Nesimultánní změny ve složkách náhodného vektoru

Daniela JARUŠKOVÁ

FSv ČVUT, Katedra matematiky, Thákurova 7, CZ 166 29 Praha 6

V časových okamžicích $i = 1, \ldots, n$ pozorujeme posloupnost nezávislých dvourozměrných vektorů $\{(X_1(i), X_2(i))\}$ takových, že $corr(X_1(i), X_2(i)) = \rho$. (Pro začátek předpokládejme, že jsou studované vektory normálně rozdělené.)

Předpokládáme, že v každé souřadnici došlo v neznámém čase ke skoku o neznámé velikosti ve střední hodnotě, přičemž doby, ve kterých dochází ke skoku se nemusí shodovat.

Odhady

Problém:

$$X_1(i) = a + e_1(i),\ i = 1, \ldots, k_1^*, \qquad X_2(i) = b + e_2(i),\ i = 1, \ldots, k_2^*,$$
$$X_1(i) = e_1(i),\ i = k_1^* + 1, \ldots, n, \qquad X_2(i) = e_2(i),\ i = k_2^* + 1, \ldots, n.$$

První nápad: Odhadněme zvlášť k_1^* použitím pouze $\{X_1(i)\}$ a k_2^* použitím pouze $\{X_2(i)\}$.

Odhad bodu změny v jedné časové řadě

Parametr a známý:

$$
\begin{aligned}
\widetilde{k} &= \operatorname{argmin}_k \left\{ \sum_{i=1}^{k} \left(X(i) - a\right)^2 + \sum_{i=k+1}^{n} X(i)^2 \right\} \\
&= \operatorname{argmax}_k \\
&\quad \Big\{ a^2(k - k^*) - 2a \sum_{i=k}^{k^*} e_i + \sum_{i=1}^{n} e_i^2 \quad \text{pro} \quad k \leq k^*, \\
&\quad - a^2(k - k^*) + 2a \sum_{i=k^*}^{k} e_i + \sum_{i=1}^{n} e_i^2 \quad \text{pro} \quad k \geq k^* \Big\}.
\end{aligned}
$$

$$a_n^2(\widetilde{k} - k^*) \xrightarrow{D} \xi,$$

kde $\xi = \operatorname{argmax}\{-|s| + 2\,W^*(s),\ s \in R_1\}\big\}$. Hustota i distribuční funkce ξ jsou známé.

Parametr a neznámý:

$$\widetilde{k} = \mathrm{argmin}_{[n\beta]\le k\le n}\left\{ \min_a \sum_{i=1}^{k} \left(X(i)-a\right)^2 + \sum_{i=k+1}^{n} X(i)^2 \right\} =$$

$$= \mathrm{argmax}_{[n\beta]\le k\le n}\left\{ \frac{1}{k}\Big(\sum_{i=1}^{k} X(i)\Big)^2 - \frac{1}{k^*}\Big(\sum_{i=1}^{k^*} X_1(i)\Big)^2 \right\} =$$

$$= \mathrm{argmax}_{[n\beta]\le k\le n}\left\{ \chi_1^2(k) - \chi_1^2(k^*) \right\} =$$

$$= \mathrm{argmax}_{k}$$

$$\Big\{ -a^2(k^*-k) - 2a\sum_{i=k}^{k^*} e_i + \frac{1}{k}\big(\sum_{i=1}^{k} e_i\big)^2 - \frac{1}{k^*}\big(\sum_{i=1}^{k^*} e_i\big)^2 \quad \text{pro} \quad k \le k^*,$$

$$-a^2(k-k^*)\frac{k^*}{k} + 2a\sum_{i=k^*}^{k} e_i + \frac{1}{k}\big(\sum_{i=1}^{k} e_i\big)^2 - \frac{1}{k^*}\big(\sum_{i=1}^{k^*} e_i\big)^2 \quad \text{pro} \quad k \ge k^*. \Big\}$$

$$
\begin{aligned}
D(k,k^*) &= a^2(k^* - k) \quad \text{pro} \quad k \le k^*, \\
D(k,k^*) &= a^2(k - k^*)\frac{k^*}{k} \quad \text{pro} \quad k \ge k^*.
\end{aligned}
$$

It holds $D(k,k^*) \ge const\, a^2 |k - k^*|$.

