Rozdelenia pravdepodobnosti odhadcov variančných parametrov vo FDSLRM

Robust 2018

Andrej Gajdoš
Martina Hančová

Prírodovedecká fakulta UPJŠ v Košiciach

25.1.2018

Čím sa zaoberáme

Kriging

- časové rady (**ČR**)
- modelovanie pomocou dostatočne všeobecnej a širokej triedy lineárnych regresných modelov (**LRM**)
- v našom výskume trieda **FDSLRM** (Štulajter 2002-2003)
- hlavná myšlienka – nájdenie najlepšej lineárnej nevychýlenej predikcie (**BLUP**)
- rôzne výskumné problémy: modelovanie, odhadovanie variančných parametrov, vplyv odhadov na prediktory, kvalita predikcií, ...

Čím sa zaoberáme

Definícia(FDSLRM)

Model časového radu $X(.)$ nazývame lineárny regresný model s konečným diskrétnym spektrom (FDSLRM), ak $X(.)$ má tvar

$$X(t) = \sum_{i=1}^{k} \beta_i f_i(t) + \sum_{j=1}^{l} Y_j v_j(t) + w(t);\ \ t = 1, 2, \ldots; k, l \in \mathbb{N}_0,$$

kde $\boldsymbol{\beta} = (\beta_1, \beta_2, \ldots, \beta_k)'$ z $\mathbb{E}^k$ je vektor regresných parametrov,
$\boldsymbol{Y} = (Y_1, Y_2, \ldots, Y_l)'$ je náhodný vektor so strednou hodnotou $E\{\boldsymbol{Y}\} = \mathbf{0}$ a kovariančnou maticou $Cov\{\boldsymbol{Y}\} = diag(\sigma_j^2)$ z $\mathbb{E}^{l\times l}$ so $\sigma_j^2 \geq 0; j = 1, 2, \ldots, l$,
$f_i(.);\ i = 1, 2, \ldots, k$ a $v_j(.);\ j = 1, 2, \ldots, l$ sú reálne funkcie definované na $\mathbb{E}^1$,
$w(.)$ je biely šum nekorelovaný s $\boldsymbol{Y}$ a s disperziou $D\{w(t)\} = \sigma^2 > 0$.

Definícia a vlastnosti LCNE

- zavedenie všeobecnejšej triedy odhadcov – **lineárna kombinácia NE** (LCNE) ▸ definícia NE

Definícia (Lineárna kombinácia NE)

Nech platia predpoklady definície NE. Lineárnou kombináciou NE, v skratke LCNE, budeme rozumieť výraz v tvare

$$\overset{*}{\sigma}_j^2(\boldsymbol{X}) = a\breve{\sigma}_j^2(\boldsymbol{X}) + b\breve{\sigma}^2(\boldsymbol{X}),\ a, b \in \mathbb{E}^1$$

kde $\breve{\sigma}_j^2(\boldsymbol{X})$, $\breve{\sigma}^2(\boldsymbol{X})$ sú príslušné NE.

- doterajší MNŠ odhadcovia vo FDSLRM – špeciálne prípady LCNE ▸ definícia DOOLSE, MDOOLSE

Definícia a vlastnosti LCNE

Veta (vlastnosti LCNE)

Majme lineárnu kombináciu NE v tvare $\overset{*}{\sigma}{}^2_j(\boldsymbol{X}) = a\breve{\sigma}^2_j(\boldsymbol{X}) + b\breve{\sigma}^2(\boldsymbol{X})$, $a, b \in \mathbb{E}^1$. *Potom títo odhadcovia sú invariatné kvadratické formy s nasledujúcimi štatistickými vlastnosťami:*

(i) $E_{\nu}\{\overset{*}{\sigma}{}^2_j(\boldsymbol{X})\} = \left[b + a\mathrm{W}_{jj}^{-1}\right]\sigma^2 + a\sigma_j^2$

(ii) $D_{\nu}\{\overset{*}{\sigma}{}^2_j(\boldsymbol{X})\} = 2a^2\left[\sigma_j^2 + \sigma^2\mathrm{W}_{jj}^{-1}\right]^2 + \frac{2b^2}{n-k-l}\sigma^4$

