ŠTATISTICKÉ VLASTNOSTI MNŠ - ODHADOV VARIANČNÝCH KOMPONENTOV V ZMIEŠANOM LINEÁRNOM MODELI

Júlia Volaufová, Viktor Witkovský

Zmiešaný lineárny model je všeobecne uvádzaný v literatúre v nasledujúcom tvare

$$Y=X\beta + U_1\xi_1+\dots+U_{p-1}\xi_{p-1}+\varepsilon ,$$

kde Y je n-rozmerný vektor pozorovaní, β je k-rozmerný neznámy vektorový parameter; ξ_i $i=1,\dots,p-1$ sú n_i rozmerné náhodné vektory (náhodné efekty), ktoré spĺňajú predpoklad

$E(\xi_i) = 0$, $E(\xi_i\xi_i') = \sigma_i^2 I_{n_i}$ $i=1,\dots,p-1$

ε je n-rozmerný vektor chýb, pre ktorý platí rovnako $E(\varepsilon)=0$, $E(\varepsilon\varepsilon')=\sigma_p^2 I_p$. Matice $X, U_1,\dots,U_{p-1}$ sú známe nenáhodné matice a $\sigma_1^2,\dots,\sigma_p^2$ sú neznáme konštantné parametre. Vo všeobecnosti je to model s varianČno - kovariančnými komponentami v tvare

(1) $\quad (Y,\ X\beta,\ \sum_{i=1}^{p}\vartheta_i V_i)$, pričom

cieľ je odhadnúť lineárnu funkciu parametra $\vartheta = (\vartheta_1,\dots,\vartheta_p)'$:

$f(\vartheta) = f'\vartheta = \sum_{i=1}^{p} f_i\vartheta_i$.

Tento problém je historicky nie nový a na jeho riešenie existuje niekoľko rôznych prístupov. Najznámejší je klasický prístup - vytvoriť kvadratický odhad $Y'AY$, može spĺňať rôzne požiadavky, ako sú nevychýlenosť, invariantnosť, lokálne najlepší nevychýlený odhad (v zmysle minimálnej disperzie) alebo odhady typu MINQUE. Uvedené metódy je možné nájsť napr. v Rao (1971, 1972), Rao, Kleffe (1980), Kubáček (1985), Rao,Kleffe (1988) a v mnohých ďalších publikáciách.

Zaujímavý je prípad založený na lineárnom modeli v závislej premennej vec YY'- vektore vytvorenom zo stĺpcov symetrickej matice YY' ich usporiadaním pod seba. (Pozri napr. Verdooren (1979,1988)).

Model má tvar:

(2) $\quad E(\text{ vec } YY') =$

$$= (X\otimes X,\ \text{vec}V_1,\dots,\ \text{vec}V_p)\begin{bmatrix}\text{vec } \beta\beta' \\ \vartheta_1 \\ \vdots \\ \vartheta_p\end{bmatrix} = Z\xi$$

Operácia "$\otimes$" je Kroneckerov súčin matíc; matica Z je známa matica typu $n^2 \times (k^2+p)$. Parameter ξ je k^2+p rozmerný, kde vec $\beta\beta'$ považujeme za rušivý. Keďže odhadujeme lineárnu funkciu parametrov druhého rádu, je prirodzené požadovať, aby odhad bol invariantný vzhľadom na posun v strednej hodnote. Odhad $T(Y)$ je invariantný práve vtedy, keď $T(Y)=T(Y+X\beta)$ pre všetky $\beta \in R^k$. Maximálnym invariantom, t.j. štatistika založená na vektore Y je vektor MY, kde $M = I - P_X$ je projektor na ortogonálny komplement stĺpcového priestoru matice X, $P_X = X(X'X)^{-1}X'$. Vytvoríme transformovaný model, založený na vektore $M\otimes M$ vecYY' v tvare:

(3) $\quad E(\ M\otimes M \text{ vec}YY') =$

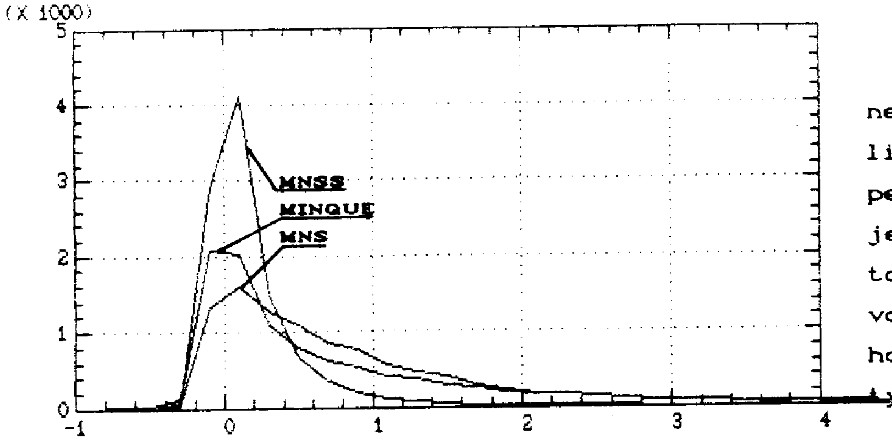
$(\text{ vec } MV_1M,\dots,\ \text{vec } MV_pM\ )\ \vartheta = Q\vartheta$.

V modeli (3) sa stratila závislosť na rušivom parametri vec $\beta\beta'$. Odhady založené na vektore $M\otimes M$ vec YY' sú invariantné. Z teórie lineárnych modelov je známe, že nutná a postačujúca podmienka pre nevychýlenú odhadnuteľnosť funkcie $f'\beta$ je $f\in R(\ Q'Q)$, kde $R(\ )$ je stĺpcový priestor matice $Q'Q$. $(\ Q'Q\ )_{ij} = \text{tr } MV_iMV_j$, čo je totožné s nutnou a postačujúcou podmienkou pre nevychýlenú invariantnú a kvadratickú odhadnuteľnosť funkcie $f'\vartheta$.

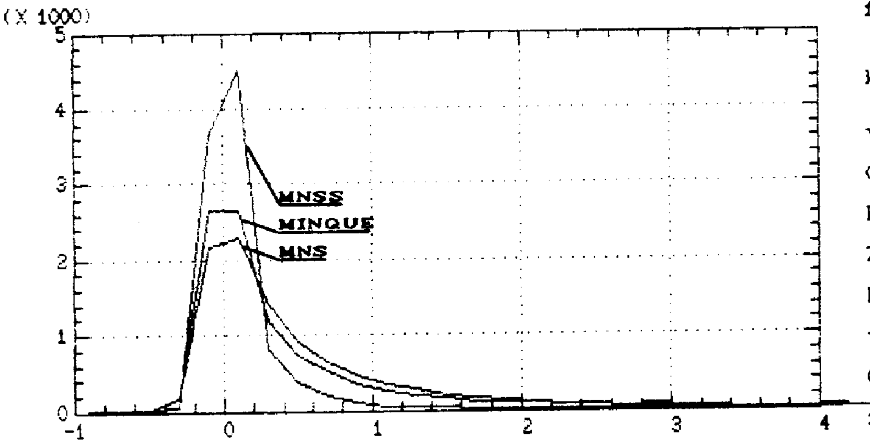
Nevychýlený a invariantný MNŠ odhad má tvar:

(4) $\quad \widehat{f'\vartheta} = f'(Q'Q)^{+}Q'M\otimes M \text{ vec } YY' =$

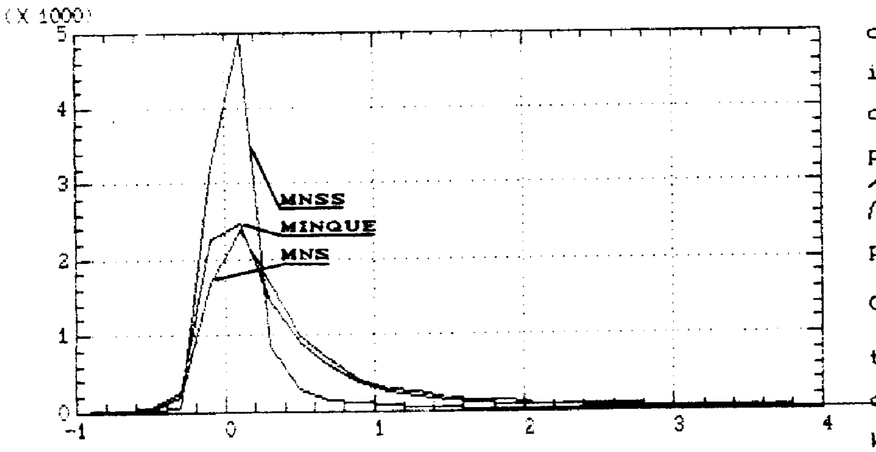
$$= \sum_{i=1}^{p} \varkappa_i Y'MV_iMY, \text{ kde } Q'Q\varkappa = f.$$



Obr. 1. Model1, odhadovaný parameter $\vartheta_1 = 0.81$, normálne rozdelenie.



Obr. 2. Model1, odhadovaný parameter $\vartheta_1 = 0.81$, transformované Studentovo rozdelenie.



Obr. 3. Model1, odhadovaný parameter $\vartheta_1 = 0.81$, transformované gama rozdelenie.

Odhad (4) vo všeobecnosti nemá žiadne vlastnosti optimality (v zmysle minimálnej disperzie) je však veľmi jednoduchý vzhľadom na výpočtovú náročnosť a nezávisí od vopred zvolenej hodnoty neznámeho parametra ϑ tak, ako odhad typu MINQUE.

MINQUE odhad (Minimum Norm Quadratic Unbiased Estimator), ktorý je invariantný, je pre MINQUE-odhadnuteľnú funkciu $f'\vartheta$ tvaru:

(5)

$$\widehat{f'\vartheta} = \sum_{i=1}^{p} \tau_i Y'(MV_oM)^+ V_i (MV_oM)^+ Y,$$

kde $V_o = V(\vartheta_o) = \sum_{i=1}^{p} \vartheta_{io} V_i$ pre vopred zvolenú hodnotu $\vartheta_o = (\vartheta_{1o}, \dots, \vartheta_{po})'$ a "+" je Moore-Penroseova inverzia matice. Zároveň platí, funkcia $f'\vartheta$ je MINQUE odhadnuteľná práve vtedy, ak $f \in R(W'W)$, kde $(W'W)_{ij} = tr(MV_oM)^+ V_i (MV_oM)^+ V_j$, z čoho ďalej platí $W'W\tau = f$.

Odhady (4) a (5) sú oba nevychýlené a invariantné, zároveň odhad (5) je v prípade normálneho rozdelenia vektora Y lokálne najlepší v bode ϑ_o.

Okrem predchádzajúcich dvoch Štulajter (1989) navrhol iný prístup k odvodeniu odhadu. Utvoril najprv MNŠ- odhad parametra β:

$\hat{\beta} = (X'X)^{-1}X'Y$ a na základe $\hat{\beta}$ položil

(6) $\overset{\square}{V} = (Y - X\hat{\beta})(Y - X\hat{\beta})'$,

teda $\overset{\square}{V}$ je "prirodzený" odhad celej kovariančnej matice $V(\vartheta)$. Štruktúra kovariančnej matice $V(\vartheta)$ je však známa, preto je prirodzené za odhad $\widehat{V(\vartheta)}$ považovať projekciu matice $\overset{\square}{V}$ do lineárneho priestoru vytvoreného množinou

$\mathcal{V} = \{ V(\vartheta), \vartheta \in \Theta \}$. Z uvedenej myšlienky dostávame

(7) $\quad f'\vartheta = \sum_{i=1}^{p} \alpha_i Y'MV_iMY$, kde $G\alpha = f$, pričom $(G)_{ij} = \operatorname{tr} V_iV_j$.

Odhad (7) na rozdiel od odhadov (4) a (5) nie je nevychýlený, len invariantný. V špeciálnom prípade, keď matice V_i sú navzájom ortogonálne Štulajter dokázal konzistenciu odhadu (7).

Je prirodzené položiť si otázku, ako sa tieto odhady (4), (5) a (7) chovajú navzájom, aké majú štatistické vlastnosti, vzhľadom na jednoduchosť výpočtov odhadov (5) a (7).