Předpokládejme, že $a_n\sqrt{n} \to \infty$, potom pro všechna $C > 0$

$$\max_{a_n^2|k-k^*|\leq C} \left| \left(\chi^2(k)-\chi^2(k^*)\right)-\left(-a_n^2\,|k-k^*|+2a_n\sum_{i=k}^{k^*} e(i)\right) \right| = o_P(1).$$

$$a_n^2(\widetilde{k} - k^*) = O_P(1).$$

Rozdělení

$$\operatorname{argmax}_{[\beta n]\le k_1\le n} \frac{1}{k_1}\Big(\sum_{i=1}^{k_1} X_1(i)\Big)^2 \text{je dáno}\{e_1(i), |i-k_1^*| \le C_1/a_n^2\},$$
$$\operatorname{argmax}_{[\beta n]\le k_2\le n} \frac{1}{k_2}\Big(\sum_{i=1}^{k_2} X_2(i)\Big)^2 \text{je dáno}\{e_2(i), |i-k_2^*| \le C_2/b_n^2\}.$$

$\frac{1}{n\,a_n^2} \to 0$ a $\frac{1}{n\,b_n^2} \to 0 \Rightarrow \widetilde{k}_1$ and $\widetilde{k}_2$ jsou asymptoticky nezávislé a oba mají stejné rozdělení jako $\xi = \operatorname{argmax}\{-|s| + 2\,W^*(s),\ s \in R_1\}\}$.

Zaveďme $Z(i) = (X_2(i) - \rho X_1(i)) / \sqrt{1-\rho^2}, i = 1, \dots, n$.
Pro $k_1 \leq k_2$

$$\begin{aligned} Z(i) &= \frac{b - \rho a}{\sqrt{1-\rho^2}} + v(i), \quad i = 1, \dots, k_1, \\ &= \frac{b}{\sqrt{1-\rho^2}} + v(i), \quad i = k_1 + 1, \dots, k_2, \\ &= v(i), \quad i = k_2 + 1, \dots, n. \end{aligned}$$

Pro $k_1 \geq k_2$

$$\begin{aligned} Z(i) &= \frac{b - \rho a}{\sqrt{1-\rho^2}} + v(i), \quad i = 1, \dots, k_2, \\ &= \frac{-\rho a}{\sqrt{1-\rho^2}} + v(i), \quad i = k_2 + 1, \dots, k_1, \\ &= v(i), \quad i = k_1 + 1, \dots, n. \end{aligned}$$

Zaveďme $S_x(k) = \sum_{i=1}^k X_1(k)$ a $S_z(k) = \sum_{i=1}^k Z(i)$.

Velikost skoku $X_1(i)$ v k_1 je $|a|$. Velikost skoku $Z(i)$ v k_1 je $\frac{|\rho a|}{\sqrt{1-\rho^2}}$ a v k_2 je $\frac{|b|}{\sqrt{1-\rho^2}}$. Kvadrát normy velikosti skoku v k_1 je $\frac{a^2}{1-\rho^2}$. Kvadrát normy velikosti skoku v k_2 je $\frac{b^2}{1-\rho^2}$.

Odhady

$$(\widehat{k}_1, \widehat{k}_2) = \operatorname{argmax}_{[\beta n] \le k_1 \le n, [\beta n] \le k_2 \le n} \chi^2(k_1, k_2),$$

kde pro $k_1 \le k_2$

$$\chi^2(k_1, k_2) = \frac{\Big(S_z(k_2)\Big)^2}{k_2} + \\ + \frac{\frac{k_2}{k_1}\Big(\sqrt{1-\rho^2}(S_x(k_1)) - \rho\big(S_z(k_1) - \frac{k_1}{k_2} S_z(k_2)\big)\Big)^2}{k_2 - \rho^2 k_1}$$

a pro $k_2 \le k_1$

$$\chi^2(k_1, k_2) = \frac{\Big(S_z(k_2)\Big)^2}{k_2} + \\ + \frac{\Big(\sqrt{1-\rho^2}(S_x(k_1)) - \rho\big(S_z(k_1) - S_z(k_2)\big)\Big)^2}{k_1 - \rho^2 k_2}.$$

$$D(k_1, k_2, k_1^*, k_2^*) \geq const(a^2 + b^2) \cdot (|k_1 - k_1^*| + |k_2 - k_2^*|)$$

$$\Delta_n^2|\widehat{k}_1 - k_1^*| = O_P(1) \quad \text{and} \quad \Delta_n^2|\widehat{k}_2 - k_2^*| = O_P(1).$$

