

# O PRE-TEST ODHADECH A STEINOVÝCH ODHADECH V REGRESI

MICHAL FRIESL A JIŘÍ REIF

*Katedra matematiky ZČU, Univerzitní 8, 306 14 Plzeň  
friesl@kma.zcu.cz, reif@kma.zcu.cz*

ABSTRACT. Comments on pre-test estimation and a short review of shrinkage estimation in the setting of classical linear regression models are presented.

## 1. ÚVOD

Jednoduchý regresní model ve tvaru přímky

$$y_i = \beta_0 + \beta_1 x_i + e_i, \quad i = 1, \dots, n,$$

s neznámými parametry  $\beta_0$  a  $\beta_1$  a náhodnými odchylkami  $e_i$  nebo model s kvadratickou závislostí

$$y_i = \beta_0 + \beta_1 x_i + \beta_2 x_i^2 + e_i, \quad i = 1, \dots, n,$$

stejně jako model  $y_i = \beta_0 + \beta_1 x_i + \beta_2 z_i + e_i$  se dvěma vysvětlujícími proměnnými, spadají do teorie lineární regrese, kde linearitou je myšlena lineární závislost vysvětlované proměnné na parametrech modelu. V případě nenáhodných vysvětlujících proměnných můžeme takový model zapsat maticově ve tvaru

$$(1.1) \quad y = X\beta + e,$$

kde  $X$  je známá pevná (nenáhodná) matice typu  $n \times p$ ,  $p < n$ , u níž budeme předpokládat hodnotu  $p$ ,  $\beta$  je vektor neznámých parametrů a  $e$  je náhodný vektor chyb. Sloupce matice  $X$  se nazývají regresory. Budeme dále předpokládat, že vektor  $e$  splňuje podmínky

$$Ee = 0 \quad \text{a} \quad \text{cov } e = \sigma^2 I_n,$$

kde  $\sigma^2$  je neznámý kladný parametr. V případě, že matice  $X$  nebude pevná, ale jejími prvky budou realizace nějakých náhodných vysvětlujících proměnných, budou naše závěry pravdivé při podmínění danými hodnotami vysvětlujících proměnných.

V domácí literatuře je tento model podrobně studován např. v monografii [70] a učebních textech [26], [27]. V tomto textu se zaměříme na popis některých odhadů vektoru regresních koeficientů  $\beta$ . Nekladli jsme si za cíl vytvořit přehled „bez mezer“, ale pouze doplnit některé informace k výše uvedeným českým publikacím.

Náš plán je následující. Po zavedení základních pojmů v odst. 2 a krátkém připomenutí některých vlastností odhadu metodou nejmenších čtverců (MNČ) v odst. 3 se v dalších dvou odstavcích zmíníme o tzv. pre-test odhadech, které se často používají v praxi, a o odhadech Steinova typu (zkrácených a sražených odhadech), které jsou lepší než odhad MNČ. Odhadům parametru  $\sigma^2$  se v tomto pojednání nebudeme věnovat a odkazujeme čtenáře např. na [33] str. 18, popř. na [39] nebo [29].

---

Tato práce vznikla za podpory výzkumného záměru MSM 235200001.

## 2. SROVNÁVACÍ KRITÉRIA

Odchylku vektoru  $\hat{\beta}$  od vektoru  $\beta$  v prostoru  $R^p$  budeme měřit pomocí kvadratické ztrátové funkce

$$(2.1) \quad L_W(\beta, \hat{\beta}) = (\hat{\beta} - \beta)' W (\hat{\beta} - \beta),$$

kde  $W$  bude nějaká symetrická nezáporně definitní matice typu  $p \times p$  (čtvercovou matici budeme nazývat nezáporně definitní, jestliže kvadratická forma určená touto maticí nabývá pouze nezáporných hodnot; v literatuře se též používá názvu pozitivně semidefinitní matice), a kvalitu odhadu  $\hat{\beta}$  vektoru  $\beta$  budeme posuzovat pomocí střední kvadratické chyby (v angl. *mean square error*) definované vztahem

$$(2.2) \quad \text{MSE}_W(\beta, \hat{\beta}) = E L_W(\beta, \hat{\beta}).$$

Výraz (2.2), chápaný jako funkci  $\beta$ , budeme nazývat rizikovou funkcí. Nebude-li pochybnost o tom, který vektor parametrů odhadujeme pomocí  $\hat{\beta}$ , budeme místo (2.2) psát stručněji  $\text{MSE}_W \hat{\beta}$ .

Uveďme nejčastější volby váhové matice  $W$ . Při volbě jednotkové váhové matice  $W = I_p$  je ztrátová funkce kvadrátem eukleidovské vzdálenosti  $\beta$  a  $\hat{\beta}$ . Pro úlohu odhadu vektoru  $\beta$  v modelu (1.1) se však častěji volí  $W = X'X$ . Protože platí

$$\text{MSE}_{X'X} \hat{\beta} = E(\hat{\beta} - \beta)' X' X (\hat{\beta} - \beta) = E(X\hat{\beta} - X\beta)' (X\hat{\beta} - X\beta),$$

je  $\text{MSE}_{X'X} \hat{\beta}$  střední hodnotou kvadrátu eukleidovské vzdálenosti vektoru „vyrovnaných“ hodnot  $X\hat{\beta}$  a vektoru teoretických regresních hodnot  $X\beta$ .