V našej simulačnej štúdii sme porovnali uvedené odhady v nasledujúcich situáciách. V dvoch rôznych modeloch sa simulovalo 10 000 realizácií vektora Y pre normálne rozdelenie, transformované Studentovo a gama rozdelenie s parametrami:

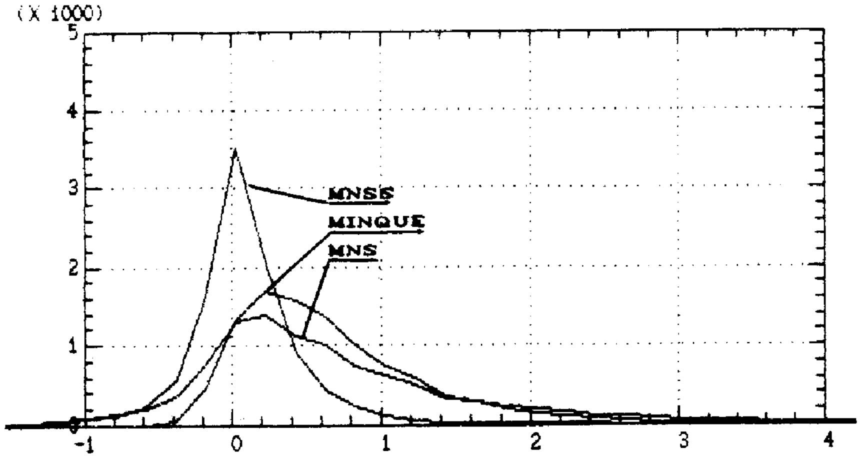
1. MODEL

$Y = X\beta + U_1\xi_1 + U_2\xi_2 + \varepsilon$, kde Y je n=24 -rozmerný vektor, matice $X_{(24\times4)}$, $U_{1,(24\times2)}$, $U_{2,(24\times6)}$ a vektor $\beta_{(4)}$ sú dané. Skutočná hodnota parametra $\vartheta_{(3)}$ použitá pri generovaní je $\vartheta=(0.81, 0.64, 0.36)'$. Pre generované vektory $\xi_{1,(2)}$, $\xi_{2,(6)}$, $\varepsilon_{(24)}$ platí:
$E(\xi_i) = 0$, $i=1,2$; $E(\varepsilon) = 0$,
$Var(\xi_i)=\vartheta_i I$ $i=1,2$ $Var(\varepsilon)=\vartheta_3 I$.

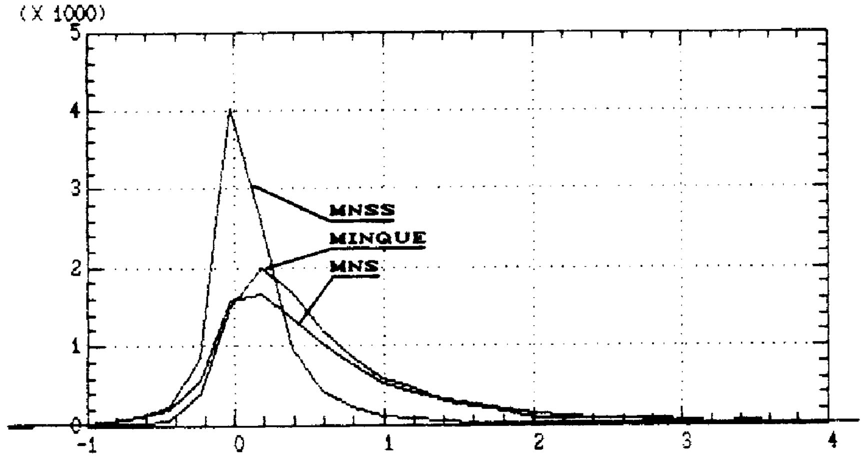
2. MODEL

$Y = X\beta + \sum_{i=1}^{p-1} U_i\xi_i + \varepsilon$, kde Y je n=12 -rozmerný vektor, matice $X_{(12\times3)}$, $U_{i,(12\times n_i)}$ a vektor $\beta_{(3)}$ sú dané, pričom $n_1=2$, $n_2=3$, $n_3=4$, $n_4=5$, $n_5=6$.

