ROBUST'98, 149 – 160

UŽITÍ POŘADOVÝCH TESTŮ PRO DETEKCI ZÁVISLOSTI V ČASOVÝCH ŘADÁCH

Jan PICEK [1] **a Jana JUREČKOVÁ**[2]

TU Liberec, KAP a MFF UK, KPMS

Abstract. This article describes non-parametric tests of independence of two autoregressive time series and tests of the order of the autoregression in a time series model. The main tool of these tests is autoregression rank scores. They are asymptoticaly distribution-free and do not require any estimation of the unknown autoregression parametres. The tests are aplied to series of daily temperatures and the daily mortality related to cardio-vascular problems.
Резюме: Эта статья описывает критерии независимости двух авторегрессий в модели временных рядов. Критерии выведёны на основе авторегрессионных ранговых меток. Эти критерии независят от распределения и для них не нужна оценка неизвестного авторегрессивного параметра. Критерии были применены на ряды дневной температуры и дневной смертности по кардио-васкуларным причинам.

1. Úvod

Regresní pořadové skóry zavedli Gutenbrunner a Jurečková [1]. Gutenbrunner, Jurečková, Koenker a Portnoy [2] použili tohoto pojmu ke konstrukci třídy lineárních regresních pořadových testů o parametrech lineárního regresního modelu s nenáhodnou regresní maticí. Jurečková [8] sestrojila nelineární testy Kolmogorov-Smirnovova typu založené na regresních pořadových skórech. Picek [14, 15] rozšířil lineární regresní pořadové testy na lineární modely s náhodnou nebo smíšenou regresní maticí. Algoritmus pro výpočet regresních pořadových skórů popsali Koenker a d'Orey [11, 12].

Koul a Saleh [13] zobecnili pojem regresních pořadových skórů na autoregresní (AR) model a zavedli *autoregresní pořadové skóry* (APS). Třídu testů lineárních hypotéz o parametrech AR modelu založených na APS sestrojili Hallin a Jurečková [5]. Tyto testy jsou v práci [7] ilustrovány na simulovaných datech a použity k identifikaci autoregresního řádu meteorologických časových řad maximálních teplot naměřených ve třech stanicích na jižní Moravě.

Hallin, Jurečková, Picek a Zahaf [6] zkonstruovali testy nezávislosti dvou autoregresních posloupností, založené na APS, a použili tyto testy ve studii vlivu vysokých teplot na úmrtí z kardio-vaskulárních příčin v bruselské oblasti v období 1980 – 89.

[1] Research supported by the Charles University Grant GAUK 198/1996 and by the Ministry of Education of Czech Republic VS97084.

[2] Research supported by the Czech Republic Grant 201/96/0230.

V této práci nejprve definujeme pojem regresních a autoregresních pořadových skórů a popíšeme jejich základní vlastnosti, nutné k pochopení struktury a fungování statistických procedur na nich založených (paragrafy 2 a 3). V paragrafu 4 popíšeme testy o identifikaci řádu autoregresní posloupnosti a v paragrafu 5 testy nezávislosti dvou autoregresních posloupností, založených na APS. V paragrafu 6 ilustrujeme tyto testy na reálných datech.

2. Regresní kvantily a regresní pořadové skóry

Uvažujme lineární regresní model

$$\mathbf{Y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{E}, \tag{2.1}$$

kde $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_p)' \in I\!R^p$ je neznámý parametr, $\mathbf{X}$ je $n \times p$ rozměrná matice, $\mathbf{Y} = (Y_1, \ldots, Y_n)'$ je vektor pozorování a $\mathbf{E} = (E_1, \ldots, E_n)'$ je vektor nezávislých stejně rozdělených chyb.

Regresní pořadové skóry v modelu (2.1) jsou úzce spjaty s *regresními kvantily,* s nimiž jsou ve vzájemně duálním vztahu. Pro jejich pochopení je proto vhodné definovat oba pojmy zároveň.

Regresním α-kvantilem $\widehat{\boldsymbol{\beta}}\,(\alpha)$ nazýváme každý vektor $\mathbf{t} \in I\!R^p$, který je řešením úlohy

$$\sum_{i=1}^{n} \rho_\alpha(Y_i - \mathbf{x}_i'\mathbf{t}) := \min, \tag{2.2}$$

kde

$$\rho_\alpha(x) = x\psi_\alpha(x), \;\; x \in I\!R^1 \tag{2.3}$$

a

$$\psi_\alpha(x) = \alpha - I[x < 0], \;\; x \in I\!R^1. \tag{2.4}$$

Tento pojem zavedli Koenker a Basset [10], kteří ukázali, že asymptotické vlastnosti regresních kvantilů jsou naprosto analogické vlastnostem empirických kvantilů v modelu se stejně rozdělenými nezávislými pozorováními. Titíž autoři charakterizovali regresní α - kvantil $\widehat{\boldsymbol{\beta}}(\alpha)$ jako složku $\widehat{\boldsymbol{\beta}}$ optimálního řešení úlohy parametrického lineárního programování:

