



# Aproximácia zovšeobecneného konfidenčného intervalu pre $\sigma_1^2$

Barbora Arendacká

barendacka@gmail.com

Ústav merania, Slovenská akadémia vied, Dúbravská cesta 9, 841 04 Bratislava

**Model:** zmiešaný lineárny model s dvomi variančnými komponentmi →

$$Y = X\beta + Zu + \epsilon$$

$$Y : n \times 1, \quad X : n \times p, \quad Z : n \times q$$

$$u \sim N_q(0, \sigma_u^2 I_q), \quad \epsilon \sim N_n(0, \sigma^2 I_n), \text{ nezávislé}$$

$$Y \sim N_n(X\beta, \sigma_1^2 ZZ^T + \sigma^2 I_n)$$

variančné komponenty:  $\sigma_1^2, \sigma^2$

**Postačujúce štatistiky:** navzájom nezávislé

$$U_i = Y^T B F_i B^T Y \sim (\lambda_i \sigma_1^2 + \sigma^2) \chi_{\nu_i}^2$$

$i = 1, \dots, r$ , kde  $\lambda_1 > \dots > \lambda_r = 0$  sú vlastné čísla  $B^T Z Z^T B$ ,  $\nu_i$  ich násobnosti a pre každé  $i$ ,  $F_i$  je ortogonálny projektor na podpriestor generovaný vlastnými vektormi patriacimi k  $\lambda_i$

$$B^T Y \sim N_{n-rank(X)}(0, \sigma_1^2 B^T Z Z^T B + \sigma^2 I)$$

**Príklad:** vyvážený model jednoduchého triedenia s náhodným efektom:

$$Y_{i,j} = \mu + \alpha_i + \epsilon_{i,j}$$

$$i = 1, \dots, I > 1, \quad j = 1, \dots, J > 1, \quad \alpha_i \sim N(0, \sigma_1^2)$$

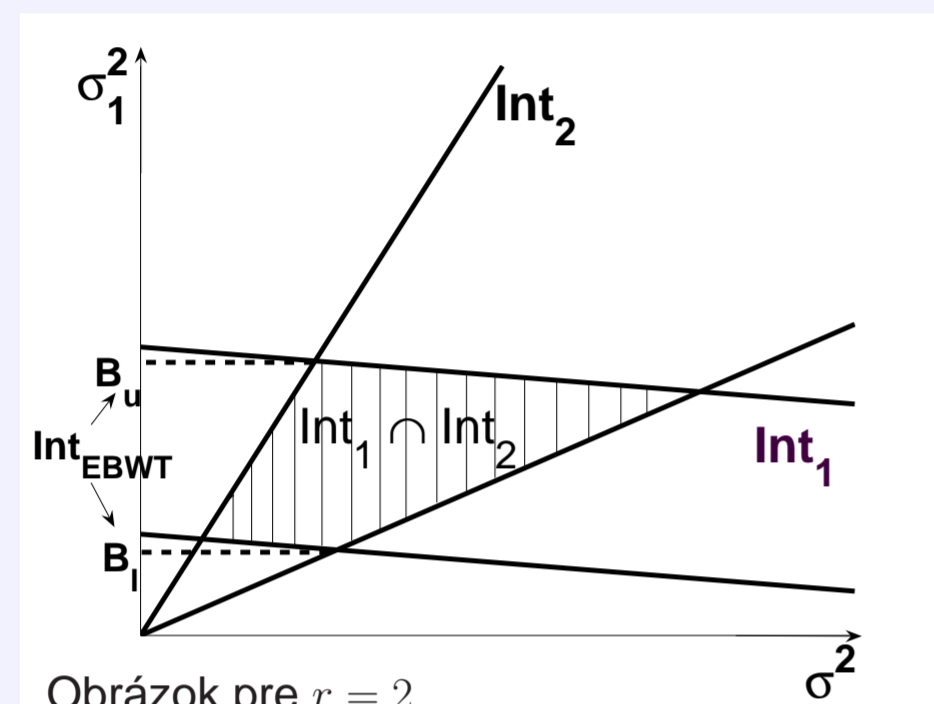
$$r = 2, \quad \lambda_1 = J, \quad \lambda_2 = 0, \quad \nu_1 = I - 1, \quad \nu_2 = I(J - 1),$$

$$U_1 = \sum_{i=1}^I J(\bar{Y}_i - \bar{Y}_{..})^2, \quad U_2 = \sum_{i=1}^I \sum_{j=1}^J (Y_{i,j} - \bar{Y}_i)^2$$