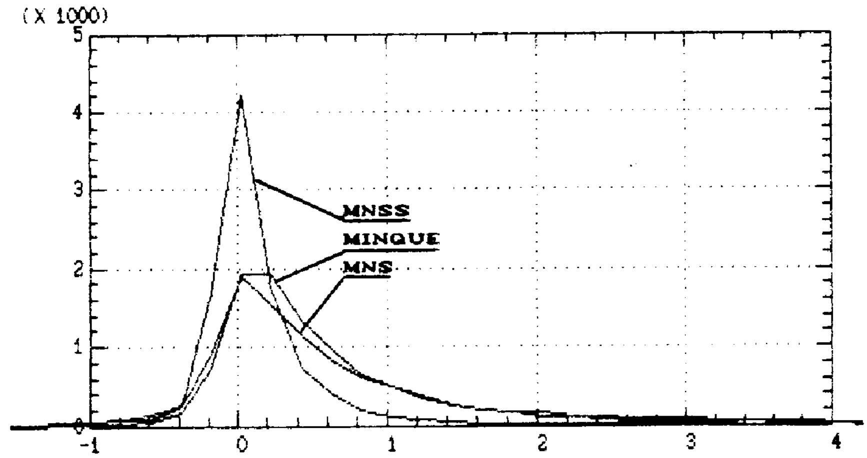
Skutočná hodnota para-



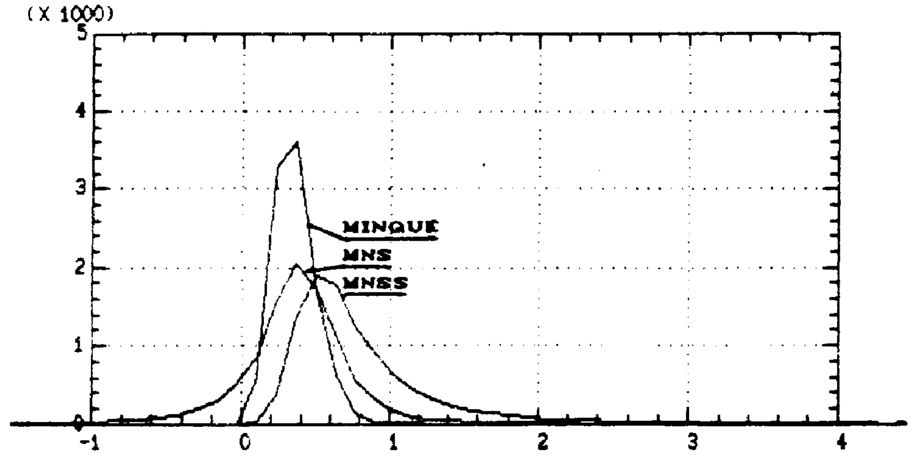
Obr. 4. Model1, odhadovaný parameter $\vartheta_2=0.64$, normálne rozdelenie.



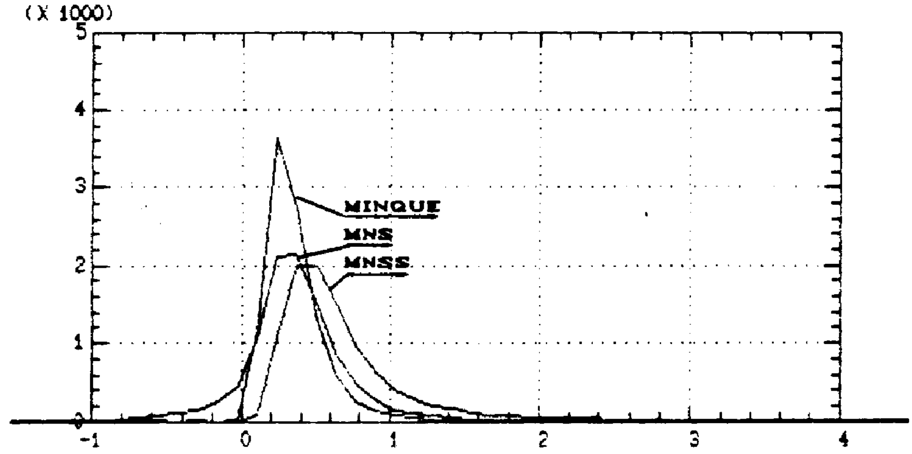
Obr. 5. Model1, odhadovaný parameter $\vartheta_2=0.64$, transformované Studentovo rozdelenie.