(iii) $Cov_{\nu}\{\overset{*}{\sigma}{}^2_i(\boldsymbol{X}), \overset{*}{\sigma}{}^2_j(\boldsymbol{X})\} = 2ac\left[\sigma^2\mathrm{W}_{ij}^{-1}\right]^2 + \frac{2bd}{n-k-l}\sigma^4$;
$\overset{*}{\sigma}{}^2_i(\boldsymbol{X}) = c\breve{\sigma}^2_i(\boldsymbol{X}) + d\breve{\sigma}^2(\boldsymbol{X})$, $i \neq j$

(iv) $MSE_{\nu}\{\overset{*}{\sigma}{}^2_j(\boldsymbol{X})\} = a^2\mathrm{W}_{jj}^{-2}\sigma^4 + 2a^2\left[\sigma_j^2 + \sigma^2\mathrm{W}_{jj}^{-1}\right]^2 + \frac{2b^2}{n-k-l}\sigma^4$

- dôkaz s využitím známych vlastností NE (Hančová, 2008) a operátorov $E\{.\}$, $D\{.\}$, $Cov\{.\}$

Definícia a vlastnosti LCNE

Veta (pravdepodobnostné rozdelenie LCNE)

Nech $\boldsymbol{X}$ je pozorovanie gaussovského FDSLRM. Potom LCNE s parametrami a, $-b$ $(a, b > 0)$ má rozdelenie $GDD(\alpha_1, \beta_1, \alpha_2, \beta_2)$ s nasledovnými parametrami

$$\alpha_1 = \frac{1}{2},\ \beta_1 = \frac{1}{2a(\sigma^2 \mathrm{W}_{jj}^{-1} + \sigma_j^2)},\ \alpha_2 = \frac{n-k-l}{2},\ \beta_2 = \frac{n-k-l}{2b\sigma^2}.$$

▸ GDD

- dôkaz s využitím lemy a projekčnej teórie v maticových (Hilbertových) priestoroch

Lema

Nech $\boldsymbol{X} \sim N(\mathrm{F}\boldsymbol{\beta}, \Sigma)$ a spĺňa lineárny regresný model v tvare $\boldsymbol{X} = \mathrm{F}\boldsymbol{\beta} + \varepsilon$, kde $E\{\varepsilon\} = 0$, $Cov\{\varepsilon\} = \Sigma$ a nech $\boldsymbol{X}'\mathrm{A}\boldsymbol{X}$ je invariatná kvadratická forma $(\mathrm{AF} = 0)$. Potom $\boldsymbol{X}'\mathrm{A}\boldsymbol{X} \sim \Gamma\left(\frac{r(\mathrm{A})}{2}, \frac{1}{2c}\right)$ práve vtedy, keď $\mathrm{A}\Sigma\mathrm{A}\Sigma = c\mathrm{A}\Sigma$.

- rozdelenia MNŠ odhadcov (NE, MNE, DOOLSE, MDOOLSE) ako dôsledky vety ▸ ďalšie rozdelenia

- ortogonálny FDSLRM: $\mathrm{F}'\mathrm{V} = 0$, $\mathrm{V}'\mathrm{V}$ je diagonálna matica

Veta

V ortogonálnom FDSLRM platia pre odhadcov variančných parametrov tieto vzťahy:

$$\text{DOOLSE} = \text{MINQUE(I)}, \quad \text{MDOOLSE} = \text{MINQUE(U,I)}.$$

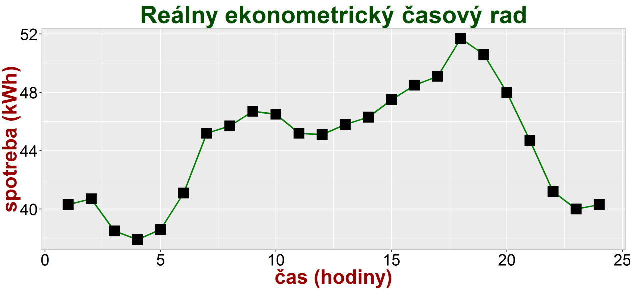
- dôkaz s využitím maticovej algebry, projekčnej teórie v Hilbertových priestoroch
- rozdelenia pravdepodobnosti pre MINQUE(I) a MINQUE(U,I) vo FDSLRM ako dôsledky našej vety

