Estymatory postselekcyjne i ich własności

Wykład III

Wisła, grudzień 2009

Estymatory postselekcyjne

$\mathcal{M}$ -zbiór (lista) modeli.
$M \in \mathcal{M}$, $M = \{P_\theta, \theta \in \Theta\}$. Metoda selekcji: na podstawie danych $\mathbf{Y} = (Y_1, \dots, Y_n)$ opisanych przez rozkład P wybieramy $\hat{M}(\mathbf{Y})$: model adekwatnie reprezentujący dane.
Często wybór $\hat{M}$ nie jest celem analizy, $\hat{M}$ jest używany do konstrukcji estymatorów, statystyk testowych, predykcji.
$\hat{\theta}(M)$ - ustalony estymator θ w modelu M (np. NW lub MNK)
Estymator postselekcyjny PMS (post-model selection estimator)

$$\tilde{\theta}_{PMS} = \sum_{M \in \mathcal{M}} \hat{\theta}(M) I\{\hat{M} = M\}.$$

Tutaj: własności predykcyjne metod selekcji opartych na minimalizacji estymatora błędu predykcji i wynikowych estymatorów postselekcyjnych.

Poprawna specyfikacja modelu

Model M poprawnie wyspecyfikowany: Istnieje θ: $P = P_\theta$.
Ma różne znaczenie w zależności od sposobu zdefiniowania M i P. Dla próby $(Y_1, \mathbf{X}_1), \ldots, (Y_n, \mathbf{X}_n)$, równość rozkładów łącznych (lub gęstości łącznych)

$$p(y_1, \mathbf{x}_1, \ldots, \mathbf{x}_n, y_n) = p_\theta(y_1, \mathbf{x}_1, \ldots, \mathbf{x}_n, y_n) \quad (*)$$

lub równość rozkładów warunkowych (lub gęstości warunkowych)

$$p(y_1, \ldots, y_n | \mathbf{x}_1, \ldots, \mathbf{x}_n) = p_\theta(y_1, \ldots, y_n | \mathbf{x}_1, \ldots, \mathbf{x}_n) \quad (**)$$

Czasami warunek słabszy niż $(**)$ np. Shao (1997)

$$E(Y_1, \ldots, Y_n | \mathbf{x}_1, \ldots, \mathbf{x}_n) = E_\theta(Y_1, \ldots, Y_n | \mathbf{x}_1, \ldots, \mathbf{x}_n) \quad (**)$$

Przykład 1 Lista modeli liniowych.
$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$, gdzie $\mathbf{X}$: deterministyczna macierz $n \times p$, $\boldsymbol{\varepsilon} \sim F \in \mathcal{F}$, F-nie zależy od $\boldsymbol{\beta}$. Dla $\tau \in \{0,1\}^P$ i $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)'$

$$M_\tau = \{P_{\beta,F} : \ \boldsymbol{\beta} \in R^p \text{ t.ż. } \beta_i = 0 \text{ gdy } \tau_i = 0, F \in \mathcal{F}\}$$

$$M(k) = \{P_{\beta,F} : \ \boldsymbol{\beta} \in R^p \text{ t.ż. } \beta_i = 0 \text{ gdy } i > k, F \in \mathcal{F}\}$$

$$\mathcal{M}_1 = \{M_\tau, \tau \in \{0,1\}^P\} \quad : 2^p \text{ modeli}$$

$$\mathcal{M}_2 = \{M(k), 0 \leqslant k \leqslant p\} \quad : p+1 \text{ modeli}$$

$$\mathcal{M}_3 = \{M_\tau, \tau \in \{0,1\}^p \& \tau_i = 1, i = 1, 2, \ldots, l\} \quad : 2^{p-l} \text{ modeli.}$$

p może zmieniać się wraz z n: $p = p(n)$.