$$\alpha\mathbf{1}_n'\mathbf{u}^+ + (1-\alpha)\mathbf{1}_n'\mathbf{u}^- := \min$$

$$\mathbf{X}\widehat{\boldsymbol{\beta}} + \mathbf{u}^+ - \mathbf{u}^- = \mathbf{Y} \tag{2.5}$$

$$\widehat{\boldsymbol{\beta}} \in I\!R^p, \mathbf{u}^+, \mathbf{u}^- \in I\!R_+^n \quad 0 < \alpha < 1$$

Optimální řešení $\hat{\mathbf{a}}(\alpha) = (\hat{a}_1(\alpha), \ldots, \hat{a}_n(\alpha))' \in \mathbb{R}^n$ duální úlohy

$$\begin{aligned} \mathbf{Y}_n'\hat{\mathbf{a}} &:= \max \\ \mathbf{X}_n'\hat{\mathbf{a}} &= (1-\alpha)\mathbf{X}'\mathbf{1}_n \qquad (2.6) \\ \hat{\mathbf{a}} &\in [0,1]^n \end{aligned}$$

nazvali Gutenbrunner a Jurečková [1] *regresními pořadovými skóry.*

Motivací názvu *pořadové* skóry je skutečnost, že v modelu polohy nabývají regresní pořadové skóry tvaru:

$$\hat{a}_{ni}(\alpha) \equiv a_n^*(R_i, \alpha) \equiv \begin{cases} 1 & \text{jestliže } \alpha \le (R_i - 1)/n \\ R_i - \alpha n & \text{jestliže } (R_i - 1)/n < \alpha \le R_i/n \\ 0 & \text{jestliže } R_i/n < \alpha, \end{cases}$$

kde R_i je pořadí Y_i mezi $Y_1, \ldots, Y_n$. Proces $a_n^*(j,\alpha),\ j = 1, \ldots, n,\ 0 < \alpha < 1$ je přesně týž, jaký použil Hájek [3] ke konstrukci nelineárních pořadových testů (viz též Hájek a Šidák [4], kapitola V.3.5).

Z duality mezi $\widehat{\boldsymbol{\beta}}(\alpha)$ a $\hat{\mathbf{a}}(\alpha)$ dostaneme následující nerovnice:

$$\hat{a}_i(\alpha) = \begin{cases} 1 & \text{jestliže } E_i > \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha)) \\ 0 & \text{jestliže } E_i < \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha)) \end{cases} \quad i = 1, \ldots, n$$

a rovnici

$$\sum_{i \in M_\alpha} \hat{a}_i(\alpha)\mathbf{x}_i = (1-\alpha)\sum_{i=1}^n \mathbf{x}_i - \sum_{i=1}^n I[E_i > \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha))]\mathbf{x}_i,$$

kde

$$M_\alpha = \{i : E_i = \mathbf{x}_i'(\widehat{\boldsymbol{\beta}}(\alpha) - \boldsymbol{\beta}(\alpha))\}, \quad \text{a} \quad \#M_\alpha = p$$

Z vlastností regresních pořadových skórů při konečném n je patrně nejdůležitější jejich *invariance vzhledem k regresi s maticí* $\mathbf{X}$:

$$\hat{\mathbf{a}}(\alpha, \mathbf{Y} + \mathbf{X}\mathbf{b}) = \hat{\mathbf{a}}(\alpha, \mathbf{Y}) \text{ pro } \forall\ b \in \mathbb{R}^p, \qquad (2.7)$$

která je rozšířením invariance pořadí (nebo pořadových skórů) vzhledem k posunutí v poloze.

Z dalších zajímavých neasymptotických vlastností regresních kvantilů a pořadových skórů uveďme (viz [9]):

- $\sum_{i=1}^n \mathbf{x}_i'\hat{\boldsymbol{\beta}}(\alpha) = -\sum_{i=1}^n Y_i \frac{d}{d\alpha}\hat{a}_i(\alpha)$ ve všech bodech spojitosti $\hat{\boldsymbol{\beta}}(\alpha)$, $0 < \alpha < 1$.
- Obě strany této rovnosti jsou neklesající schodovité funkce $\alpha \in (0,1)$.