$$\mathcal{M}_C = \{(k_1, k_2); |k_1 - k^*| \le C/\Delta_n^2 \cap |k_2 - k^*| \le C/\Delta_n^2\}.$$

$$\max_{(k_1,k_2)\in\mathcal{M}_C} \Bigg| \chi^2(k_1, k_2) - \chi^2(k_1^*, k_2^*) - \Bigg(- \frac{a_n^2|k_1 - k_1^*|}{1-\rho^2} - \frac{b_n^2|k_2 - k_2^*|}{1-\rho^2} +$$
$$+ \frac{2a_n}{\sqrt{1-\rho^2}} \operatorname{sign}(k_1 - k_1^*) \sum_{i=\min(k_1,k_1^*)+1}^{\max(k_1,k_1^*)} \left(\sqrt{1-\rho^2} e_1(i) - \rho v(i)\right) +$$
$$+ \frac{2b_n}{\sqrt{1-\rho^2}} \operatorname{sign}(k_2 - k_2^*) \sum_{i=\min(k_2,k_2^*)+1}^{\max(k_2,k_2^*)} v(i) \Bigg) \Bigg| = o_P(1). \qquad (1)$$

$k_1^* = [n\tau_1^*],\ k_2^* = [n\tau_2^*],\ \tau_1^* \neq \tau_2^*$

Asymptotické rozdělení $(\widehat{k}_1, \widehat{k}_2)$ as $n \to \infty$. Předpokládejme, že $a_n \to 0$, $b_n \to 0$ tak, aby $0 < L_1 < |b_n/a_n| < L_2$. Označme $\Delta_n^2 = a_n^2 + b_n^2$ a předpokládejme, že $\Delta_n \sqrt{n} \to \infty$.

Věta

Za shora uvedených předpokladů

$$\frac{a_n^2}{1-\rho^2}(\widehat{k}_1 - k_1^*) \xrightarrow{D} \xi_1,$$
$$\frac{b_n^2}{1-\rho^2}(\widehat{k}_2 - k_2^*) \xrightarrow{D} \xi_2,$$

kde ξ_1 a ξ_2 mají asymptoticky stejné rozdělení jako $\xi = \text{argmax}\{-|s| + 2\,W^*(s),\ s \in R_1\}\}$. Navíc jsou odhady $\widehat{k}_1$ a $\widehat{k}_2$ asymptoticky nezávislé.

Co se stane, když $\tau_1^* = \tau_2^* = \tau^*$?

Marginální rozdělení $a_n^2(\widetilde{k}_1 - k^*)$ a $b_n^2(\widetilde{k}_2 - k^*)$ se nezmění, ale odhady budou korelované.

Předpokládejme, že $a_n/b_n \to L$, pak vektor $\left(a_n^2(\widehat{k}_1 - k^*)/(1-\rho^2), b_n^2(\widehat{k}_2 - k^*)/(1-\rho^2)\right)$ konverguje v distribuci

$$(t^0, s^0) = \operatorname{argmax}_{-\infty<t<\infty, -\infty<s<\infty} X(t, s).$$

$$
X(t,s) =
\begin{aligned}
&= t + s(1-2L\rho) + 2W_1^*(t) + 2W_2^*(s), && t<0, s<0, |t| \geq L^2|s|,\\
&= t(1-2\frac{1}{L}\rho) + s + 2W_1^*(t) + 2W_2^*(s), && t<0, s<0, |t| < L^2 s,\\
&= -t + s + 2W_1^*(t) + 2W_2^*(s), && t \geq 0, s \leq 0,\\
&= t - s + 2W_1^*(t) + 2W_2^*(s), && t \leq 0, t \geq 0,\\
&= -t(1-2\frac{1}{L}\rho) - s + 2W_1^*(t) + 2W_2^*(s), && t>0, s>0, t<L^2 s,\\
&= -t - s(1-2L\rho) + 2W_1^*(t) + 2W_2^*(s), && t>0, s>0, t>L^2 s,
\end{aligned}
$$

kde $\{W_1^*(t), \infty < t < \infty\}$ a $\{W_2^*(t), \infty < t < \infty\}$ jsou náhodné procesy definované:

$$\begin{aligned} W_1^*(t) &= W_{11}(t),\ t \geq 0, & \qquad W_2^*(t) &= W_{21}(t),\ t \geq 0, \\ &= W_{12}(-t),\ t < 0, & &= W_{22}(-t),\ t < 0, \end{aligned}$$