Przykład 2 Model jednoczynnikowej ANOVA. $\mathbf{Y} = \mathbf{X}\beta + \varepsilon$

$$\mathbf{X}_n = \begin{pmatrix} \mathbf{1}_r & \mathbf{0} & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{1}_r & \mathbf{1}_r & \mathbf{0} & \dots & \mathbf{0} \\ \mathbf{1}_r & \mathbf{0} & \mathbf{1}_r & \dots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{1}_r & \mathbf{0} & \mathbf{0} & \dots & \mathbf{1}_r \end{pmatrix}$$

p poziomów czynnika, na każdym r obserwacji.
$\boldsymbol{\beta} = (\mu_1, \mu_2 - \mu_1, \dots, \mu_p - \mu_1)'$.
$M_1 :\ \boldsymbol{\mu} = (\mu_1, \mu_2, \dots, \mu_p) : \mu_1 = \mu_2 = \dots = \mu_p$.
$M_2 :\ \boldsymbol{\mu} \in R^p$.
$\mathcal{M} = \{M_1, M_2\}$.
Przypadek $p = p(n) \to \infty$. Model dobrze wyspecyfikowany może również zmieniać się wraz z n !

Zgodność i konserwatywność reguły selekcji

Zakładamy, że rodzina modeli $\mathcal{M}$ jest dobrze wyspecyfikowana tzn. istnieje $M \in \mathcal{M}$ i $P_\theta = P \in M$ opisujący dane $\mathbf{Y}$ oraz, że jest domknięta ze względu na przecięcia $(M_1, M_2 \in \mathcal{M} \Rightarrow M_1 \cap M_2 \in \mathcal{M})$.

Wtedy istnieje minimalny model M_{min} o tej własności (minimalny model dobrze wyspecyfikowany).

Reguła wyboru modelu $\hat{M}$ jest zgodna jeśli $P(\hat{M} = M_{min}) \to 1$ gdy $n \to \infty$.

Reguła wyboru modelu $\hat{M}$ jest konserwatywna jeśli *nie* jest zgodna, ale

$$P(M_{min} \subset \hat{M}) \to 1.$$

Uwaga Zgodność (tak rozumiana) ma sens tylko dla list dobrze wyspecyfikowanych !

Problem zgodności AIC dla modelu liniowego

$M_1 \subset M_2$, M_1 oparty na p_0 predyktorach , M_2 na $p_0 + p_1$ predyktorach.
$\varepsilon \sim N(0, \sigma^2 \mathbf{I})$.
$P \in M_1$: Oba modele dobrze wyspecyfikowane.

$$AIC = \frac{RSS_i}{n} + \frac{2\hat{\sigma}_n^2 \times (\text{liczba par. modelu i})}{n}$$

Ale $RSS_i = \varepsilon'(\mathbf{I} - \mathbf{H}_i)\varepsilon$ ($\mathbf{H}_i$- macierz daszkowa dla modelu i)

$$\hat{M} = M_1 \quad \Longleftrightarrow \quad 2\hat{\sigma}_n^2 \Delta > \varepsilon'(\mathbf{H}_2 - \mathbf{H}_1)\varepsilon$$

$\Delta = p_1$ - różnica liczb parametrów. Ale

$$\varepsilon'(\mathbf{H}_2 - \mathbf{H}_1)\varepsilon \sim \sigma^2 \chi^2_\Delta$$

i $\hat{\sigma}_n^2 \to \sigma^2(P)$. Jeśli $\Delta = p_1(n) \to \infty$ $(*)$, to AIC zgodna.

Rozpatrzmy sytuację, gdy $\Delta = p_1$- stała.
Wtedy

$$P(\sigma^2 \chi^2_\Delta > C\sigma^2 \Delta) > 0,$$

zatem

$$\liminf_n P(\hat{M} = M_2) > 0$$

Kryterium AIC <u>nie</u> jest zgodne: stała $C = C_n$ (w miejsce 2) musi dążyć do ∞ !