3. Autoregresní pořadové skóry

Uvažujme autoregresní model

$$X_t = \theta_1 X_{t-1} + \ldots + \theta_p X_{t-p} + \varepsilon_t, \quad t = 0, \pm 1, \ldots, \tag{3.1}$$

kde $\{\varepsilon_t\}$ je posloupnost nezávislých a stejně rozdělených náhodných veličin (bílý šum) s distribuční funkcí F a hustotou f (inovační hustota), obecně neznámou. Předpokládáme, že $\theta_1, \ldots, \theta_p$ splňují obvyklé předpoklady týkající se stacionarity:

$$1 - \sum_{i=1}^{p} \theta_i z^i \neq 0 \qquad |z| \leq 1.$$

Autoregresní kvantily modelu (3.1) jsou definovány jako řešení minimalizace

$$\sum_{t=1}^{n} \rho_\alpha (X_t - b_0 - Y_{t-1} b_1 - \ldots - Y_{t-p} b_p) := \min, \ \mathbf{b} = (b_0, \ldots, b_p)' \in \mathbb{R}^{p+1}$$

kde ρ_α je funkce definovaná v (2.3).

Autoregresní pořadové skóry v modelu (3.1), duální k autoregresním kvantilům, jsou definovány jako optimální řešení $\hat{\mathbf{a}}_n(\alpha) = (\hat{a}_{n1}(\alpha), \ldots, \hat{a}_{nn}(\alpha))'$ úlohy parametrického lineárního programování

$$\begin{aligned}
& X_1 \hat{a}_{n1}(\alpha) + X_2 \hat{a}_{n2}(\alpha) + \ldots + X_n \hat{a}_{nn}(\alpha) := \max \\
& \sum_{t=1}^{n} \hat{a}_{nt}(\alpha) = (1-\alpha) n \\
& \sum_{t=1}^{n} X_{t-i} \hat{a}_{nt}(\alpha) = (1-\alpha) \sum_{t=1}^{n} X_{t-i}, \ \ i = 1, \ldots, p \\
& \hat{\mathbf{a}}_n(\alpha) \in [0,1]^n, \quad 0 \leq \alpha \leq 1.
\end{aligned} \tag{3.2}$$

Algebraické vlastnosti regresních kvantilů a pořadových skórů platí i v modelu (3.1). Hlavní výhodou autoregresních pořadových skórů je, podobně jako v lineárním modelu, že jsou invariantní vzhledem k rušivým autoregresním parametrům, a že tedy autoregresní pořadové skóry, vypočtené pro $X_1, \ldots, X_n$, jsou rovny autoregresním pořadovým skórům, vypočteným pro (nepozorovatelný) bílý šum.

Ostatní vlastnosti vyplývají z paragrafu 2; dále

- $\hat{a}_{nt}(\,.\,)$ jsou spojité, po částech lineární funkce α, $\alpha \in [0,1]$;
- $\hat{a}_{nt}(0) = 1$ a $\hat{a}_{nt}(1) = 0$;
- pro libovolné pevné $\alpha \in (0,1)$, přesně p hodnot $\hat{a}_{n1}(\alpha), \ldots, \hat{a}_{nn}(\alpha)$ leží mezi 0 a 1.

4. Testy o řádu AR modelu založené na autoregresních pořadových skórech

Uvažujme nejprve testování hypotézy $AR(p-1)$ proti $AR(p)$:

$$\text{(4.1)} \qquad \mathbf{H}_0 : \theta_p = 0, \qquad (\theta_1, \ldots, \theta_{p-1}) \text{ nespecifikované.}$$

$$\mathbf{H}_1 : \theta_p \neq 0, \qquad (\theta_1, \ldots, \theta_p) \text{ nespecifikované.}$$

Za platnosti $\mathbf{H}_0$ má uvažovaný autoregresní model tvar

$$\text{(4.2)} \qquad X_t = \theta_1 X_{t-1} + \ldots + \theta_{p-1} X_{t-p+1} + \varepsilon_t \quad t = 1, \ldots, n.$$

Test konstruujeme v několika následujících krocích:

- Vypočteme autoregresní pořadové skóry $\hat{a}_{n1}(\alpha), \ldots, \hat{a}_{nn}(\alpha)$ za platnosti $\mathbf{H}_0$, *tj.* pro model (4.2).
- Vybereme neklesající *skórovou funkci*

  $\varphi : (0,1) \mapsto I\!R^1$ takovou, že $\varphi(1-u) = -\varphi(u),\ 0 < u < 1$ a

  $$0 < A^2(\varphi) = \int_0^1 \varphi^2(u) du < \infty,$$

  a vypočítáme *skóry* $\hat{b}_{n1}, \ldots, \hat{b}_{nn}$ generované skórovou funkcí φ následovně:

$$\text{(4.3)} \qquad \hat{b}_{nt} = -\int_0^1 \varphi(u) d\hat{a}_{nt}(u), \quad t = 1, \ldots, n,$$