## Približný konfidenčný interval pre $\sigma_1^2$

„stredné sumy štvorcov“:  $S_i = U_i/\nu_i, s = \sum_{i=1}^{r-1} \nu_i, \alpha_u - \alpha_l = 1 - \alpha$

El-Bassiouniho-Williamsov-Tukeyho interval (viď [3], [7], [1])



Obrázok pre  $r = 2$

$B_l, B_u$  sú nezáporné riešenia alebo nuly, ak nezáporné riešenia neexistujú:

$$\sum_{i=1}^{r-1} \frac{\nu_i S_i}{\lambda_i B_l + S_r \frac{s F_{s, \nu_r; \alpha_u}}{\chi_{s; \alpha_u}^2}} = \chi_{s; \alpha_u}^2 \quad \sum_{i=1}^{r-1} \frac{\nu_i S_i}{\lambda_i B_u + S_r \frac{s F_{s, \nu_r; \alpha_l}}{\chi_{s; \alpha_l}^2}} = \chi_{s; \alpha_l}^2$$

**Int<sub>1</sub>** : presný interval pre  $\sigma_1^2$  pri danom  $\sigma^2$  založený na  $\sum_{i=1}^{r-1} \frac{\nu_i S_i}{\lambda_i \sigma_1^2 + \sigma^2} \sim \chi_s^2$

**Int<sub>2</sub>** : presný interval pre  $\sigma_1^2$  pri danom  $\sigma^2$  získaný z presného intervalu pre  $\sigma_1^2/\sigma^2$  vynásobením jeho hraníc hodnotou  $\sigma^2$

**Int<sub>EBWT</sub>** obsahuje **Int<sub>1</sub> ∩ Int<sub>2</sub>** bez ohľadu na  $\sigma^2$

## Zovšeobecnený konfidenčný interval pre $\sigma_1^2$

„stredné sumy štvorcov“:  $S_i = U_i/\nu_i, s = \sum_{i=1}^{r-1} \nu_i, \alpha_u - \alpha_l = 1 - \alpha$

Parkov-Burdickov zovšeobecnený pivot (viď [5], tiež [4], [6])

Pre dané pozorovania je zovšeobecnený interval tvorený kvantilmi  $[q_{\alpha_l}, q_{\alpha_u}]$  rozdelenia zovšeobecneného pivota **R**, ktorý je pre dané dáta definovaný ako nezáporné riešenie alebo nula, ak nezáporné riešenie neexistuje:

$$\sum_{i=1}^{r-1} \frac{\nu_i S_i}{\lambda_i \mathbf{R} + \nu_r S_r / Q_2} = Q_1,$$

kde  $s_1, \dots, s_r$  sú napozorované hodnoty  $S_1, \dots, S_r$ ,  $Q_1 \sim \chi_s^2, Q_2 \sim \chi_{\nu_r}^2$  sú navzájom nezávislé.

**Odvodenie:** pivotálne rovnice

$$\sum_{i=1}^{r-1} \frac{\nu_i S_i}{\lambda_i \sigma_1^2 + \sigma^2} = E_1 \sim Q_1$$