Zgodność Generalised Information Criterion (GIC)

Twierdzenie 1 Niech $|\mathcal{M}| = p$-ustalona

$$\hat{M} = \mathrm{argmin}_{M \in \mathcal{M}} GIC(M) =: \mathrm{argmin}_{M \in \mathcal{M}} \Big(\frac{RSS(M)}{n} + \frac{p(M) C_n \hat{\sigma}_n^2}{n} \Big),$$

gdzie $p(M)$ liczba zmiennych w modelu M. $\varepsilon \sim N(0, \sigma^2 \mathbf{I}_n)$ i istnieje model $M \in \mathcal{M}$, t.że $P = P_{(\beta, \sigma^2)} = N(\boldsymbol{\beta}, \sigma^2 \mathbf{I}_n) \in M$.
Jeśli $\mathbf{X}'\mathbf{X}$ dodatnio określona dla każdego n !! oraz $n^{-1}\mathbf{X}'\mathbf{X}$ zbiega do pewnej macierzy dodatnio określonej $\mathbf{M}$, $\hat{\sigma}_n^2$ zgodny estymator σ^2 i

(i) $C_n \to \infty$ i $C_n/n \to 0$, to $\hat{M}$ -zgodna;

(ii) $\limsup_n |C_n| < \infty$ to $\hat{M}$ -konserwatywna.

Zgodność a efektywność predykcji

Uwaga Często nie zgodność jest własnością, która interesuje nas najbardziej. Dla $\boldsymbol{\mu}_n = E(\mathbf{Y}|\mathbf{X})$

$$\tilde{M} = \mathrm{argmin}_{M \in \mathcal{M}} Err_{in}(M) = \mathrm{argmin}_{M \in \mathcal{M}} n^{-1} E_{\mathbf{Y}^0}(||\mathbf{Y}^0 - \mathbf{X}(M)\hat{\beta}_M||^2 | \mathbf{X}) =$$

$$\mathrm{argmin}_{M \in \mathcal{M}} \underbrace{(n^{-1}||\boldsymbol{\mu}_n - \mathbf{H}(M)\boldsymbol{\mu}_n||^2 + n^{-1}\varepsilon' \mathbf{H}(M)\varepsilon)}_{L_n(M)}$$

$$P(Err_{in}(\tilde{M}) = Err_{in}(\hat{M})) \to 1 \quad ?? \qquad (*)$$

lub

$$Err_{in}(\hat{M})/Err_{in}(\tilde{M}) \to 1 \ (P) \quad ?? \qquad (**)$$

Oba problemy dobrze postawione również dla list błędnie wyspecyfikowanych.
$(*)$ jest inną własnością zgodności (zgodność predykcyjna).
$(**)$ - efektywność predykcyjna.
Lista może nie zawierać poprawnych modeli, a $(**)$ będzie spełniona (Shao (1997)), tak jest np. przy pewnych warunkach dla k. Akaike.

Twierdzenie 2 (Shao (1997)) Jesli $\hat{\sigma}_n^2$ zgodny estymator wariancji,

$$C_1 n \leqslant \lambda_{1,\mathbf{X}_n'\mathbf{X}_n},\ \lambda_{p,\mathbf{X}_n'\mathbf{X}_n} \leqslant C_2 n,$$

$$\sum_{M:P\notin M} (nE(||\boldsymbol{\mu}_n - \mathbf{X}(M)\hat{\beta}(M)||^2))^{-l} \to 0 \ \text{ dla pewnego } l \in N,$$

to

$$AIC(M) = \frac{\varepsilon'\varepsilon}{n} + \frac{2p\hat{\sigma}^2}{n} - \frac{\varepsilon'\mathbf{H}(M)\varepsilon}{n}$$

dla każdego M takiego, że $P \in M$
oraz

$$AIC(M) = \frac{\varepsilon'\varepsilon}{n} + L_n(M) + o_P(L_n(M))$$

jeśli $P \notin M$, gdzie $L_n(M) = (||\boldsymbol{\mu}_n - \mathbf{H}(M)\boldsymbol{\mu}_n||^2 + \varepsilon'\mathbf{H}(M)\varepsilon)/n$

przy czym $o_P(L_n(M))$ jest jednostajne po wszystkich modelach M takich, że $P \notin M$.