- Na základě těchto skórů určíme *lineární autoregresní pořadovou skórovou statistiku*

$$\text{(4.4)} \qquad S_\varphi^{(n)} = n^{-1/2} \sum_{t=1}^{n} X_{t-p} \hat{b}_{nt}.$$

- Nakonec spočteme *testovou statistiku*, která má potom tvar

$$\text{(4.5)} \qquad T_\varphi^{(n)} = \frac{S_\varphi^{(n)}}{A(\varphi) D_n},$$

  kde

  $$D_n^2 = n^{-1} \mathbf{Y}_{II}' [\mathbf{I}_n - \mathbf{Y}_I (\mathbf{Y}_I' \mathbf{Y}_I)^{-1} \mathbf{Y}_I'] \mathbf{Y}_{II},$$
  $$\mathbf{Y}_{II} = (X_{1-p}, X_{2-p}, \ldots, X_{n-p})'$$

  a

  $$\mathbf{Y}_I = \begin{bmatrix} X_0 & X_1 & \ldots & X_{p-2} \\ X_1 & X_2 & \ldots & X_{p-1} \\ \ldots\ldots & & & \\ X_{n-1} & X_n & \ldots & X_{p-3+n} \end{bmatrix}$$

  je matice $n \times (p-1)$.

Výběr skórové funkce φ je podobný jako v případě běžných pořadových testů; typické výběry φ jsou:

i) *Wilcoxonovy skóry* (optimální pro logistické rozdělení): $\varphi(u) = u - \frac{1}{2}$, $0 < u < 1$ a skóry jsou rovny $\hat{b}_{nt} = -\int_0^1 (u - \frac{1}{2}) d\hat{a}_{nt}(u) = \int_0^1 \hat{a}_{nt}(u)du - \frac{1}{2}$ a $A^2(\varphi) = \frac{1}{12}$.

ii) *Normální (Van der Waerdenovy) skóry* (optimální pro normální rozdělení): $\varphi(u) = \Phi^{-1}(u)$, $0 < u < 1$, kde Φ je distribuční funkce normálního rozdělení $\mathcal{N}(0,1)$; $A^2(\varphi) = 1$.

iii) *Mediánové (znaménkové) skóry* (optimální pro dvojitě-exponenciální [Laplaceovo] rozdělení): $\varphi(u) = \frac{1}{2}\mathrm{sign}(u - \frac{1}{2})$, skóry $\hat{b}_{nt} = \hat{a}_{nt}(\frac{1}{2}) - \frac{1}{2}$, $t = 1, \ldots, n$, a $A^2(\varphi) = 1$.

Předpokládejme, že inovační hustota f s distribuční funkcí F vyhovují následujícím podmínkám:

(F1) $\int_{-\infty}^{\infty} x f(x) dx = 0, \qquad \int_{-\infty}^{\infty} x^2 f(x) dx < \infty.$

(F2) $0 < f(x) < C, C > 0$ a f má omezené derivace $f', \ f'', \quad x \in \mathbb{R}^1$.

(F3) f je klesající pro $x \to \pm\infty$, a

$$\lim_{x \to -\infty} \frac{-\log F(x)}{b|x|^r} = 1, \qquad \lim_{x \to +\infty} \frac{-\log(1 - F(x))}{bx^r} = 1$$

pro nějaké $b > 0, \ r \geq 1$.

Označme $\mathcal{F}$ systém hustot, vyhovujícím (F1) - (F3). Hallin and Jurečková [5] dokázali, že pro hustoty $f \in \mathcal{F}$ má testová statistika $T_\varphi^{(n)}$, za hypotézy $\mathbf{H}_0$, asymptoticky normální rozdělení $\mathcal{N}(0,1)$, přičemž Pitmanova vydatnost testu proti lokálním alternativám je stejná jako u příslušných pořadových testů bez rušivých parametrů.

Tedy hypotézu $\mathbf{H}_0$ zamítáme na hladině významnosti $\alpha \in (0,1)$, když

$$|T_\varphi^{(n)}| \geq u_{\alpha/2} \overset{def}{=} \Phi^{-1}(1 - \frac{\alpha}{2}). \tag{4.6}$$

Třída $\mathcal{F}$ obsahuje nejen hustotu normálního rozdělení, ale také všechny hladké unimodální hustoty s exponenciálně klesajícími chvosty. Numerické ilustrace ukazují, že testy pracují i pro Cauchyho rozdělení.

