Explicitný tvar momentov aproximácie MSE pre EBLUP v lineárnych regresných modeloch časových radov

Prezentácia Robust 2016

Andrej Gajdoš
Martina Hančová

Prírodovedecká fakulta UPJŠ v Košiciach

15.9.2016

Čím sa zaoberáme

- skúmanie časových radov (ČR)
- predikcie ČR lineárnymi regresnými modelmi (FDSLRM)

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 $\beta = (\beta_1, \beta_2, \ldots, \beta_k)'$ z $\mathbb{E}^k$ je vektor regresných parametrov,
$Y = (Y_1, Y_2, \ldots, Y_l)'$ je náhodný vektor so strednou hodnotou $E\{Y\} = \mathbf{0}$ a kovariančnou maticou $Cov\{Y\} = diag(\sigma_j^2)$ 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 Y a s disperziou $D\{w(t)\} = \sigma^2 > 0$.

Čím sa zaoberáme

Definícia (Najlepší lineárny nevychýlený prediktor - BLUP)

Nech $\boldsymbol{X} = (X(1), X(2), \dots, X(n))'$ je konečné pozorovanie časového radu $X(.)$ popísané lineárnym regresným modelom s neznámym parametrom $\boldsymbol{\beta}$:

$$\boldsymbol{X} = F\boldsymbol{\beta} + \boldsymbol{\varepsilon}, \quad E(\boldsymbol{\varepsilon}) = 0, \quad \mathrm{Cov}(\boldsymbol{X}) = \Sigma_n > 0,$$

pričom $\boldsymbol{\beta} \in \mathbb{E}^k$, $F \in \mathbb{E}^{n\mathrm{x}k}$ je tzv. matica plánu s plnou hodnosťou. Prediktor $X^*(n+d)$ nazývame **najlepším lineárnym nevychýleným prediktorom** zložky $X(n+d)$, ak je:

(a) lineárny vzhľadom na $\boldsymbol{X}$, t.j. $X^*(n+d) = \boldsymbol{a}'\boldsymbol{X} + a_0; \boldsymbol{a} \in \mathbb{E}^n, a_0 \in \mathbb{E}$,

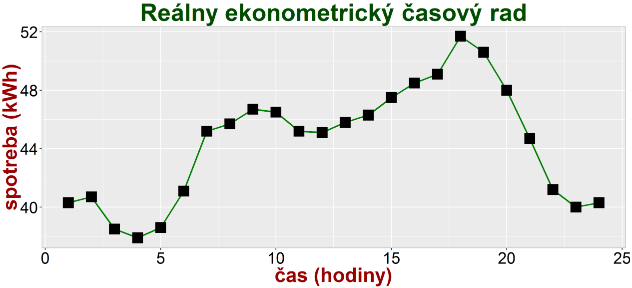
(b) nevychýlený pre všetky $\boldsymbol{\beta}$, t.j. $E_{\boldsymbol{\beta}}(X^*(n+d)) = E_{\boldsymbol{\beta}}(X(n+d))$

(c) najlepší, t.j. minimalizuje MSE v triede všetkých lineárnych nevychýlených prediktorov $\tilde{U}(n+d)$

$$E(X^*(n+d) - X(n+d))^2 \leq E(\tilde{U}(n+d) - X(n+d))^2.$$

Reálne dáta

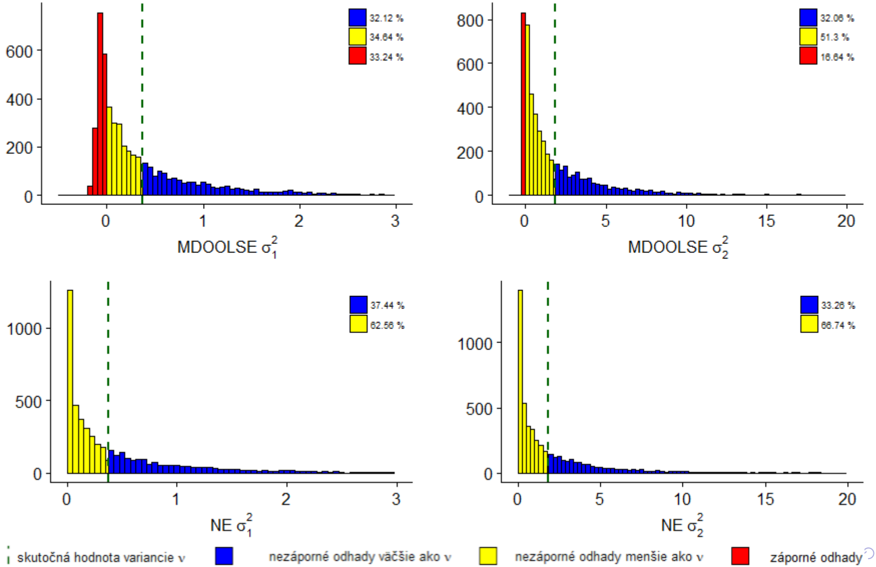
- 24 hodinové pozorovanie **X** spotreby el. energie v obchodnom dome (Štulajter & Witkovský, 2004):