Jeśli $P \in M$ dla dokładnie jednego modelu M i $p_n(M) \to \infty$ to $\left(\frac{p\hat{\sigma}_n^2}{n} - \frac{\varepsilon' \mathbf{H}(M)\varepsilon}{n} = o_P(L_n(M))\right)$

$$\Downarrow$$

$$AIC(M) = \frac{\varepsilon'\varepsilon}{n} + \frac{2p\hat{\sigma}_n^2}{n} - \frac{\varepsilon'\mathbf{H}(M)\varepsilon}{n} = \frac{\varepsilon'\varepsilon}{n} + L_n(M) + o_P(L_n(M))$$

jednostajnie po wszystkich modelach M

$$\Downarrow$$

predykcyjna efektywność AIC

Podobnie $\limsup p_n(M) < \infty \Rightarrow$ zgodność predykcyjna $\Rightarrow$ efektywność predykcyjna.

Ogólniej $Err_{in}(\hat{M})/Err_{in}(\tilde{M}) \to 1\,(P)$ dla $\hat{M}$ wybranego w oparciu o AIC, gdy

- P nie należy do więcej niż jednego modelu na liście (w szczególności żadnego);
- $\#\{M : P \in M\}$ jest ograniczona i $p_n(M_{min}) \to \infty$ lub

  $$min_{M:P\in M, M\neq M_{min}}(p_n(M) - p_n(M_{min})) \to \infty,$$

  gdzie M_{min} jest dobrze wyspecyfikowanym modelem o minimalnej liczbie parametrów.

Przykład 2. Kontynuacja $p \to \infty$, r-ustalone. ($p/n = r^{-1}$ nie dąży do 0 !)

Z tw. Shao wynika, że AIC efektywne predykcyjnie, natomiast dla GIC przy $C_n \to \infty$ $\mathrm{P}(\hat{M}_{GIC} = M_1) \to 1$ i GIC jest efektywne predykcyjnie tylko gdy M_1 jest dobrze wyspecyfikowany.

Dla sytuacji $r \to \infty$, p-ustalone, wymiary obu modeli są ustalone, AIC nie jest efektywne predykcyjnie, gdy średnia odpowiedź nie zależy od poziomu czynnika, a GIC jest zawsze, gdy $C_n \to \infty$ i $C_n/n \to 0$.

Shao (1997): ogólna konkluzja dla modeli liniowych: AIC bardziej predykcyjnie użyteczne, gdy M_{min} nie ma ustalonego wymiaru (lub M_{min} nie istnieje), w przeciwnym przypadku GIC z $C_n \to \infty$ bardziej użyteczne.

Przykład 2. Kontynuacja Efektywność n-krotnej kroswalidacji

$$L_n(M) = \Delta_n(M) + \frac{\varepsilon' \mathbf{H}(M) \varepsilon}{n},$$

gdzie dla $\mu_i = EY_i$

$$\Delta_n(M) = \frac{1}{k} \sum_{i=1}^{k} (\mu_i - \bar{\mu})^2.$$

Jeśli $\Delta_n(M) \to \Delta > 0$, to kroswalidacja n-krotna nie jest efektywna predykcyjnie, jeśli

$$\frac{\sigma^2}{r} < \Delta \leqslant \frac{\sigma^2}{r-1}$$

Odtąd $\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \boldsymbol{\varepsilon}$. Popatrzmy na ryzyko średniokwadratowe

$$E(||\mathbf{Y}^0 - \mathbf{X}\tilde{\boldsymbol{\beta}}||^2|\mathbf{X}) = n\sigma^2 + E(||\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\tilde{\boldsymbol{\beta}}||^2|\mathbf{X})$$

$R_n = E(||\mathbf{X}\boldsymbol{\beta} - \mathbf{X}\tilde{\boldsymbol{\beta}}||^2|\mathbf{X})$ średniokwadratowa miara jakości predykcji.