Numerické prístupy pre GDD

- tri spôsoby numerického výpočtu distribučnej funkcie GDD:
  1. numerická integrácia integrálu distribučnej funkcie
  2. numerická aproximácia pomocou váženej sumy χ^2 rozdelení
  3. Monte Carlo simulácie (Gajdoš, Hančová, Hanč 2017)
- aplikácia odvodenej teórie a navrhnutých numerických prístupov (problém výskytu záporných odhadov variancií)
- numerická/simulačná štúdia založená na prevzatých dátach (Štulajter, Witkovský, 2004)
- pravdepodobnosti, kvantily, intervaly spoľahlivosti

Numerické prístupy pre GDD

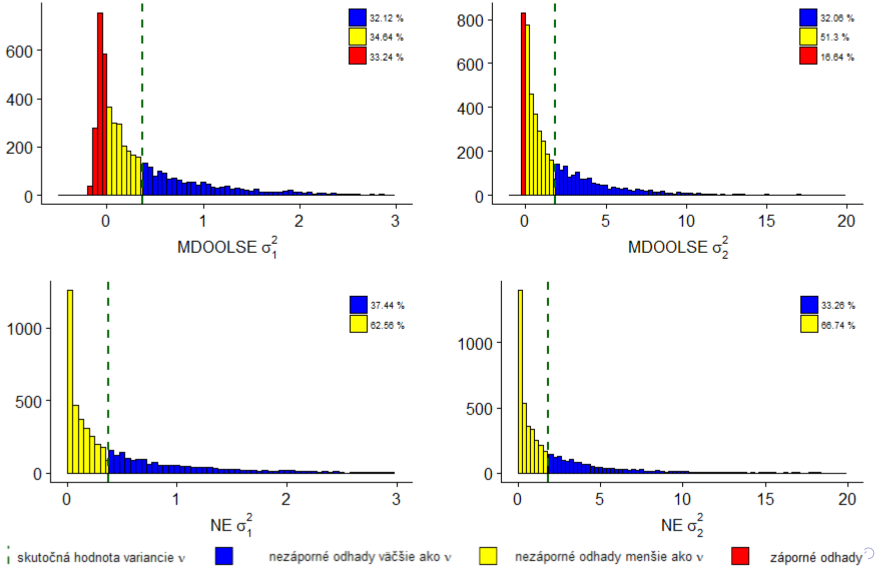
- 24 hodinové pozorovanie $\boldsymbol{X}$ spotreby el. energie v obchodnom dome (Štulajter & Witkovský, 2004):



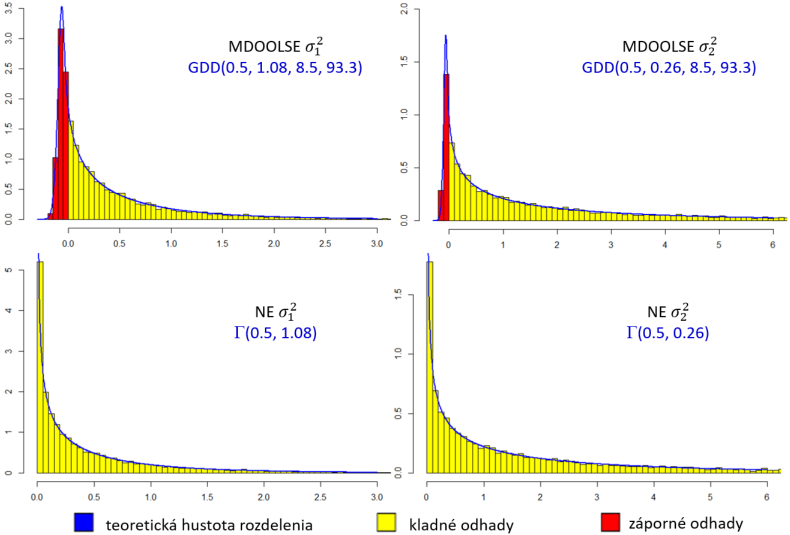
$$X_m = \alpha + \sum_{i=1}^{m} (\beta_i \cos \lambda_i t + \gamma_i \sin \lambda_i t) + \sum_{j=1}^{3-m} (Y_j \cos \lambda_j t + Z_j \sin \lambda_j t) + \\ + w(t);\ \lambda_i, \lambda_j \in \mathbb{E}^1,\ m \in \{0, 1, 2, 3\}$$