Opakovanou aplikací testu provedeme **identifikaci řádu autoregrese**: Provedeme posloupnost testů $AR(p)$ proti $AR(p+1)$, $p = 0, 1, 2, \ldots$ na dané hladině významnosti α; jako řád sledovaného procesu potom přijmeme nejmenší číslo p_0, pro které hypotéza $AR(p_0)$ není zamítnuta ve prospěch $AR(p_0 + 1)$; je-li $p_0 = 0$ je proces identifikován jako bílý šum.

5. Testy nezávislosti dvou autoregresních časových řad založené na ARS

Nechť $(X_1, Y_1), (X_2, Y_2), \ldots, (X_n, Y_n)$ jsou pozorování odpovídající následujícím autoregresním modelům:

$$(5.1)\quad X_t = \theta_1 X_{t-1} + \ldots + \theta_p X_{t-p} + \varepsilon_t$$

$$t = 0, \pm 1, \pm 2, \ldots,$$

$$(5.2)\quad Y_t = \vartheta_1 Y_{t-1} + \ldots + \vartheta_q Y_{t-q} + \delta_t,$$

kde $\{\varepsilon_t\}$ a $\{\delta_t\}$ jsou dvě posloupnosti nezávislých stejně rozdělených náhodných veličin s hustotami f a g. Předpokládáme, že tyto hustoty a parametry $\theta_1, \ldots, \theta_p$ a $\vartheta_1, \ldots, \vartheta_q$ nejsou známé, pouze o nich předpokládáme, že splňují podmínky stacionarity, tj.

$$1 - \sum_{i=1}^{p} \theta_i z^i \neq 0 \qquad \text{a} \qquad 1 - \sum_{i=1}^{q} \vartheta_i z^i \neq 0, \qquad |z| \leq 1.$$

Chceme testovat hypotézu $\mathbf{H}_0$ o nezávislosti mezi dvěma autoregresními posloupnostmi $\{X_t\}$ a $\{Y_t\}$ bez znalosti f a g a rušivých parametrů $\theta_1, \ldots, \theta_p$ a $\vartheta_1, \ldots, \vartheta_q$, t.j.

$$\mathbf{H}_0 : \; X_{t_1}, Y_{t_2} \quad \text{nezávislé } \forall t_1, t_2.$$

Následující neparametrické testy $\mathbf{H}_0$, založené na *autoregresních pořadových skórech*, navrhli Hallin, Jurečková, Picek, Zahaf [6]. Test konstruujeme v několika krocích:

- Spočteme autoregresní pořadové skóry v obou časových řadách, tj. najdeme optimální řešení $\hat{\mathbf{a}}_n^X(\alpha) = (\hat{a}_{n1}^X(\alpha), \ldots, \hat{a}_{nn}^X(\alpha))'$, resp. $\hat{\mathbf{a}}_n^Y(\alpha) = (\hat{a}_{n1}^Y(\alpha), \ldots, \hat{a}_{nn}^Y(\alpha))'$ úloh parametrického lineárního programování

$$X_1 \hat{a}_{n1}^X(\alpha) + X_2 \hat{a}_{n2}^X(\alpha) + \ldots + X_n \hat{a}_{nn}^X(\alpha) := \max$$

$$\sum_{t=1}^{n} \hat{a}_{nt}^X(\alpha) = (1-\alpha)n$$

$$\sum_{t=1}^{n} X_{t-i} \hat{a}_{nt}^X(\alpha) = (1-\alpha) \sum_{t=1}^{n} X_{t-i}, \quad i = 1, \ldots, p$$

$$\hat{\mathbf{a}}_n^X(\alpha) \in [0,1]^n, \quad 0 \leq \alpha \leq 1.$$

resp.

$$\begin{aligned}
&Y_1\hat{a}^Y_{n1}(\alpha)+Y_2\hat{a}^Y_{n2}(\alpha)+\ldots+Y_n\hat{a}^Y_{nn}(\alpha):=\max\\
&\sum_{t=1}^{n}\hat{a}^Y_{nt}(\alpha)=(1-\alpha)n\\
&\sum_{t=1}^{n}Y_{t-i}\hat{a}^Y_{nt}(\alpha)=(1-\alpha)\sum_{t=1}^{n}Y_{t-i},\quad i=1,\ldots,q\\
&\hat{\mathbf{a}}^Y_n(\alpha)\in[0,1]^n,\quad 0\le\alpha\le 1.
\end{aligned}$$