Związki między zgodnością (w sensie wyboru minimalneo modelu dobrze wyspycyfikowanego) a zachowaniem się ryzyka średniokwadratowego dla estymatora PMS $\tilde{\boldsymbol{\beta}}$.

Lista modeli liniowych zawierających skończoną liczbę predyktorów, $\mathbf{j}_0 = \{1, \ldots, k_0\}$ zbiór indeksów predyktorów odpowiadających niezerowym współczynnikom.

$$\mathbf{J}_1 = \{\mathbf{j} : \mathbf{j}_0 \setminus \mathbf{j} \neq \emptyset\};$$

$$\mathbf{J}_2 = \{\mathbf{j} : \mathbf{j}_0 \subset \mathbf{j}\}.$$

$\hat{\mathbf{j}}$ - reguła wyboru, $\tilde{\beta}$-odpowiadający jej PMS.

Twierdzenie (Nishii (1984)) Jeśli $n^{-1}\mathbf{X}'\mathbf{X}$ zbiega do pewnej macierzy $\mathbf{M}$

Jeśli

i) $nP(\hat{\mathbf{j}} = \mathbf{j}) \to 0$ dla $j \in \mathbf{J}_1$

(ii) $P(\hat{\mathbf{j}} = \mathbf{j}) \to 0$ dla $j \in \mathbf{J}_2 \setminus \mathbf{j}_0$

to

$$R_n(\mathbf{j}_0^c) = E(||\mathbf{X}\tilde{\beta} - \mathbf{X}\beta||^2 I\{\hat{\mathbf{j}} \neq \mathbf{j}_0\}|\mathbf{X}) \to 0.$$

Dwa modele liniowe

$M_1 \subset M_2$, M_1 oparty na p_0 predyktorach , M_2 na $p_0 + p_1$ predyktorach

$$\boldsymbol{\beta} = (\boldsymbol{\beta}_1', \boldsymbol{\beta}_2')'$$

$M_1 : \ \boldsymbol{\beta}_2 = \mathbf{0}$.
$\hat{\mathbf{j}}$: reguła wyboru, $\tilde{\boldsymbol{\beta}}_2$ odpowiadający jej PMS. Z tw. Nishii

$$P(\tilde{\boldsymbol{\beta}}_2 = 0) \to 0$$

implikuje

$$E(||\mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}||^2) \to \sigma^2 p_0 \quad \text{jeśli } \boldsymbol{\beta}_2 = 0$$

i (*)

$$E(||\mathbf{X}\tilde{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}||^2) \to \sigma^2 (p_0 + p_1) \quad \text{jeśli } \boldsymbol{\beta}_2 \neq 0$$

Rozpatrzmy (nieobserwowalną) metodę selekcji:

$$\boldsymbol{\beta}_2 = 0 \Rightarrow \hat{\mathbf{j}} = \{1, 2, \ldots, p_0\} \ i \ \boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}_{MNK,1}$$

$$\boldsymbol{\beta}_2 = 0 \Rightarrow \hat{\mathbf{j}} = \{1, 2, \ldots, p_0 + p_1\} \ i \ \boldsymbol{\beta}^* = \hat{\boldsymbol{\beta}}_{MNK,2}$$

Ryzyko $\boldsymbol{\beta}^*$ ma własność $(*)$.
Ryzyko $\tilde{\boldsymbol{\beta}}$ zachowuje się tak samo jak ryzyko nieobserwowalnego estymatora $\boldsymbol{\beta}^*$ (dla <u>ustalonego</u> $\boldsymbol{\beta}$). 'Własność prorocza'.

Tak jest tylko dla ustalonego $\boldsymbol{\beta}$!

$\mathcal{M}$: lista zawierająca wszystkie podmodele modelu liniowego ustalonego wymiaru p. Błędy o gęstości f, takiej, że $\infty > I(f) > 0$.