$$\nu_r S_r / \sigma^2 = E_2 \sim Q_2$$

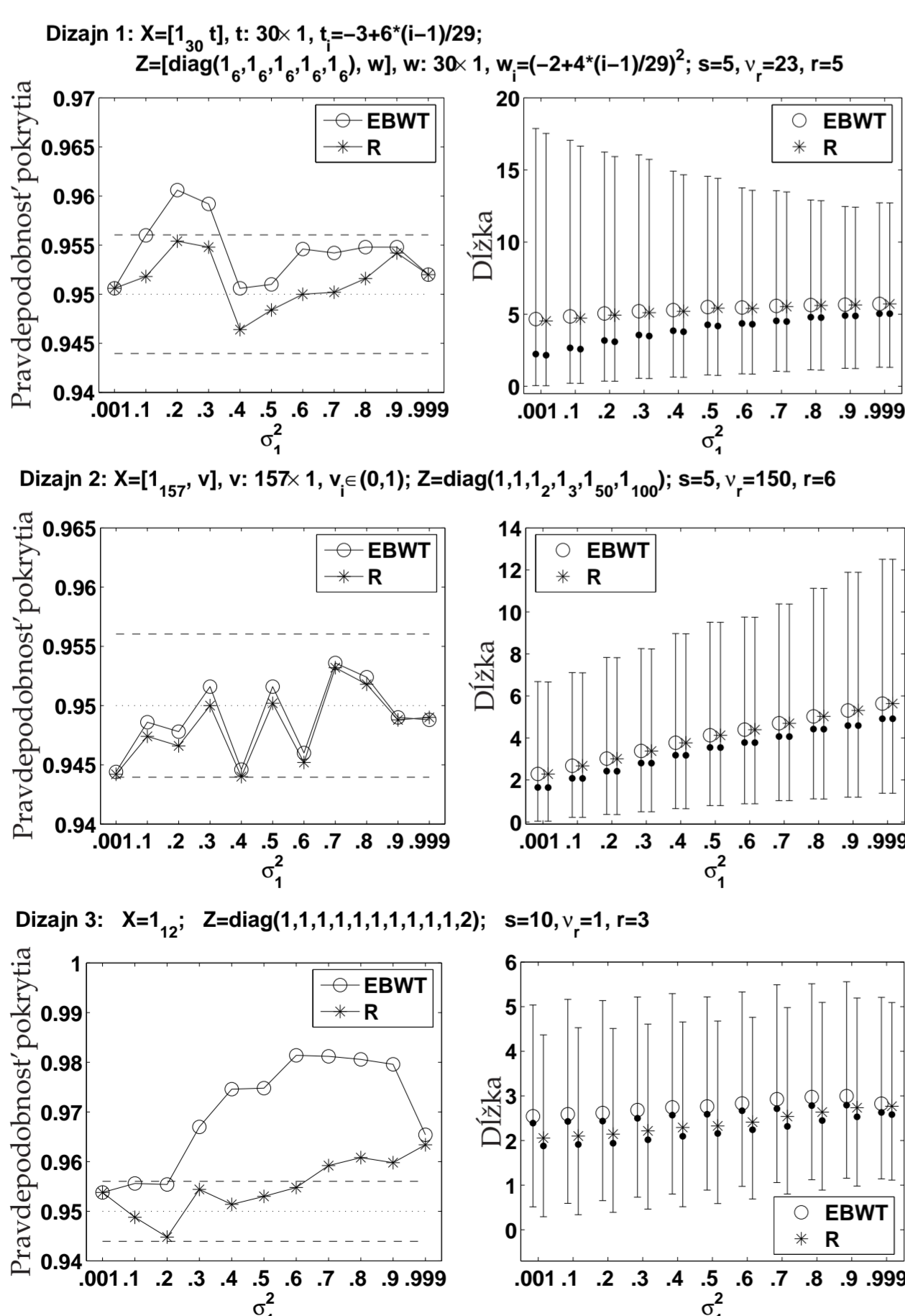
$S_1, \dots, S_r$  pozorujeme,  $E_1, E_2$  generujeme → získavané riešenia systému pre  $\sigma_1^2, \sigma^2$  (obmedzené na parametrický priestor) určujú rozdelenie ( $\sigma_1^2, \sigma^2$ ) pri daných dátach.

Interval pre  $\sigma_1^2$  potom prirodzene tvorí dolný a horný kvantil rozdelenia tohto parametra.