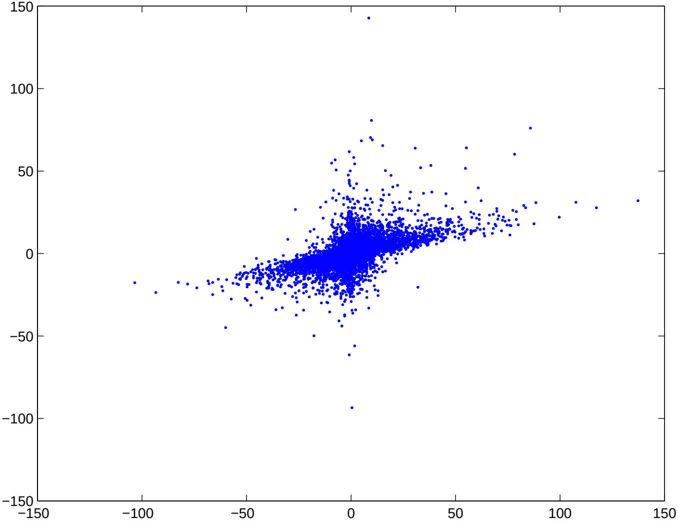
a $\{W_{11}(t), t \geq 0\}$, $\{W_{12}(t), t \geq 0\}$, $\{W_{21}(t), t \geq 0\}$, $\{W_{22}(t), t \geq 0\}$ jsou Wienerovy procesy takové, že

$$cov\big(W_{11}(t), W_{12}(s)\big) = -\rho \min(t, s), \qquad cov\big(W_{12}(t), W_{21}(s)\big) = 0,$$
$$cov\big(W_{11}(t), W_{21}(s)\big) = 0, \qquad cov\big(W_{12}(t), W_{22}(s)\big) = 0,$$
$$cov\big(W_{11}(t), W_{22}(s)\big) = 0, \qquad cov\big(W_{21}(t), W_{22}(s)\big) = -\rho \min(t, s).$$

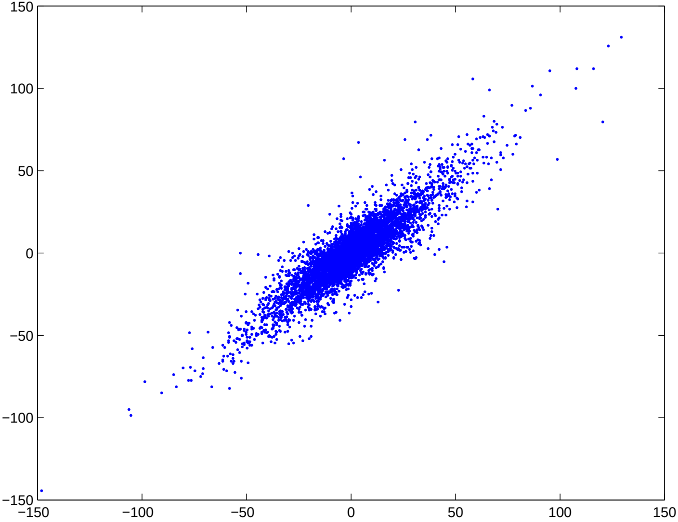
$$\xi(t) = \inf\{s; X(t,s) = \max_{0\le s} X(t,s)\},$$
$$t^0 = \inf\{\widetilde{t}; X\big(\widetilde{t}, \xi(\widetilde{t})\big) = \max_{0\le t} X\big(t, \xi(t)\big)\},$$
$$s^0 = \xi(t^0).$$

V důsledku zákona o iterovaném logaritmu pro Wienerovy procesy platí $P(t^0 < \infty \cup s^0 < \infty) = 1$.