- štyri FDSLRM dizajny na základe spektrálnej analýzy ČR (R Core Team, 2018)

Numerické prístupy pre GDD



Numerické prístupy pre GDD



Numerické prístupy pre GDD

model	$m=0$		$m=1$		$m=2$	
odhad	DOOLSE	MDOOLSE	DOOLSE	MDOOLSE	DOOLSE	MDOOLSE
σ_1^2	0.073	0.075	0.124	0.135	0.299	0.337
σ_2^2	0.065	0.067	0.159	0.173	0.148	0.168
σ_3^2	0.131	0.135	0.313	0.337	$\times$	$\times$
σ_4^2	0.168	0.173	0.155	0.168	$\times$	$\times$
σ_5^2	0.328	0.337	$\times$	$\times$	$\times$	$\times$
σ_6^2	0.164	0.168	$\times$	$\times$	$\times$	$\times$

Význam a prínos výsledkov

- zavedenie všeobecnejšej triedy odhadcov variančných parametrov vo FDSLRM
- odvodenie momentových charakteristík a rozdelení pravdepodobnosti pre LCNE
- možnosť všeobecnejšie skúmať odhadcov i empirické prediktory vo FDSLRM
- navrhnuté a implementované numerické prístupy pre GDD
- exaktná kvantifikácia problému výskytu záporných odhadov vo FDSLRM

Literatúra

Gajdoš, A., Hančová, M., Hanč, J. (2017). *Kriging Methodology and Its Development in Forecasting Econometric Time Series*. Statistica: Statistics and Economy Journal, 97(1), 59–73.

Hančová, M. (2008). *Natural estimation of variances in a general finite discrete spectrum linear regression model*. Metrika, 67(3), 265-276.

Klar, B. (2015). *A Note on Gamma Difference Distributions*. Journal of Statistical Computation and Simulation, 85(18), 3708–3715.

R Core Team (2018). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/.

Štulajter, F. (2002). *Predictions in Time Series Using Regression Models* (Springer).

Štulajter, F. (2003). *The MSE of the BLUP in a Finite Discrete Spectrum LRM*. Tatra Mountains Mathematical Publications, 26(1), 125–131.

Štulajter, F., Witkovský, V. (2004). *Estimation of variances in orthogonal finite discrete spectrum linear regression models*. Metrika, 60(2), 105–118.

Ďaka za pozornosť

Definícia (NE), Hančová (2008)

Uvažujme pre pozorovanie $\boldsymbol{X}$ ČR $X(.)$ CLRM. Odhadca $\breve{\nu}(\boldsymbol{X}) = (\breve{\sigma}^2(\boldsymbol{X}), \breve{\sigma}_1^2(\boldsymbol{X}), \dots, \breve{\sigma}_l^2(\boldsymbol{X}))$ sa nazýva prirodzený odhadca, ak:

$$\breve{\sigma}^2(\boldsymbol{X}) = \frac{1}{n-k-l}[X - \mathrm{F}\hat{\beta}(\boldsymbol{X}) - \mathrm{V}\hat{y}(\boldsymbol{X})]'[X - \mathrm{F}\hat{\beta}(\boldsymbol{X}) - \mathrm{V}\hat{y}(\boldsymbol{X})],$$

$$\breve{\sigma}_j^2(\boldsymbol{X}) = \hat{y}_j(\boldsymbol{X})^2, \; j = 1, 2, \dots, l,$$

kde $(\hat{\beta}(\boldsymbol{X}), \hat{y}(\boldsymbol{X}))' = (\hat{\beta}_1(\boldsymbol{X}), \dots, \hat{\beta}_k(\boldsymbol{X}), \hat{y}_1(\boldsymbol{X}), \dots, \hat{y}_l(\boldsymbol{X}))'$ je odhad regresného parametra $(\boldsymbol{\beta}, \boldsymbol{y})' \in \mathbb{E}^{k+l}$ metódou najmenších štvorcov.

