \beta_j$$

potrzeba zadać diagonalną macierz C^{-1} , a po temu – także diagonalną macierz $V(\beta)$, gdyż

$$C^{-1} = \frac{n_{\theta} - 2}{s_{\theta}} V(\beta)$$

Wtedy też: ~~$p(\beta|\tau) = f_N^k(\beta|a, \tau^{-1}C^{-1}) = \prod_{i=1}^k f_N^1(\beta_i|a_i, [C^{-1}]_{ii})$~~

➔ ~~Gęstość 1-wymiarowego rozkładu normalnego $N(\mu, \sigma^2)$ w argumencie $x \in \mathbb{R}$:~~

$$f_N^1(x|\mu, \sigma^2) = \text{ROZKŁ. NORMALNY}(x; \mu; \sqrt{\sigma^2}; 0)$$

- ~~Podobnie możemy zdekomponować rozkład próbkowy (funkcję wiarygodności), ponieważ – zgodnie z założeniami KMNRL – obserwacje są między sobą niezależne (przy ustalonym θ):~~

$$p(y|\theta) = f_N^T(y|X\beta, \tau^{-1}I_T) = \prod_{t=1}^N f_N^1(y_t|x_t\beta, \tau^{-1})$$

- ~~Dla rozkładu a posteriori: $f_N^k\left(\beta \middle| \bar{a}, \underbrace{\tau^{-1}\bar{C}^{-1}}_{\equiv \Sigma}\right) = (2\pi)^{-\frac{k}{2}} (\det \Sigma)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2}(y - \bar{a})^T \Sigma^{-1}(y - \bar{a})\right\}$~~

Testowanie istotności parametrów strukturalnych 4

➤ Obliczywszy $p(y|M_0)$ (czyli $p(y|H_0)$) oraz $p(y|M_1)$ (czyli $p(y|H_1)$), możemy wyznaczyć iloraz szans *a posteriori*

$$POD_{01} = \frac{\Pr(M_0|y)}{\Pr(M_1|y)} = \frac{\Pr(M_0) p(y|M_0)}{\Pr(M_1) p(y|M_1)}$$

który przy jednakowych *p*-stwach *a priori* obydwu testowanych modeli redukuje się do czynnika Bayesa

$$BF_{01} = \frac{p(y|M_0)}{p(y|M_1)}$$

informującego – w odniesieniu do hipotez – o tym, ilokrotnie bardziej/mniej wiarygodna jest hipoteza H_0 (nieistotność parametrów zgromadzonych w $\beta_{(0)}$ – lub pojedynczego β_i) w stosunku do H_1 (istotność). Tę kwestię możemy również sformułować w kategoriach redukcji modelu M_1 do M_0 : H_0 orzeka, że taka redukcja jest zasadna, podczas gdy H_1 opowiada się za modelem pełnym.

➤ Zauważmy, że $p(y|M)$ możemy wyznaczyć dla dowolnego modelu – tj. z dowolnym zestawem zmiennych objaśniających \Rightarrow możemy porównać moc wyjaśniającą ich wszystkich (**ranking**)

➤ Równanie: $y_t = \beta_1 x_{t1} + \beta_2 x_{t2} + \dots + \beta_k x_{tk} + \varepsilon_t$; jeden z parametrów (np. β_1) jest wyrazem wolnym – zawsze uwzględniany w modelu; pozostaje $k - 1$ regresorów, z których możemy tworzyć różne zestawy

→ Ile jest możliwych takich zestawów:

$$C_{k-1}^1 + C_{k-1}^2 + C_{k-1}^{k-2} + C_{k-1}^{k-1} + \underbrace{C_{k-1}^0}_{\substack{\text{model z samym} \\ \text{tylko wyr. wolnym}}} = 2^{k-1} + 1$$

Testowanie istotności parametrów strukturalnych 5

➤ Skale do porównywania testowanych modeli (umowne!)

→ dla $\log BF_{01} > 0$ (czyli $p(y|M_0) > p(y|M_1)$)

→ określają, z jaką siłą dane świadczą przeciw modelowi M_1 (odrzucają go) na rzecz M_0

| H. Jeffreys (1967) | R.E. Kass i A.E. Raftery (1995) |
|--|---|
| • $0 < \log_{10} BF_{01} \leq 0,5 \rightarrow$ słabo | • $0 < \log_{10} BF_{01} \leq 1 \rightarrow$ słabo |
| • $0,5 < \log_{10} BF_{01} \leq 1 \rightarrow$ solidnie | • $1 < \log_{10} BF_{01} \leq 3 \rightarrow$ dane pozytywnie świadczą przeciw M_0 |
| • $1 < \log_{10} BF_{01} \leq 1,5 \rightarrow$ silnie | • $3 < \log_{10} BF_{01} \leq 5 \rightarrow$ silnie |
| • $1,5 < \log_{10} BF_{01} \leq 2 \rightarrow$ bardzo silnie | • $5 < \log_{10} BF_{01} \rightarrow$ bardzo silnie |
| • $2 < \log_{10} BF_{01} \rightarrow$ zdecydowanie | |