Kvantily  $q_{\alpha_l}, q_{\alpha_u}$  môžeme nájsť simulačne, či numerickým výpočtom. Oba spôsoby sú pomerne zdĺhavé. Pre praktické účely by však možno stačilo nájsť aproximáciu  $q_{\alpha_l}, q_{\alpha_u}$ . Tieto kvantily vieme presne vyjadriť v špeciálnych situáciách, pozri tiež [2], čo nás usmerňuje pri hľadaní približných foriem:

Ak	$\nu_r \rightarrow \infty, (s_1, \dots, s_r \text{ sa považujú za pevné}),$	$\frac{s_i}{s_r} \rightarrow \infty, s_i / \sum_{j=1}^{r-1} s_j \rightarrow a_i > 0, i = 1, \dots, r - 1$	$\sum_{i=1}^{r-1} \nu_i s_i / s_r \leq s F_{s, \nu_r; \alpha_u},$
tak	$q_{1-\alpha_u} \rightarrow t, \text{ kde } \sum_{i=1}^{r-1} \frac{\nu_i s_i}{\lambda_i t + s_r} = \chi_{s; \alpha_u}^2, t \geq 0$	$q_{1-\alpha_u} / \sum_{i=1}^{r-1} s_i \rightarrow \sum_{i=1}^{r-1} a_i \nu_i / \lambda_i / \chi_{s; \alpha_u}^2,$	$q_{1-\alpha_u} = 0,$
	$B_l \rightarrow t, \text{ kde } \sum_{i=1}^{r-1} \frac{\nu_i s_i}{\lambda_i t + s_r} = \chi_{s; \alpha_u}^2, t \geq 0$	$B_l / \sum_{i=1}^{r-1} s_i \rightarrow \sum_{i=1}^{r-1} a_i \nu_i / \lambda_i / \chi_{s; \alpha_u}^2,$	$B_l = 0$

⇒ pre  $\alpha_l = \alpha/2, \alpha_u = 1 - \alpha/2$ , môžeme položiť  $q_{\alpha/2} \approx B_l, q_{1-\alpha/2} \approx B_u$