- Vybereme neklesající *skórovou funkci*

  $\varphi:(0,1)\mapsto \mathbb{R}^1$ takovou, že $\varphi(1-u)=-\varphi(u),\ 0<u<1$ a

$$0<A^2(\varphi)=\int_0^1\varphi^2(u)du<\infty.$$

- Vypočítáme *skóry* pro první řadu generované skórovou funkcí φ,

$$\hat{b}^X_{nt}=-\int_0^1\varphi(u)d\hat{a}^X_{nt}(u),\quad t=1,\ldots,n.$$

- Podobně vypočteme skóry pro druhou řadu

$$\hat{b}^Y_{nt}=-\int_0^1\varphi(u)d\hat{a}^Y_{nt}(u),\quad t=1,\ldots,n.$$

- Na základě těchto skórů spočteme statistiky

$$S^{(n)}_\nu=n^{-1/2}(A(\varphi))^{-2}\sum_{t=1}^{n}\hat{b}^X_t\hat{b}^Y_{t-\nu},\quad \nu=k_1,\ldots,k_2,$$

  kde k_1 a k_2 jsou předepsaná celá čísla.
- Nakonec vypočítáme testové kritérium, které má tvar

$$T_{n;k_1,k_2}=\sum_{\nu=k_1}^{k_2}(S^{(n)}_\nu)^2,$$

  přičemž hypotézu $\mathbf{H}_0$ zamítáme na hladině významnosti $\alpha\in(0,1)$, je-li

$$T_{n;k_1,k_2}>\chi^2_{k_2-k_1+1}(\alpha),$$

  kde $\chi^2_k(\alpha)$ označuje $(1-\alpha)$ kvantil χ^2 rozdělení s k stupni volnosti.

Poznamenejme, že $n^{-1/2}S^{(n)}_\nu$ lze interpretovat jako pořadovou míru korelace mezi ε_t and $\delta_{t-\nu}$. Užijeme-li Wilcoxonovu skórovou funkci ($\varphi(u)=u-\frac{1}{2}$), pak v tomto kontextu lze hovořit o *Spearmanově pořadovém korelačním koeficientu.*

Wilcoxonovy skóry mají tvar

$$\begin{aligned} \hat{b}^X_{nt} &= -\int_0^1 (u-\frac{1}{2})d\hat{a}^X_{nt}(u) &= \int_0^1 \hat{a}^X_{nt}(u)du - \frac{1}{2} \\ \hat{b}^Y_{nt} &= -\int_0^1 (u-\frac{1}{2})d\hat{a}^Y_{nt}(u) &= \int_0^1 \hat{a}^Y_{nt}(u)du - \frac{1}{2} \end{aligned} \qquad t = 1, \ldots, n,$$

a $A^2 = \frac{1}{12}$. Statistika Spearmanova typu $S^{(n)}_\nu$ má potom tvar

$$S^{(n)}_\nu = 12n^{-1/2} \sum_{t=1}^n \hat{b}^X_t \hat{b}^Y_{t-\nu}, \quad \nu = 0, 1, \ldots, K.$$

6. Numerická ilustrace

Hallin et al. [6] studovali vliv vysokých teplot na úmrtí z kardio-vaskulárních příčin. Využijeme dat z této studie tj. průměrných denních teplot a počtu úmrtí v letech 1980-1989 v bruselské oblasti pro ilustraci uvedených testů. Budeme na rozdíl od uvedené studie zkoumat závislost pouze mužské úmrtnosti ve vztahu k teplotě.

Abychom mohli použít test pro detekci závislosti, potřebujeme určit řád autoregrese jednotlivých časových řad, a to postupem popsaným ve čtvrtém paragrafu. Poznamenejme, že teploty i počet úmrtí jsou centrovány jejich denními průměry (vypočtenými z dostupných 10-ti let) a že se omezíme v jednotlivých letech pouze na letní období.

Postupně jsme uvažovali model $AR(1)$ $X_t = \theta_1 X_{t-1} + \varepsilon_t$, kde jsme testovali hypotézu

$$\mathbf{H}_0 : \theta_1 = 0 \quad \text{proti} \quad \mathbf{H}_1 : \theta_1 \neq 0;$$

Dále model $AR(2)$ $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \varepsilon_t$, s testem

$$\mathbf{H}^*_0 : \theta_2 = 0, \quad \theta_1 \text{ nespecifikované},$$

proti

$$\mathbf{H}^*_1 : \theta_2 \neq 0, \quad \theta_1, \theta_2 \text{ nespecifikované}$$

a nakonec model $AR(3)$ $X_t = \theta_1 X_{t-1} + \theta_2 X_{t-2} + \theta_3 X_{t-3} + \varepsilon_t$, kde jsme testovali hypotézu

$$\mathbf{H}^{**}_0 : \theta_3 = 0, \quad \theta_1, \theta_2 \text{ nespecifikované},$$

proti

$$\mathbf{H}^{**}_1 : \theta_3 \neq 0, \quad \theta_1, \theta_2, \theta_3 \text{ nespecifikované}.$$