Dla zgodnej reguły selekcji i $\hat{\boldsymbol{\beta}} = \hat{\boldsymbol{\beta}}_{PMS}$ na niej opartej

$$\operatorname{limsup}_{\boldsymbol{\beta} \in R^p} E||\mathbf{X}\hat{\boldsymbol{\beta}} - \mathbf{X}\boldsymbol{\beta}||^2 = \infty.$$

(Leeb i Pötscher (2008), również Yang (2005))

Twierdzenie ma znaczenie, gdy parametry model modelu dobrze wyspecyfikowanego zależą od n: $\mathbf{Y}^{(n)} = \mathbf{X}^{(n)}\boldsymbol{\beta}^{(n)} + \varepsilon^{(n)}$.

Twierdzenie $r(\boldsymbol{\beta}) = (r_1, r_2, \ldots, r_p)$, $r_i = I(\beta_i \neq 0)$.
$\hat{\boldsymbol{\beta}}$ jst estymatorem $\boldsymbol{\beta}$ takim, że dla dowolnego $\boldsymbol{\beta} \in R^p$

$$P_{\boldsymbol{\beta}}(r(\hat{\boldsymbol{\beta}}) \leqslant r(\boldsymbol{\beta})) \to 1 \quad n \to \infty$$

($\hat{\boldsymbol{\beta}}$ prawidłowo identyfikuje nieistotne zmienne z pr. dążącym do 1, w szczególności wystarczy, aby $\hat{\boldsymbol{\beta}}$ zgodny).
$\ell : R^p \to R^+$ dowolna nieujemna funkcja straty. Wówczas

$$\sup_{\boldsymbol{\beta}} E(\ell(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \to sup_{s \in R^p} \ell(s).$$

W szczególności dla kwadratowej funkcji straty

$$sup_{\boldsymbol{\beta}} E(n(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \to \infty$$

Ponadto

$$sup_{\beta} E((\mathbf{X}\hat{\beta} - \mathbf{X}\beta)'(\mathbf{X}\hat{\beta} - \mathbf{X}\beta)) \to \infty$$

dla estymatora MNK w pełnym modelu mamy

$$E(n(\hat{\boldsymbol{\beta}}_{MNK} - \boldsymbol{\beta}_{MNK})'(\hat{\boldsymbol{\beta}}_{MNK} - \boldsymbol{\beta}_{MNK})) = tr((n^{-1}\mathbf{X}'\mathbf{X})^{-1}) \to tr(\mathbf{M}^{-1}).$$

Własność estymatora w twierdzeniu odpowiednik złego zachowania się superefektywnego estymatora Hodgesa.

$X_1, \dots, X_n \sim N(\theta, 1)$, $\bar{X}_n^a = \bar{X}\{|\bar{X}| \geqslant n^{-1/4}\} + a\{|\bar{X}| < n^{-1/4}\}$.

Wariancja asymptotyczna $\bar{X}_n^a$ dla $\theta = 0$ równa a^2, natomiast $E_{\theta_n}(\bar{X}_n^a - \theta_n)^2 \to \infty$, gdy $\theta_n = n^{-1/4}$.

Dowód

Dla dowolnego $\mathbf{s} \in R^p$, $\boldsymbol{\beta}_n = -n^{-1/2}\mathbf{s}$

$$\sup_{\beta \in R^p} E_\beta \ell(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \geqslant E_{\beta_n}(\ell(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}_n)) \geqslant$$

$$\ell(-n^{1/2}\boldsymbol{\beta}_n)P_{\beta_n}(r(\hat{\boldsymbol{\beta}}) = 0) = \ell(s)P_{\beta_n}(r(\hat{\boldsymbol{\beta}}) = 0)$$

ale

$$P_{\mathbf{0}}(r(\hat{\boldsymbol{\beta}}) = 0) \to 1$$

i model liniowy jest modelem Asymptotyczne Lokalnie Normalnym (LAN) $\Rightarrow$ ciąg P_{β_n} jest kontygualny względem $P_{\mathbf{0}}$. Zatem

$$P_{\beta_n}(r(\hat{\boldsymbol{\beta}}) = 0) \to 1.$$

$\mathbf{s}$-dowolny, zatem $\sup_{\beta \in R^p} E_\beta \ell(n^{1/2}(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})) \geqslant \sup_{\mathbf{s} \in R^p} \ell(\mathbf{s})$.