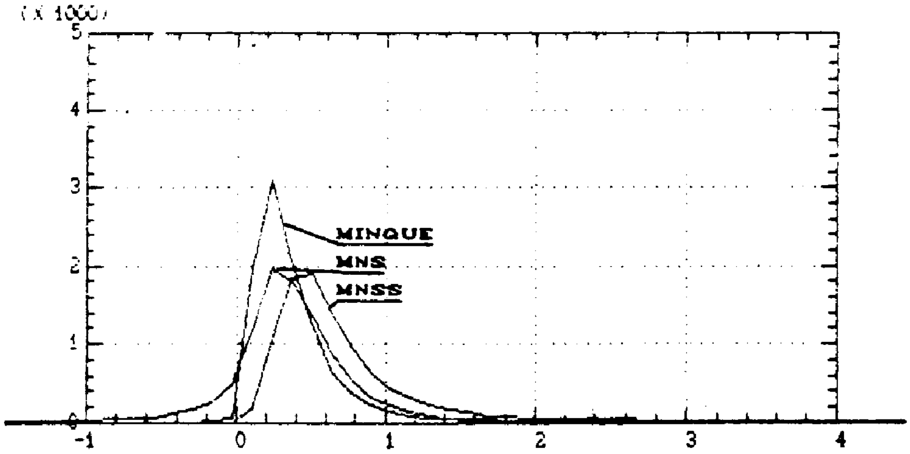
Obr. 6. Model1, odhadovaný parameter $\vartheta_2=0.64$, transformované gama rozdelenie.



Obr. 7. Model1, odhadovaný parameter $\theta_3 = 0.36$, normálne rozdelenie.



Obr. 8. Model1, odhadovaný parameter $\theta_3 = 0.36$, transformované Studentovo rozdelenie.



Obr. 9. Model1, odhadovaný parameter $\theta_3 = 0.36$, transformované gama rozdelenie.

metra $\theta_{(6)}$ pre generovanie je $\theta=(0.25, 0.81, 1, 0.49, 0.36, 0.09)'$. Pre generované vektory $\xi_{i,(n_i)}$, $\varepsilon_{(12)}$ platí: $E(\xi_i) = 0$, $Var(\xi_i)=\theta_i I_{(n_i)}$ pre $i=1,\ldots,5$ a pre ε platí $E(\varepsilon) = 0, Var(\varepsilon)=\theta_6 I_{(12)}$.

Na obrázkoch 1. - 12. sú znázornené polygóny pre jednotlivé odhady: MINQUE, MNŠ, MNŠŠ je odhad navrhnutý Štulajterom. Vo väčšine prípadov, okrem obr. 7. - 9. sa ukázalo, že napriek vychýlenosti MNŠŠ odhad má výrazne menšiu strednú kvadratickú chybu (MSE) než MINQUE a MNŠ-odhad, ktoré sú nevychýlené. Vo všeobecnosti rozdelenie MINQUE ani iných kvadratických odhadov nie je známe, len v niektorých špeciálnych prípadoch a za predpokladu normality vektora Y. Z obrázkov je možné vidieť, že rozdelenia MINQUE a MNŠ sú si blízke v zmysle podobnosti polygónov.

Ukázalo sa tiež, že rozdelenia odhadov sú si blízke v uvedenom zmysle aj pre rôzne typy rozdelenia (symetrické, napr. Studentovo aj nesymetrické, napr. gama rozdelenie).

V tabuľkách 1. - 3. sú uvedené výberové priemery, disperzie a stredná kvadratická chyba odhadov parametrov θ_1, θ_2 a θ_3. V tabuľke 4 je to isté pre odhad funkcie $f(\theta) = \sum_{i=1}^{6} \theta_i = 3.0$.