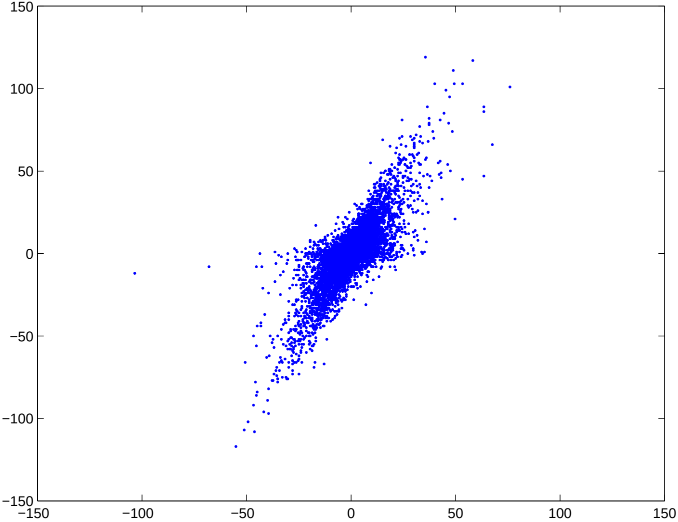
Rozdělení (t^0, s^0) závisí na L. Za předpokladu, že $\tau_1^* = \tau_2^* = \tau^*$ rozdělení $(\widehat{k}_1 - k^*, \widehat{k}_2 - k^*)$ nezávisí na τ^* a typu rozdělení $\{\big(e_1(i), e_2(i)\big)\}$.



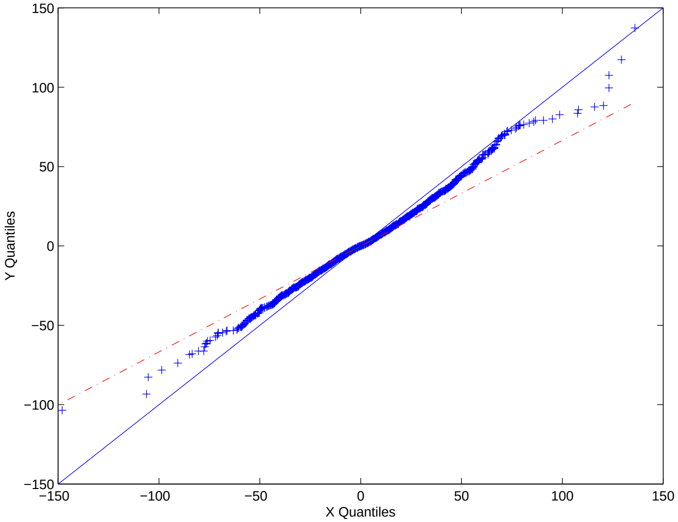
scatterplot of $(\widehat{k}_1, \widehat{k}_2)$ for $a_n = 0.4,\ b_n = 0.2,\ \rho = 0.8$



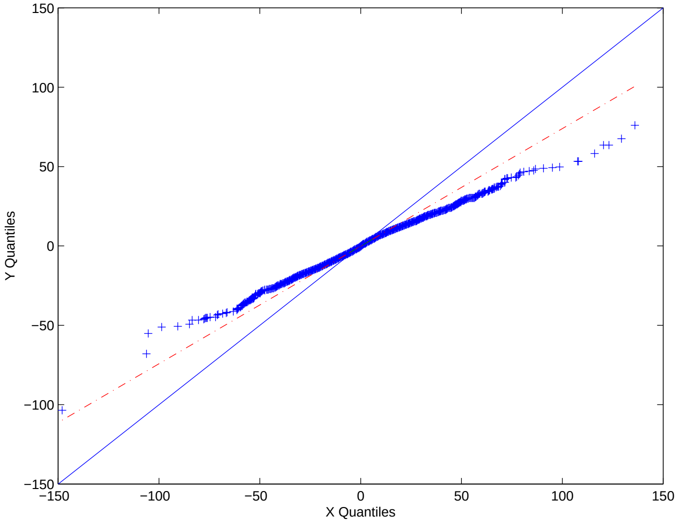
scatterplot of $(\widehat{k}_1, \widehat{k}_2)$ for $a_n = 0.4,\ b_n = 0.4,\ \rho = 0.8$



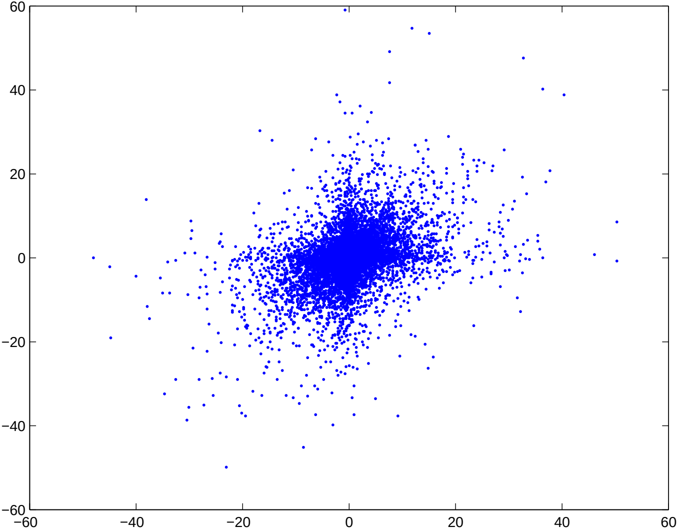
scatterplot of $(\widehat{k}_1, \widehat{k}_2)$ for $a_n = 0.4,\ b_n = 0.6,\ \rho = 0.8$



