

Estymacja funkcji wariancji dla procesu autoregresji

Piotr Borkowski, Jan Mielniczuk

Instytut Podstaw Informatyki PAN

10 września 2008

Niech dany będzie model liniowy z błędami heteroskedastycznymi

$$Y = X\beta + \varepsilon$$

gdzie Y jest wektorem zmiennej objaśnianej o wymiarze $n \times 1$, $X = [x_{ij}]$ macierzą o wymiarze $n \times p$, ε wektorem nzl błędów wymiaru $n \times 1$ takim, że $\mathbb{E}\varepsilon = 0$ o macierzy kowariancji $\Sigma_\varepsilon = \text{diag}(\exp(Z\eta))$.

Macierz Z o wymiarze $n \times q$ jest macierzą eksperymentu dla macierzy kowariancji, może być ona różna od macierzy X .

Niech f_θ będzie łączną gęstością próby $Y_1, \dots, Y_n$ zależną od parametru $\theta = (\beta', \eta')'$. Funkcja wiarygodności próby $Y_1, \dots, Y_n$ pochodzącej z rozkładu sparametryzowanego wektorem parametrów θ ma postać

$$\mathcal{L}(Y, \theta) = f_\theta(Y_1, \dots, Y_n)$$

Metody szukania ekstremów dla funkcji $f = -\log \mathcal{L}(Y, \theta)$

- Metoda największego spadku

$$\theta_{i+1} = \theta_i - \lambda_i \nabla f(\theta_i), \lambda_i > 0$$

- Metoda Gaussa-Newtona

$$\theta_{i+1} = \theta_i - (Hf(\theta_i))^{-1} \nabla f(\theta_i),$$

- Metoda Levenberga-Marquardta

$$\theta_{i+1} = \theta_i - (Hf(\theta_i) + \lambda I)^{-1} \nabla f(\theta_i),$$

Model liniowy wykorzystamy, by modelować sytuację dla szeregu autoregresyjnego $(X_t)_{t \in \mathbb{N}}$ spełniającego

$$X_{t+1} = m(X_t) + \sigma(X_t)\varepsilon_{t+1}, \quad t = 0, 1, \dots,$$

gdzie $m(\cdot)$ i $\sigma(\cdot)$ są funkcjami rzeczywistymi, $(\varepsilon_t)_{t \in \mathbb{N}}$ ciąg iid taki, że $\mathbb{E}(\varepsilon_t) = 0$, $\text{Var}(\varepsilon_t) = 1$ oraz ε_{t+1} jest niezależny od historii procesu $\mathcal{F}_t = \sigma(X_0, \dots, X_t)$ do czasu t .

Zauważmy, że $\mathbb{E}(X_{t+1}|X_t) = m(X_t)$ tzn. że $m(\cdot)$ jest funkcją regresji X_{t+1} przy danym X_t , oraz $\text{Var}(X_{t+1}|X_t) = \sigma^2(X_t)$.

Zakładamy, że gęstość $f_\varepsilon(\cdot)$ istnieje oraz $\sigma(x) > 0$ dla każdego $x \in \mathbb{R}$. Dla rozpatrywanego procesu gęstość warunkowa X_{i+1} przy zadanym $X_i = x_i$ jest równa

$$f(x_{i+1}|x_i) = \frac{1}{\sigma(x_i)} f_\varepsilon\left(\frac{x_{i+1} - m(x_i)}{\sigma(x_i)}\right).$$

Pozwala to dla danej trajektorii $X_1, \dots, X_n$ wyliczyć jej funkcję wiarygodności przy ustalonym $X_1 = x_1$:

$$\mathcal{L}_n(\theta) = \log f_\theta(x_1, x_2, \dots, x_n | x_1) = \sum_{i=1}^{n-1} \log f_\theta(x_{i+1} | x_i)$$