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

- štyri FDSLRM dizajny a ilustrácia výsledkov pre dizajn $m = 2$ (R Core Team, 2015)

Problém - záporné odhady



Korekcia MSE pre EBLUP

- BLUP $\rightarrow X^*_{\boldsymbol{\nu}}(n+d) = \boldsymbol{f}'\boldsymbol{\beta}^*_{\boldsymbol{\nu}}(\mathbf{X}) + \mathbf{r}'_{\boldsymbol{\nu}}\Sigma^{-1}_{\boldsymbol{\nu}}(\mathbf{X} - F\boldsymbol{\beta}^*_{\boldsymbol{\nu}}(\mathbf{X}))$, pričom $\boldsymbol{\beta}^*_{\boldsymbol{\nu}}(\mathbf{X}) = (F'\Sigma^{-1}F)^{-1}F'\Sigma^{-1}_{\boldsymbol{\nu}}\mathbf{X}$

- $$\mathrm{MSE}\{X^*_{\boldsymbol{\nu}}(n+d)\} = \mathrm{D}_{\boldsymbol{\nu}}\{X(n+d)\} - \boldsymbol{r}'_{\boldsymbol{\nu}}\Sigma^{-1}_{\boldsymbol{\nu}}\boldsymbol{r}_{\boldsymbol{\nu}} + \\ + \parallel \boldsymbol{f} - F'\Sigma^{-1}_{\boldsymbol{\nu}}\boldsymbol{r}_{\boldsymbol{\nu}} \parallel_{(F'\Sigma^{-1}_{\boldsymbol{\nu}}F)^{-1}}$$

- EBLUP $\rightarrow X^*_{\tilde{\boldsymbol{\nu}}}(n+d) = \boldsymbol{f}'\boldsymbol{\beta}^*_{\tilde{\boldsymbol{\nu}}}(\mathbf{X}) + \mathbf{r}'_{\tilde{\boldsymbol{\nu}}}\Sigma^{-1}_{\tilde{\boldsymbol{\nu}}}(\mathbf{X} - F\boldsymbol{\beta}^*_{\tilde{\boldsymbol{\nu}}}(\mathbf{X}))$

- $$\boxed{\mathrm{E}_{\boldsymbol{\nu}_0}[X^*_{\tilde{\boldsymbol{\nu}}}(n+d) - X^*_{\boldsymbol{\nu}_0}(n+d)]^2 \approx \mathrm{E}_{\boldsymbol{\nu}_0}\left[\frac{\partial X^*_{\boldsymbol{\nu}}(n+d)}{\partial \boldsymbol{\nu}'}\Big|_{\boldsymbol{\nu}=\boldsymbol{\nu}_0}(\tilde{\boldsymbol{\nu}} - \boldsymbol{\nu}_0)\right]^2}$$

- Potreba explicitných vyjadrení momentov až do 6. rádu $\mathrm{E}\{\mathbf{X}'A\mathbf{X}\}$ resp. $\mathrm{E}\{\mathbf{X}'B\mathbf{X}\}$, $\mathrm{E}\{\mathbf{X}\mathbf{X}'\}$, $\mathrm{E}\{\mathbf{X}\mathbf{X}'A\mathbf{X}\mathbf{X}'\}$ resp. $\mathrm{E}\{\mathbf{X}\mathbf{X}'B\mathbf{X}\mathbf{X}'\}$, $\mathrm{E}\{\mathbf{X}\mathbf{X}'A\mathbf{X}\mathbf{X}'B\mathbf{X}\mathbf{X}'\}$ a ďalšie...

Explicitný tvar momentov

Veta

Nech je dané konečné pozorovanie časového radu lineárnym regresným modelom: $X \sim N(F\beta, \Sigma)$, kde $F \in \mathbb{E}^{n\times k}$, $\beta \in \mathbb{E}^{k}$ a $\Sigma \in \mathbb{E}^{n\times n}$, $\Sigma > 0$.
Nech $Q_A \equiv X'AX$, $Q_B \equiv X'BX$, $A, B \in \mathbb{E}^{n\times n}$ sú invariantné kvadratické formy, t.j. $AF = BF = 0$, $\mathrm{E}(Q_A) = \mathrm{tr}(A\Sigma)$, $\mathrm{E}(Q_B) = \mathrm{tr}(B\Sigma)$ a $\mathrm{Cov}(Q_A, Q_B) = 2\mathrm{tr}(A\Sigma B\Sigma)$.
Potom pre centrované $Q_A^* \equiv Q_A - \nu_A$ a $Q_B^* \equiv Q_B - \nu_B$, $\nu_A, \nu_B \in \mathbb{E}^1$ platí:

1. $\mathrm{E}(Q_A^*) = \mathrm{E}(Q_A) - \nu_A$, $\mathrm{E}(XQ_A^*) = \mathrm{E}(Q_A^*)F\beta$,
   $\mathrm{E}(XQ_A^*X') = \mathrm{E}(Q_A^*)[\Sigma + F\beta(F\beta)'] + 2\Sigma A\Sigma$
2. $\mathrm{E}(Q_A^*Q_B^*) = \mathrm{Cov}(Q_A, Q_B) + \mathrm{E}(Q_A^*)\mathrm{E}(Q_B^*)$,
   $\mathrm{E}(XQ_A^*Q_B^*) = \mathrm{E}(Q_A^*Q_B^*)F\beta$,
   $$\mathrm{E}(XQ_A^*Q_B^*X') = 2\Sigma[\mathrm{E}(Q_A^*)B + \mathrm{E}(Q_B^*)A + 2A\Sigma B + 2B\Sigma A]\Sigma + \\ + \mathrm{E}(Q_A^*Q_B^*)[\Sigma + F\beta(F\beta)']$$

Explicitný tvar momentov

- dôkaz s využitím algebraických metód viacrozmernej štatistiky – vektorizácia, komutačné matice, kroneckerov súčin a vzťahy medzi nimi (Ghazal & Neudecker 2000, Kollo & Rosen 2005)

Význam a prínos výsledkov

- potreba explicitných tvarov momentov pre ďalšie teoretické štúdium aproximácie MSE a iných vlastností EBLUPov
- Doteraz teoreticky študované aproximácie MSE pre EBLUPy vo FDSLRM len pre nevychýlené invariantné odhady:
  - ortogonálny prípad FDSLRM – nevychýlené DOOLSE=REMLE (Štulajter, 2007)
  - všeobecný prípad FDSLRM – nevychýlené NE (Hančová, 2011)
- moderný a elegantný prístup k odvádzaniu výsledkov
- výsledné tvary výhodné pre počítačovú implementáciu v R (aplikácia na reálne dáta, počítačové simulácie, bootstrap)

Literatúra

Ghazal, A.G., Neudecker, H. (2000). *On second-order and fourth-order moments of jointly distributed random matrices: a survey*. Linear Algebra and its Applications 321, 61–93.

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

Hančová, M. (2011).*Empirical Predictors in Finite Discrete Spectrum Linear Regression Models*. PROBASTAT 2011.

Hančová, M., Hanč, J. & Gajdoš., A (2015). *A Simulation Study of Bootstrap Methods for Kriging in Time Series Forecasting*. PROBASTAT 2015.

Harville, D.A. (2008).*Accounting for the Estimation of Variances and Covariances in Prediction Under a General Linear Model: an Overview*. Tatra Mountains Mathematical Publications 39, pp. 1-15.

Kreiss, J.P. & Lahiri, S.N. (2012). *Bootstrap methods for time series*. Handbook of Statistics – Time Series Analysis: Methods and Applications. Elsevier.

Literatúra

Kollo, T., von Rosen., D. (2005). *Advanced Multivariate Statistics with Matrices*.

R Core Team (2016). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria.
URL https://www.R-project.org/.

Srivastava, V.K., Tiwari, R. (1975). *Evaluation of Expectations of Products of Stochastic Matrices*. Scand J Statist 3, 135-138.

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

Štulajter, F., Witkovský, V. (2004). *Estimation of variances in orthogonal finite discrete spectrum linear regression models*. Metrika, vol. 60, no. 2, pp. 105–118.

Štulajter, F. (2007) *Mean squared error of the empirical best linear unbiased predictor in an orthogonal finite discrete spectrum linear regression model*. Metrika 65, 2007, pp. 331-348.

Vďaka za pozornosť

- pre MSE EBLUPu platí (Štulajter 2002)

$$\mathrm{E}_\Sigma[\tilde{U} - X_{n+d}]^2 = \mathrm{E}_\Sigma[U^*_\Sigma - X_{n+d}]^2 + \mathrm{E}_\Sigma[U^*_\Sigma - \tilde{U}]^2.$$

- výpočet $\mathrm{E}_\Sigma[U^*_\Sigma - \tilde{U}]^2$ vo všeobecnosti je **otvorený problém**, v praxi aproximácia – Taylorov rozvoj:

$$\tilde{U} \approx U^*_\Sigma + \frac{\partial \tilde{U}(\mathbf{X}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}'}|_{\boldsymbol{\nu}=\nu_0}(\tilde{\boldsymbol{\nu}} - \nu_0)$$

odkiaľ

$$\tilde{U} - U^*_\Sigma \approx \frac{\partial \tilde{U}(\mathbf{X}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}'}|_{\boldsymbol{\nu}=\nu_0}(\tilde{\boldsymbol{\nu}} - \nu_0)$$

teda MSE $\tilde{U} - U^*_\Sigma$ sa dá **aproximovať** vzťahom

$$\mathrm{E}_{\nu_0}[\tilde{U} - U^*_{\nu_0}]^2 \approx \mathrm{E}_{\nu_0}\left[\frac{\partial \tilde{U}(\mathbf{X}, \boldsymbol{\nu})}{\partial \boldsymbol{\nu}'}|_{\boldsymbol{\nu}=\nu_0}(\tilde{\boldsymbol{\nu}} - \nu_0)\right]^2$$