Ako uzáver je možno povedať, že v situáciách, keď výpočtová náročnosť MINQUE odhadu je veľká (výpočet matice $(MVM)^+$), je lepšie uprednostniť MNŠ alebo MNŠŠ odhad, a to v závislosti od typu úlohy, najmä v prípade odhadu lineárnej funkcie parametra ϑ.

L I T E R A T Ú R A

[1] **Kubáček,L.**: Locally best quadratic estimators. Math. Slovaca 35,(1985), No.4,pp.393-408.

[2] **Rao,C.R.**: Estimation of variance and covariance components-MINQUE theory. J. Multivar. Anal. 1, (1971), pp. 257-275.

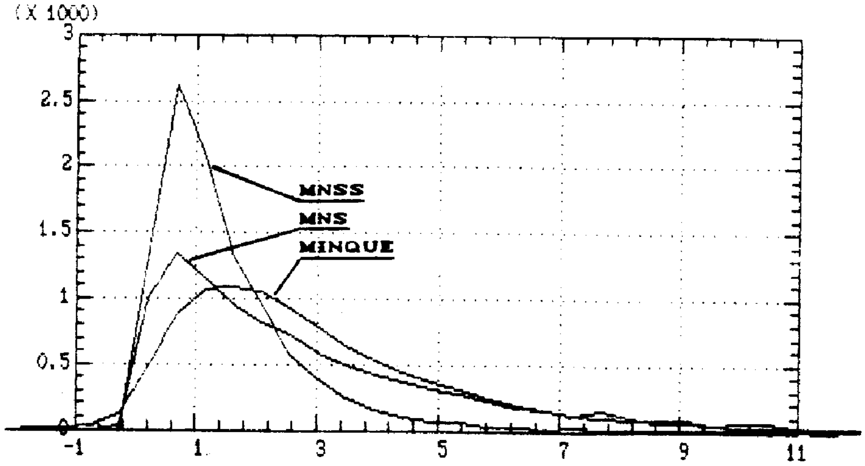
[3] **Rao,C.R.**: Estimation of variance and covariance components in linear models. JASA 67, (1972), pp. 112-115.

[4] **Rao,C.R.,Kleffe,J.**: Esimation of variance components. Handbook of Statistics, Vol.1, P.R. Krishnaiah, ed., North Holland Pub. Comp.,(1980) pp. 1-40.

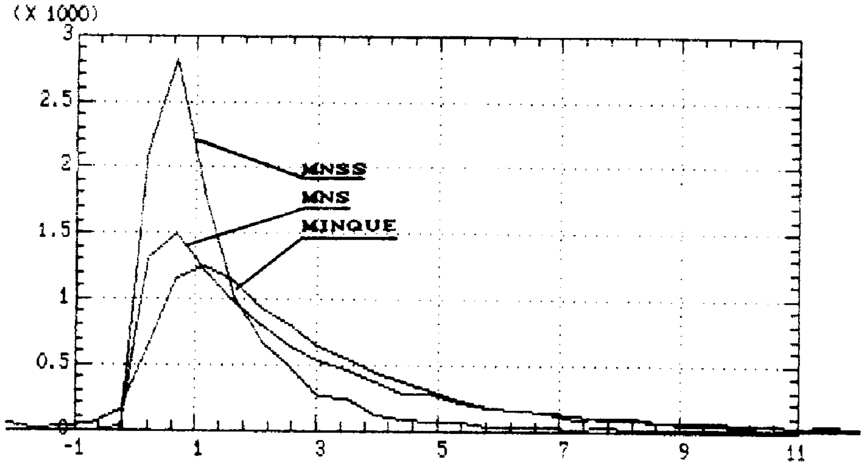
[5] **Rao,C.R.,Kleffe,J.**: Estimation of variance components and applications. North Holland (1988) New York.

[6] **Verdooren,L.R.**: Practical aspects of variance component estimation. Invited lecture for the 4th International Summer School on Problems of Model Choice and Parameter Estimation in Regression Analysis. 7th - 16th May 1979, Mulhausen, GDR.