spät'

Definícia (DOOLSE, MDOOLSE), Štulajter, Witkovský (2004)

Nech je daný lineárny regresný model

$$\boldsymbol{X} = \mathrm{F}\boldsymbol{\beta} + \boldsymbol{\varepsilon}; \mathrm{F} \in \mathbb{E}^{n\times k}, \boldsymbol{\beta} \in \mathbb{E}^{k}, E\{\boldsymbol{\varepsilon}\} = 0, Cov\{\boldsymbol{X}\} = E\{\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{'}\} = \Sigma, \Sigma \geq 0,$$

ktorého kovariančná matica $\Sigma = \Sigma_{\boldsymbol{\nu}}$; $\boldsymbol{\nu} \in \Upsilon \subset \mathbb{E}^{l}$, je parametrizovaná pomocou parametra $\boldsymbol{\nu}$. Pre ľubovoľnú realizáciu $\boldsymbol{x}$ vektora $\boldsymbol{X}$ sú definované odhady DOOLSE a MDOOLSE parametra $\boldsymbol{\nu}$ nasledovne:

$$\text{DOOLSE} \quad \widetilde{\nu}(\boldsymbol{x}) = \arg\min_{\boldsymbol{\nu}\in\Upsilon} \|\mathrm{S}(\boldsymbol{x}) - \Sigma_{\boldsymbol{\nu}}\|^2,$$

$$\text{MDOOLSE} \quad \dot{\nu}(\boldsymbol{x}) = \arg\min_{\boldsymbol{\nu}\in\Upsilon} \|\mathrm{S}(\boldsymbol{x}) - \mathrm{M_F}\Sigma_{\boldsymbol{\nu}}\mathrm{M_F}\|^2,$$

kde matica $\mathrm{S}(\boldsymbol{x})$ je daná $\mathrm{S}(\boldsymbol{x}) = (\boldsymbol{x} - \mathrm{F}\hat{\boldsymbol{\beta}})(\boldsymbol{x} - \mathrm{F}\hat{\boldsymbol{\beta}})^{'}$; $\mathrm{M_F} = \mathrm{I} - \mathrm{F}(\mathrm{F}^{'}\mathrm{F})^{-1}\mathrm{F}^{'}$; $\boldsymbol{x}$ je realizácia $\boldsymbol{X}$ a $\hat{\boldsymbol{\beta}}$ je odhad parametra $\boldsymbol{\beta}$ metódou najmenších štvorcov (OLSE)

$$\hat{\boldsymbol{\beta}}(\boldsymbol{x}) = \arg\min_{\boldsymbol{\beta}\in\mathbb{E}^{k}} \|\boldsymbol{X} - \mathrm{F}\boldsymbol{\beta}\|^2.$$

späť

Gama rozdielové rozdelenie – $GDD(\alpha_1, \beta_1, \alpha_2, \beta_2)$, Klar (2015)

- hustota rozdelenia

$$f(z) = \begin{cases} ce^{\beta_2 z} \int_z^\infty x^{\alpha_1 - 1}(x-z)^{\alpha_2 - 1} e^{-(\beta_1+\beta_2)x} dx, & z > 0 \\ ce^{-\beta_1 z} \int_{-z}^\infty x^{\alpha_2 - 1}(x-z)^{\alpha_1 - 1} e^{-(\beta_1+\beta_2)x} dx, & z < 0 \end{cases}; \ c = \frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}}{\Gamma(\alpha_1)\Gamma(\alpha_2)}$$

$$f(z) = \begin{cases} \frac{\tilde{c}}{\Gamma(\alpha_1)} z^{\frac{\alpha_1+\alpha_2}{2}-1} e^{\frac{\beta_2-\beta_1}{2}z} W_{\frac{\alpha_1-\alpha_2}{2}, \frac{1-\alpha_1-\alpha_2}{2}}((\beta_1+\beta_2)z), & z > 0 \\ \frac{\tilde{c}}{\Gamma(\alpha_2)} (-z)^{\frac{\alpha_1+\alpha_2}{2}-1} e^{\frac{\beta_1-\beta_2}{2}(-z)} W_{\frac{\alpha_2-\alpha_1}{2}, \frac{1-\alpha_1-\alpha_2}{2}}((\beta_1+\beta_2)(-z)), & z < 0 \end{cases}$$