V r. 1981 nám postupně testové statistiky (4.5) pro teploty vyšly 8.05 -3.56 1.58. Tedy na hladině $\alpha = 0.05$ zamítáme hypotézy $\mathbf{H}_0$, $\mathbf{H}^*_0$ a nezamítáme hypotézu $\mathbf{H}^{**}_0$. Protože i v dalších letech dostáváme analogické výsledky, budeme uvažovat, že letní teplotní řada má řád 2. Pro řadu úmrtí vycházejí v uvažovaných modelech tyto hodnoty testové statistiky -2.36 1.26 2.06.

V tomto případě na hladině $\alpha = 0.05$ zamítáme hypotézu $\mathbf{H}_0$ a nezamítáme hypotézy $\mathbf{H}_0^*$ a $\mathbf{H}_0^{**}$. Můžeme tedy uvažovat, že tato řada má řád 1.

Dále jsme prováděli analýzu závislostí těchto dvou řad v jednotlivých letech postupem popsaným v předešlém paragrafu. Výsledky shrneme do tří tabulek. V první tabulce jsou uvedeny hodnoty Wilcoxonovy (Spearmanovy) autoregresní pořadové statistiky S_ν, měřící korelaci mezi teplotou v čase t a úmrtím v čase $t + \nu$, $\nu = 0, \ldots, 5$, pro 10 letních období v letech 1980 až 1989. Pro úplnost poznamenejme, že $1 - \frac{\alpha}{2}$ kvantil normálního rozdělení je 1.96 pro $\alpha = 0.05$ a 2.58 pro $\alpha = 0.01$. V posledním řádku tabulky jsou uvedeny součty $\sum_{1980}^{1989} S_\nu^2$ (pro pevné ν). Opět pro úplnost $1 - \alpha$ kvantil χ^2 rozdělení s 10 stupni volnosti je 18.31 pro $\alpha = 0.05$ a 23.21 pro $\alpha = 0.01$. Významné hodnoty na hladině $\alpha = 0.05$ jsou označeny hvězdičkou.

Rok	$\nu = 0$	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$
1980	-0.6644	-0.0540	1.2076	-0.5588	0.1605	-0.1367
1981	-1.3993	* -2.2177	1.4633	1.0186	0.8693	-0.8808
1982	-0.6259	0.9271	1.0291	-1.4873	0.5564	1.6725
1983	-0.4081	1.2246	0.0295	0.3021	-0.4320	-0.3230
1984	-1.2078	*-2.1626	0.2215	0.8073	-0.3486	-0.6200
1985	-0.4017	-0.1525	-0.2346	-0.2144	-0.8872	-0.6044
1986	0.7570	0.4308	-0.0862	0.0172	1.6723	-1.0295
1987	0.8162	1.1426	-0.1445	-0.3784	0.4188	0.3734
1988	1.1662	0.2481	-0.2147	0.1090	0.9116	1.9449
1989	1.7603	0.6470	0.5478	* 2.2447	-0.1546	1.2453
$\sum_{1980}^{1989} S_\nu^2$	10.2767	13.9522	5.1386	9.5452	6.0138	10.9794

Ve druhé tabulce je totéž, pouze užíváme k výpočtu normální (Van der Waerdenovy) skóry.

Rok	$\nu = 0$	$\nu = 1$	$\nu = 2$	$\nu = 3$	$\nu = 4$	$\nu = 5$
1980	-0.6644	0.7973	0.6691	-0.3709	1.4861	1.2511
1981	-1.3993	0.0798	-1.4124	-0.9717	1.5990	0.7866
1982	-0.6259	-1.8359	0.4108	-1.5023	-0.4050	-1.6044
1983	-0.4081	-1.7282	-0.0530	-1.0644	-1.8369	-0.1864
1984	-1.2078	-1.4045	*-2.5651	0.6198	* -2.1322	-1.5499
1985	-0.4017	-1.1035	-0.5125	-0.7047	-0.6322	-0.5008
1986	0.7570	-0.2522	-0.5351	0.2810	-0.5173	1.5558
1987	0.8162	-1.6480	-1.3156	1.8226	1.9002	0.0411
1988	1.1662	-1.9433	-0.2473	-0.2081	1.7210	-0.4703
1989	1.7603	-0.4128	-1.4903	0.0165	-1.1744	-0.1655
$\sum_{1980}^{1989} S_\nu^2$	10.2767	16.917	13.7569	8.7973	*21.4699	10.1173

V poslední tabulce uvádíme Wilcoxonovy (Spearmanovy) autoregresní pořadové statistiky $T_{n;1;k}$, měřící korelaci mezi teplotou v čase t a úmrtností v čase $t+1, \ldots, t+k$, $k = 1, \ldots, 5$. , dále jsou zde uvedeny součty $\sum_{1980}^{1989} T_{n;1;k}$ pro pevné k a pro pohodlí též příslušné kvantily χ^2 rozdělení. Významné hodnoty jsou ohvězdičkovány.