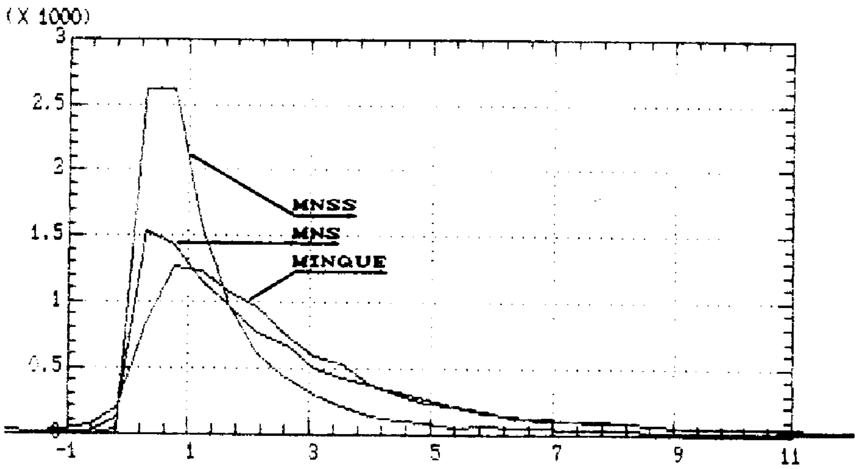
[7] **Verdooren,L.R.**: Least squares estimators and non-negative estimators of variance components. Commun. Statist.-Theory Meth., 17(4), (1988), pp. 1027-1051.



Obr.10. Model2. odhadovaný parameter ϑ=3.0, normálne rozdelenie.



Obr.11. Model2, odhadovaný parameter ϑ=3.0, transformované Studentovo rozdelenie.



Obr.12. Model2, odhadovaný parameter ϑ=3.0, transformované gama rozdelenie.

Parameter	Rozdelenie	Odhad	Priemer	Disperzia	MSE
$\vartheta_1 = 0.81$	Normálne	MINQUE MNŠ MNŠŠ	0.8198 0.8127 0.1732	0.99417 1.5680 0.0858	0.4913
	Studentovo	MINQUE MNŠ MNŠŠ	0.8027 0.7630 0.1613	20.0972 13.3103 0.7419	1.1627
	Gama	MINQUE MNŠ MNŠŠ	0.8115 0.8084 0.1717	4.0910 7.1408 0.3956	0.8030

Tab. 1.

Parameter	Rozdelenie	Odhad	Priemer	Disperzia	MSE
$\vartheta_2 = 0.64$	Normálne	MINQUE MNŠ MNŠŠ	0.6464 0.6562 0.0819	0.3736 0.9082 0.1189	0.4304
	Studentovo	MINQUE MNŠ MNŠŠ	0.6309 0.6240 0.0667	0.9142 1.9894 0.9050	1.2337
	Gama	MINQUE MNŠ MNŠŠ	0.6250 0.6357 0.0762	0.9547 1.6542 0.3089	0.6268

Tab. 2.

Parameter	Rozdelenie	Odhad	Priemer	Disperzia	MSE
$\vartheta_3 = 0.36$	Normálne	MINQUE MNŠ MNŠŠ	0.3571 0.3561 0.7929	0.0191 0.1425 0.2300	0.4174
	Studentovo	MINQUE MNŠ MNŠŠ	0.3573 0.3578 0.7883	0.0508 0.1894 4.1380	4.3214
	Gama	MINQUE MNŠ MNŠŠ	0.3642 0.3667 0.8100	0.0662 0.2026 1.2803	1.4828

Tab. 3.

Parameter	Rozdelenie	Odhad	Priemer	Disperzia	MSE
$\vartheta = 3.0$	Normálne	MINQUE MNŠ MNŠŠ	3.0082 3.0288 1.4540	5.7118 8.9197 1.4098	3.7999
	Studentovo	MINQUE MNŠ MNŠŠ	2.9655 2.9474 1.4075	10.8945 13.2648 3.3239	5.8600
	Gama	MINQUE MNŠ MNŠŠ	2.9918 2.9698 1.4339	11.3257 14.0818 2.6078	5.0605

Tab. 4.

Adresa: Ústav merania a meracej techniky SAV, Dúbravská 9,84219 Bratislava