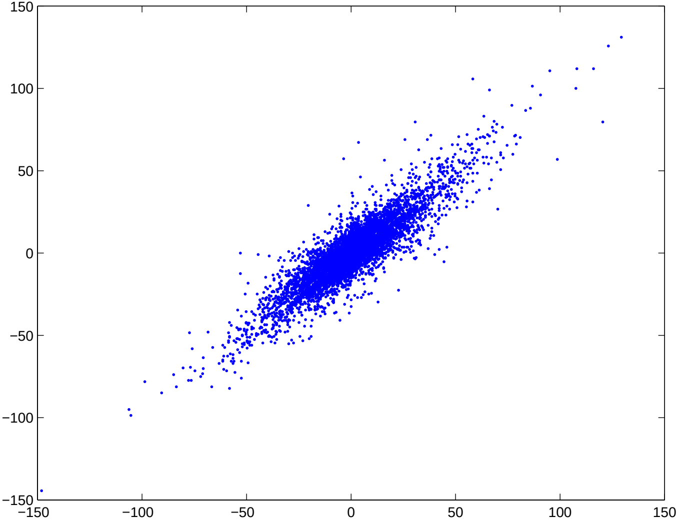
qqplot of $\widehat{k_1}$ for $a_n = 0.4,\ b_n = 0.4$ and $\widehat{k_1}$ for $a_n = 0.4,\ b_n = 0.2, \quad \rho = 0.8$



qqplot of $\widehat{k}_1$ for $a_n = 0.4,\ b_n = 0.4$ and $\widehat{k}_1$ for $a_n = 0.4,\ b_n = 0.6, \quad \rho = 0.8$



scatterplot of $(\widehat{k}_1, \widehat{k}_2)$ for $a_n = 0.4,\ b_n = 0.4,\ \rho = 0.4$



scatterplot of $(\widehat{k}_1, \widehat{k}_2)$ for $a_n = 0.4,\ b_n = 0.4,\ \rho = 0.8$

$$a_n = b_n = 0.4, \quad \rho = 0.4 \quad Var\left(\frac{0.4^2}{1-0.4^2}|k_1 - k_1^*|\right) = 41.68$$

$$a_n = b_n = 0.4, \quad \rho = 0.8 \quad Var\left(\frac{0.4^2}{1-0.8^2}|k_1 - k_1^*|\right) = 264.76$$

Chceme odhadnout rozdíl $\tau_1^* - \tau_2^*$ za předpokladu $a_n = b_n$

Jestliže $\tau_1^* \neq \tau_2^*$, $\widehat{k}_1 - \widehat{k}_2$ rozdíl dvou asymptoticky nezávislých veličin se známým asymptotickým rozdělením.

Jestliže $\tau_1^* = \tau_2^*$, rozdělení pomocí simulace

	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
$a_n = 0.4$	11.8	11.4	11.8
$a_n = 0.6$	11.8	10.9	12.0

Tabulka 95% horních kvantilů $a_n^2(\widehat{k}_1^L - \widehat{k}_2^L)/(1-\rho^2)$.

	$\rho = 0$	$\rho = 0.4$	$\rho = 0.8$
$a_n = 0.4$	16.3	15.6	15.6
$a_n = 0.6$	15.8	14.4	16.0

Tabulka 97.5% horních kvantilů $a_n^2(\widehat{k}_1^L - \widehat{k}_2^L)/(1-\rho^2)$.

1. Bai J., Perron P.: Estimating and testing linear models with multiple structural changes, Econometrica 66 (1998), 47 – 78.

2. Hušková M.: Estimators for epidemic alternatives, Comment. Math. Univ.Carolinae 36, 2 (1995), 279 – 291.

3. Yao Y.- C.: Approximating the distribution of the maximum likelihood estimate of the change-point in a sequence of independent random variables, Ann. Statist. 15 (1987), 1321 – 1328.