Twierdzenie prawdziwe, gdy supremum brane po otoczeniach zera $\theta \in K(0, \rho_n)$ takich, że $n^{1/2}\rho_n \to \infty$.

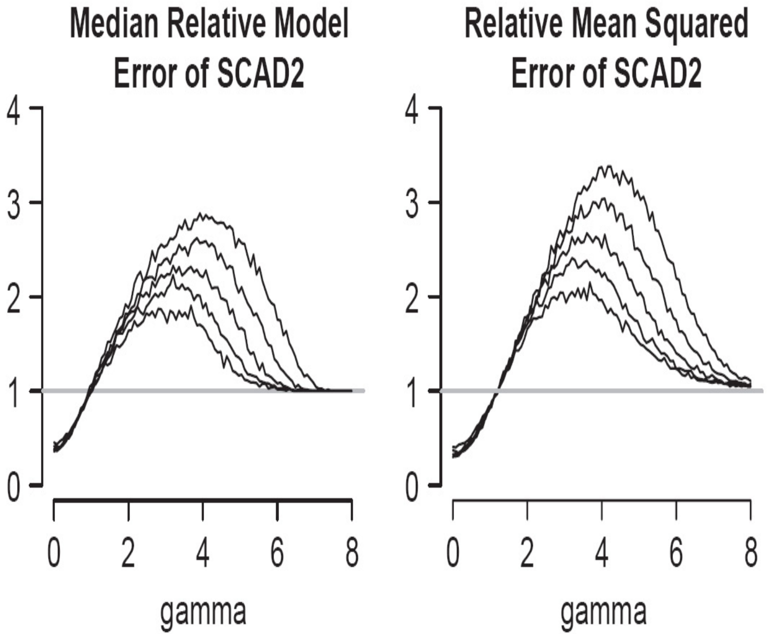
Eksperyment symulacyjny dla modelu liniowego, $p = 8$, $\Sigma_\varepsilon = \mathbf{I}$, , $\mathbf{x}_1 \sim N(0, \Sigma)$, $\Sigma_{\mathbf{X}} = (0.5^{|i-j|})$.
Porównanie estymatora MNK z estymatorem SCAD2 (Fan i Li (2001): wersja regularyzowanego estymatora MNK, mająca własność zgodności.

$\boldsymbol{\beta}_n = \boldsymbol{\beta}_0 + (\gamma/\sqrt{n})(0, 0, 1, 1, 0, 1, 1, 1)'$, $\boldsymbol{\beta}_0 = (3, 1.5, 0, 0, 2, 0, 0, 0)'$, $n = 60, 120, 240, 480, 960$.

Relative Mean Squared Error:
$E(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}))/E(\hat{\boldsymbol{\beta}}_{MNK} - \boldsymbol{\beta}_{MNK})'(\hat{\boldsymbol{\beta}}_{MNK} - \boldsymbol{\beta}_{MNK})$

Model Relative Error: $Me(\hat{\boldsymbol{\beta}})/Me(\hat{\boldsymbol{\beta}}_{MNK})$, gdzie

$$Me(\hat{\boldsymbol{\beta}}) = (\hat{\boldsymbol{\beta}} - \boldsymbol{\beta})'\Sigma(\hat{\boldsymbol{\beta}} - \boldsymbol{\beta}).$$


Median Relative Model
Error of SCAD2
4
3
2
1
0
0
2
4
6
8
gamma
Relative Mean Squared
Error of SCAD2
4
3
2
1
0
0
2
4
6
8
gamma