Rok	$k = 1$	$k = 2$	$k = 3$	$k = 4$	$k = 5$
1980	0.0029	1.4613	1.7736	1.7994	1.8181
1981	*4.9185	*7.0598	* 8.0973	8.8530	9.6289
1982	0.8595	1.9188	4.1309	4.4406	7.2381
1983	1.4996	1.5005	1.5918	1.7785	1.8828
1984	*4.6768	4.7259	5.3777	5.4992	5.8837
1985	0.0232	0.0783	0.1243	0.9114	1.2768
1986	0.1856	0.1930	0.1933	2.9902	4.0501
1987	1.3055	1.3264	1.4697	1.6451	1.7845
1988	0.0615	0.1077	0.1196	0.9507	4.7336
1989	0.4186	0.7188	5.7575	5.7814	7.3323
$\sum_{1980}^{1989} T_{n;1;k}$	13.9522	19.0909	28.6362	34.6500	45.6294
$\chi^2_{k;0.95}$	3.84	5.99	7.81	9.48	11.07
$\chi^2_{k;0.99}$	6.63	9.21	11.34	13.28	15.09
$\chi^2_{10k;0.95}$	18.31	31.41	43.77	55.76	67.50
$\chi^2_{10k;0.99}$	23.21	37.57	50.89	63.69	76.15

Z uvedeného je vidět, že v jednotlivých případech vyšly významné hodnoty, ale obecný závěr, že by počet úmrtí u mužů a teplota spolu souvisela, zřejmě udělat nejde.

Reference

[1] Gutenbrunner, C. and Jurečková, J. (1992). "Regression rank scores and regression quantiles." *Ann. Statist.* 20, 305-330.

[2] Gutenbrunner, C., Jurečková, J., Koenker, R. and Portnoy, S. (1993). "Tests of linear hypotheses based on regression rank scores." *J. Nonpar. Statist.* 2, 307-331.

[3] Hájek, J. (1965). "Extension of the Kolmogorov-Smirnov test to regression alternatives." *Bernoulli-Bayes-Laplace* (L.LeCam and J.Neyman, eds.), 45-60. *Springer-Verlag, New York.*

[4] Hájek, J. and Šidák, Z. (1967). "Theory of Rank Tests." *Academia, Prague.*

[5] Hallin, M. and Jurečková, J. (1996). "Optimal tests for autoregressive models based on regression rank scores." Submitted.

[6] Hallin, M., Jurečková, J., Picek, J. and Zahaf, T. (1997). "Nonparametric tests of independence of two autoregressive time series based on autoregression rank scores." *J. Statist. Planning Inference* (to appear).

[7] Hallin, M., Zahaf, T., Jurečková, J., Kalvová, J. and Picek, J. (1997). "Nonparametric tests in AR models with applications to climatic data." *Environmetrics* 8, 651-660.

[8] Jurečková, J. (1991). "Tests of Kolmogorov-Smirnov type based on regression rank scores." *Transactions of the 11th Prague Conf. on Inform. Theory, Statist. Decision Functions and Random Processes* (J.Á.Víšek, ed.), 41-49. Academia, Prague.

[9] Jurečková, J. (1997). "Regression rank scores tests applied to heavy-tailed distributions." Submitted.

[10] Koenker, R. and Bassett, G. (1978). "Regression quantiles." *Econometrica* 46, 33-50.

[11] Koenker, R. and d'Orey, V. (1987). "Computing regression quantiles." *Appl. Statist.* 36, 383-393.

[12] Koenker, R. and d'Orey, V. (1994). "A remark on algorithm AS 229: Computing dual regression quantiles and regression rank scores." *Appl. Statist.* 42, 410-414.

[13] Koul, H.L. and Saleh, A.K.Md.E. (1995). "Autoregression quantiles and related rank score processes." *Ann. Statist.* 23, 670-689.

[14] Picek, J. (1997). "Regression rank scores tests for regression with random regressors." Submitted.

[15] Picek, J. (1996). "Statistické postupy založené na regresních pořadových skórech" *Kandidátská disertační práce, MFF UK Praha.*