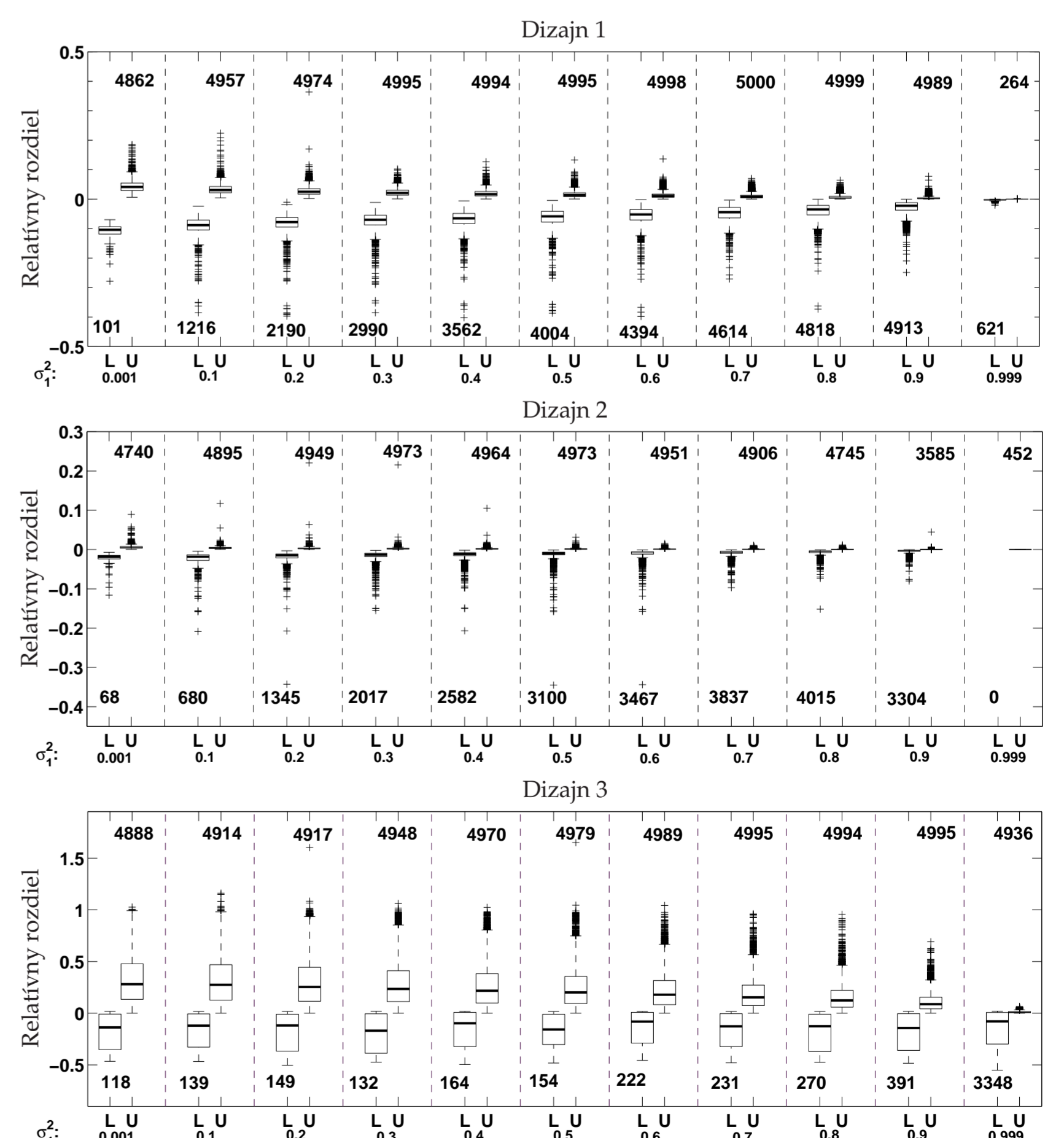


**Simulačné porovnanie.**  $1 - \alpha = 0.95$ , 5000 realizácií  $Y$  pre každú dvojicu skutočných hodnôt  $\sigma_1^2, \sigma^2$  ( $\sigma^2 = 1 - \sigma_1^2$ )

**Vľavo:**  $[q_{\alpha/2}, q_{1-\alpha/2}] \approx [B_l, B_u]$   
Efekt nahradenia jedného intervalu druhým. Pravdepodobnosti pokrytia a dĺžky zovšeobecneného intervalu a jeho aproximácie. Zobrazené sú priemery (o,\*), mediány (bodky) a 5. a 95. percentily dĺžok (spojené čiarou).

**Vpravo:**  $q_{\alpha/2} \approx B_l, q_{1-\alpha/2} \approx B_u$   
Presnosť aproximácie. Boxploty relatívnych rozdielov:  $(B_l - q_{\alpha/2})/q_{\alpha/2}, (L); (B_u - q_{1-\alpha/2})/q_{1-\alpha/2}, (U)$ ; pre dvojice dolných (horných) hraníc s absolútnym rozdielom viac ako  $0.5 \cdot 10^{-3}$ , čo je chyba, s ktorou boli  $q_{\alpha/2}, q_{1-\alpha/2}$  spočítané metódou bisekcie. Počet uvažovaných dvojíc udávajú čísla pod boxplotmi. (Pozn. v Dizajne 3 je veľa nulových hraníc, v prípade ktorých je aproximácia presná.)

Aproximácia je najhoršia pre dizajn, kde  $s > \nu_r$ , a podľa očakávania sa zlepšuje s rastúcim  $\nu_r$ .



**Podakovanie.** Práca bola podoporená grantmi VEGA 1/3016/06 a APVV RPEU-0008-06.

### Referencie.

[1] Arendacká, B. (2007) - A modification of the Hartung-Knapp interval on the variance component in two-variance component models, *Kybernetika*, 43, 471–480  
[2] Bross, I. (1950) - Fiducial intervals for variance components, *Biometrics*, 6, 136–144

[3] El-Bassiouni, M.Y. (1994) - Short confidence intervals for variance components, *Commun. in Stat. Theory and Methods*, 23, 1951–1933  
[4] Hannig, J. (2008) - On generalized fiducial inference, *Statistica Sinica*, to appear  
[5] Park, D.J., Burdick, R.K. (2003) - Performance of confidence in-

tervals in regression models with unbalanced one-fold nested error structures, *Commun. in Stat. Simul. and Comp.*, 32, 717–732  
[6] Weerahandi, S. (1995) - Exact Statistical Methods for Data Analysis, Springer-Verlag, New York  
[7] Williams, J.S. (1962) - A confidence interval for variance components, *Biometrika*, 49, 278–281