Do modelowania gęstości $f(x_2|x_1)$ wykorzystamy rodzinę parametryczną $\mathcal{F}_{kl} = \{f_\theta(x_2|x_1)\}_{\theta \in \Theta_{kl}}$, gdzie $\Theta_{kl} \subset \mathbb{R}^{k+l}$ i $f_\theta(x_2|x_1)$ jest zdefiniowana z $m(\cdot) = m_\theta(\cdot)$ oraz $\sigma(\cdot) = \sigma_\theta(\cdot)$, gdzie

$$m_\theta(x) = \sum_{i=1}^k \beta_i \phi_i(x),$$

$$\sigma_\theta(x) = \exp \left\{ \sum_{i=1}^l \eta_i \psi_i(x) \right\},$$

$(\phi_i(x))_{i=1}^k$ i $(\psi_i(x))_{i=1}^l$ dwa zbiory funkcji liniowo niezależnych

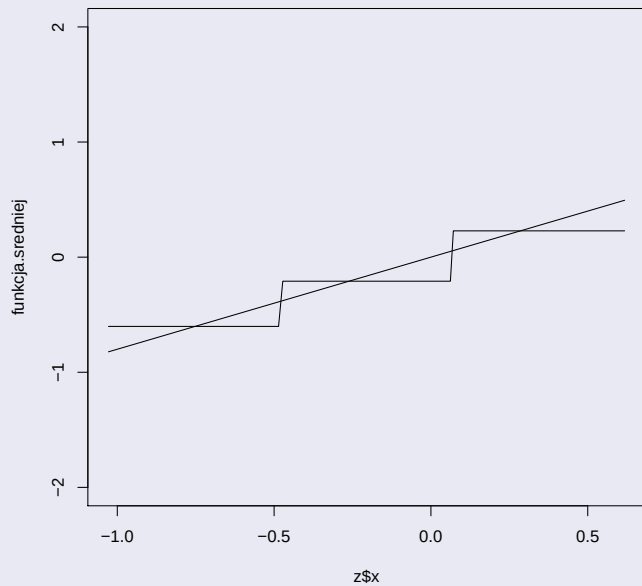
Rozpatrujemy parametryzację $f(x_2|x_1)$ na ustalonym odcinku $[a, b]$.

Dzielimy go na p przedziałów, gdzie $1 \leq p \leq P$ oraz

$\phi_i(x) = (\phi_{i,0}(x), \dots, \phi_{i,s-1}(x))'$ oznacza wektor wielomianów Legendre'a rzędu $0, 1, \dots, s-1$, określonych dla i -tego przedziału podziału. Niech $b_i = (\beta_{i,0}, \dots, \beta_{i,s-1})'$ będzie wektorem współczynników.

$$m_{sp}(x) = \sum_{i=1}^p b_i' \phi_i(x).$$

$k = sp$ -liczba parametrów



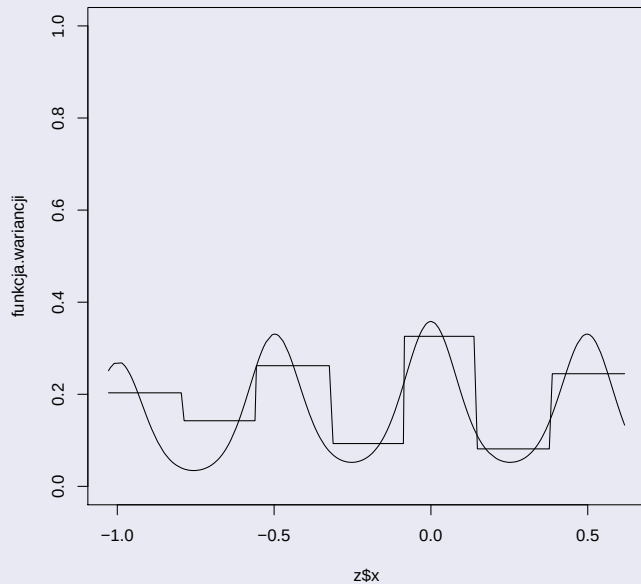
Analogicznie dla logarytmu funkcji wariancji.