$$\tilde{c} = \frac{\beta_1^{\alpha_1}\beta_2^{\alpha_2}}{(\beta_1+\beta_2)^{\frac{\alpha_1+\alpha_2}{2}}} \qquad W_{k,m} \text{ – Whittakerova funkcia}$$

- momentová vytvárajúca a distribučná funkcia

$$M_X(t) = \left(1 - \frac{t}{\beta_1}\right)^{-\alpha_1} \left(1 + \frac{t}{\beta_2}\right)^{-\alpha_2}, \quad |t| < \min\{\beta_1, \beta_2\}$$

$$F(t) = d \int_{\max\{0,-t\}}^\infty x^{\alpha_2-1} e^{-\beta_2 x} \gamma(\alpha_1, \beta_1(x+t)) dx, \quad t \in \mathbb{E}$$

- stredná hodnota a disperzia

$$\mu = \frac{\alpha_1}{\beta_1} - \frac{\alpha_2}{\beta_2}, \quad \sigma^2 = \frac{\alpha_1}{\beta_1^2} + \frac{\alpha_2}{\beta_2^2}$$

spät'

NE

$$\breve{\sigma}^2(\boldsymbol{X}) \sim \Gamma\left(\frac{n-k-l}{2}, \frac{n-k-l}{2\sigma^2}\right)$$

$$\breve{\sigma}_j^2(\boldsymbol{X}) \sim \Gamma\left(\frac{1}{2}, \frac{1}{2(\sigma^2 \mathrm{W}_{jj}^{-1} + \sigma_j^2)}\right)$$

MNE

$$\breve{\sigma}^2(\boldsymbol{X}) \sim \Gamma\left(\frac{n-k-l}{2}, \frac{n-k-l}{2\sigma^2}\right)$$

$$\breve{\sigma}_j^2(X) \sim GDD(\alpha_1, \beta_1, \alpha_2, \beta_2)$$

s parametrami $\alpha_1 = \frac{1}{2}$, $\beta_1 = \frac{1}{2(\sigma^2 \mathrm{W}_{jj}^{-1} + \sigma_j^2)}$, $\alpha_2 = \frac{n-k-l}{2}$, $\beta_2 = \frac{n-k-l}{2\sigma^2 \mathrm{W}_{jj}^{-1}}$

DOOLSE

$$\tilde{\sigma}^2(\boldsymbol{X}) \sim \Gamma\left(\frac{n-k-l}{2}, \frac{n-l}{2\sigma^2}\right)$$

$$\tilde{\sigma}_j^2(\boldsymbol{X}) \sim GDD(\alpha_1, \beta_1, \alpha_2, \beta_2)$$

kde $\alpha_1 = \frac{1}{2}$, $\beta_1 = \frac{1}{2(\sigma^2 / \left\|\boldsymbol{v_j}\right\|^2 + \sigma_j^2)}$, $\alpha_2 = \frac{n-k-l}{2}$, $\beta_2 = \frac{(n-l)||\boldsymbol{v_j}||^2}{2\sigma^2}$

MDOOLSE

$$\dot{\sigma}^2(\boldsymbol{X}) \sim \Gamma\left(\frac{n-k-l}{2}, \frac{n-k-l}{2\sigma^2}\right)$$

$$\dot{\sigma}_j^2(\boldsymbol{X}) =\sim GDD(\alpha_1, \beta_1, \alpha_2, \beta_2)$$

kde $\alpha_1 = \frac{1}{2}$, $\beta_1 = \frac{1}{2(\sigma^2 / \left\|\boldsymbol{v_j}\right\|^2 + \sigma_j^2)}$, $\alpha_2 = \frac{n-k-l}{2}$, $\beta_2 = \frac{(n-k-l)\left\|\boldsymbol{v_j}\right\|^2}{2\sigma^2}$