Dla ustalonego t i $q \leq Q$ rozpatrujemy podział $[a, b]$ na q części, na każdej z nich definiujemy wielomiany Legendre'a $\psi_j(x)$ rzędu $t - 1$.

$e_j = (\eta_{j,0}, \dots, \eta_{j,t-1})' \in \mathbb{R}^t$ - wektor współczynników.

$$\sigma_{tq}(x) = \exp \left\{ \sum_{j=1}^q e_j' \psi_j(x) \right\}.$$

$q \leq Q$, $l = tq$ liczba parametrów



Powyższe parametryzacje pozwalają przedstawić model autoregresyjny jako model liniowy z odpowiednią macierzą X oraz Z .

- estymator ML
- estymator REML (Restricted Maximum Likelihood)

$$\mathcal{L}_n(\theta) = \log f_{\theta}(x_1, x_2, \dots, x_n | x_1) = \sum_{i=1}^{n-1} \log f_{\theta}(x_{i+1} | x_i)$$

Definiujemy $\hat{\theta}_{kl}$ estymator ML

$$\hat{\theta}_{ML}^{kl} = \operatorname{argmax}_{\theta \in \Theta_{kl}} \mathcal{L}_n(\theta)$$

Funkcja kryterialna z karą za liczbę parametrów (BIC)

$$-\mathcal{L}_n^{kl}(\hat{\theta}_{kl}) + 1/2 \cdot (k + l) \log n$$

Estymator *PMS* (Post-model-selection), to estymator ML w modelu dla którego funkcja kryterialna jest najmniejsza, tzn rozpatrujemy $\hat{\theta}_{ML}^{\tilde{k}\tilde{l}}$, gdzie $\tilde{k}\tilde{l} = \operatorname{argmin}\{-\mathcal{L}_n^{kl}(\hat{\theta}_{kl}) + 1/2 \cdot (k + l) \log n, 1 \leq k \leq k_0, 1 \leq l \leq l_0, \}$

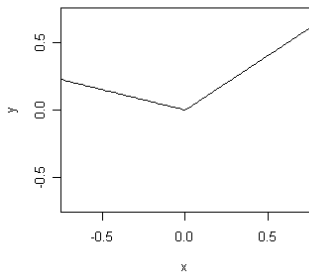
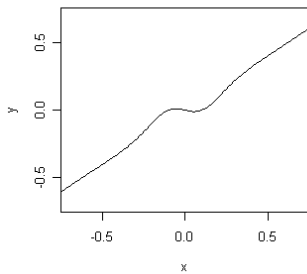
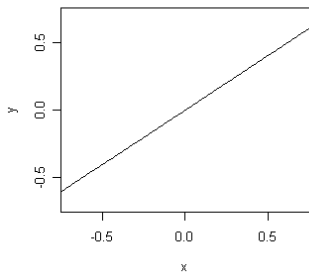
Iteracyjna metoda rozwiązywania równania ML

- Zastępujemy hesjan logarytmu funkcji wiarygodności przez jego wartość oczekiwaną (znamy postać macierzy informacji Fishera)
- Estymatory $\hat{\beta}$ i $\hat{\eta}$ wyznaczone są naprzemiennie. $\beta_{(m+1)}$ jest określony m.in na podstawie $\eta_{(m)}$, natomiast $\eta_{(m+1)}$ na podstawie $\beta_{(m)}$ i $\eta_{(m)}$
- Zamiast metody Gaussa-Newtona stosowaliśmy metodę Levenberga-Marquardta

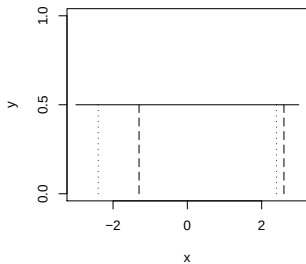
- Idea REML: rzutujemy wektor Y na podprzestrzeń ortogonalną do kolumn X . Rzut Y nie zależy wtedy od parametrów b . Estymacja Σ jest oparta na rzutowanym Y a nie bezpośrednio na Y .
- Implementacja estymatora REML jest dostępna jako algorytm *remlscore* Smith'a (2002) w pakiecie *statmod*.

Badane estymatory

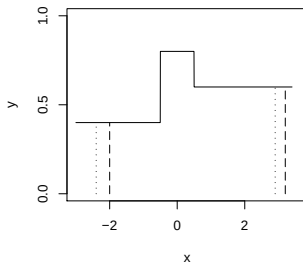
- Estymator PMS dla ML ($S1$) z $(s, t) = (2, 1)$ dla $k_0 = 2 \cdot 10, l_0 = 1 \cdot 15$
- Zmodyfikowany PMS ($S1m$) (łączymy środki otrzymanego histogramu)
- Estymator lokalnie liniowy ($LL1$) (wykorzystuje *dpill* z pakietu *kernsmooth*)
- Estymator lokalnie liniowy ($LL2$) (oparty na kodzie Fana i Yao (1998))



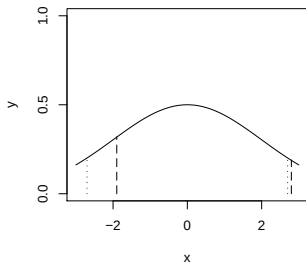
Std. dev. σ_1



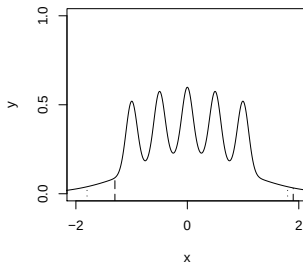
Std. dev. σ_2

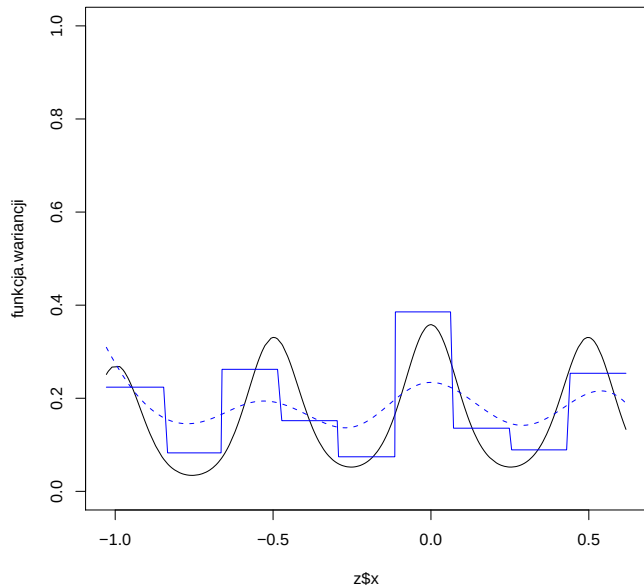


Std. dev. σ_3



Std. dev. σ_4





Porównanie estymatorów

Średnie oraz odchylenia standardowe dla $EISE \times 10^4$

	m_1			
	S1	S1m	LL1	LL2
σ_1	2.89(4.5)	2.78(3.95)	9.64(11.32)	17.28(9.43)
σ_2	135.99(52.91)	165.03(60.95)	140.61(34.78)	116.6(38.31)
σ_3	86.89(41.42)	59.9(51.43)	25.62(18.6)	45.9(27.83)
σ_4	61.15(19.46)	63.29(21.33)	100.64(19.51)	92.68(20.92)
	m_2			
	S1	S1m	LL1	LL2
σ_1	3.39(5.55)	3.01(4.14)	13.53(64.09)	16.48(8.72)
σ_2	141.23(67.86)	147.99(64.13)	101.22(40.65)	96.5(32.57)
σ_3	83.88(38.56)	54.79(42.48)	33.59(73.75)	43.25(26.46)
σ_4	53.51(15.33)	53(16.09)	100.95(15.87)	92.66(22.01)

$$EISE = \frac{1}{n-1} \sum_{i=2}^n (\tilde{\sigma}^2(X_i) - \sigma^2(X_i))^2,$$