Je-li  $\ell \in R^p$  a chceme odhadovat parametrickou kombinaci  $\ell'\beta$ , je vhodné volit váhovou matici  $W = \ell\ell'$ , pro kterou pak platí

$$\text{MSE}_{\ell\ell'}(\beta, \hat{\beta}) = E(\hat{\beta} - \beta)' \ell\ell' (\hat{\beta} - \beta) = E(\ell'\hat{\beta} - \ell'\beta)^2.$$

**Definice 2.1.** *Nechť  $\hat{\beta}$ ,  $\tilde{\beta}$  jsou dva odhady vektoru  $\beta \in R^p$  a  $W$  je symetrická nezáporně definitní matice typu  $p \times p$ . Řekneme, že  $\hat{\beta}$  dominuje  $\tilde{\beta}$  ve smyslu  $\text{MSE}_W$ , jestliže pro všechny  $\beta \in R^p$  platí  $\text{MSE}_W \hat{\beta} \leq \text{MSE}_W \tilde{\beta}$ . Platí-li navíc pro některé  $\beta$  ostrá nerovnost, řekneme, že  $\hat{\beta}$  ostře dominuje  $\tilde{\beta}$  ve smyslu  $\text{MSE}_W$ .*

Někteří autoři porovnávají odhady vektoru  $\beta$  pomocí maticové MSE.

**Definice 2.2.** *Nechť  $\hat{\beta}$  je odhad vektoru  $\beta$ . Matici*

$$\text{MSE} \hat{\beta} = \text{MSE}(\beta, \hat{\beta}) = E(\hat{\beta} - \beta)(\hat{\beta} - \beta)'$$

*budeme nazývat maticovou střední kvadratickou chybou.*

**Věta 2.3.** *Nechť  $\hat{\beta}$  a  $\tilde{\beta}$  jsou odhady vektoru  $\beta \in R^p$  a nechť  $\text{MSE} \hat{\beta}$  a  $\text{MSE} \tilde{\beta}$  jsou jejich maticové střední kvadratické chyby. Následující podmínky jsou ekvivalentní.*

- (i) *Pro všechny  $\beta \in R^p$  je matice  $\text{MSE} \tilde{\beta} - \text{MSE} \hat{\beta}$  nezáporně definitní.*
- (ii) *Pro všechny  $\ell \in R^p$  a  $\beta \in R^p$  platí  $E(\ell'\tilde{\beta} - \ell'\beta)^2 \leq E(\ell'\hat{\beta} - \ell'\beta)^2$ .*
- (iii) *Pro všechny symetrické nezáporně definitní matice  $W$  typu  $p \times p$  odhad  $\hat{\beta}$  dominuje  $\tilde{\beta}$  ve smyslu  $\text{MSE}_W$  (viz def. 2.1).*

*Důkaz.* Podmínka (i) znamená, že pro každý vektor  $\ell \in R^p$  a  $\beta \in R^p$  platí

$$\ell'(\text{MSE} \tilde{\beta} - \text{MSE} \hat{\beta})\ell \geq 0,$$

neboli podle definice  $\text{MSE} \tilde{\beta}$  a  $\text{MSE} \hat{\beta}$

$$(2.3) \quad E \ell'(\tilde{\beta} - \beta)(\tilde{\beta} - \beta)'\ell - E \ell'(\hat{\beta} - \beta)(\hat{\beta} - \beta)'\ell \geq 0.$$

Protože součiny  $\ell'(\tilde{\beta} - \beta) = (\tilde{\beta} - \beta)'\ell$  a  $\ell'(\hat{\beta} - \beta) = (\hat{\beta} - \beta)'\ell$  patří do  $R$ , je (2.3) ekvivalentní (ii), tedy (i) je ekvivalentní (ii). Z téhož důvodu je (i) ekvivalentní podmínce, že  $\hat{\beta}$  dominuje  $\tilde{\beta}$  ve smyslu  $\text{MSE}_{\ell\ell'}$  pro všechna  $\ell \in R^p$ . Z této podmínky však plyne (iii), neboť každá symetrická matice  $W$  typu  $p \times p$  lze vyjádřit ve tvaru tzv. spektrálního rozkladu

$$W = \sum_{i=1}^p \lambda_i \ell_i \ell_i',$$

kde  $\lambda_i$  jsou vlastní čísla matice  $W$  (a tedy  $\lambda_i \geq 0$ , je-li  $W$  nezáporně definitní) a  $\ell_i$  jsou odpovídající vlastní vektory (např. [70], str. 21–22). Opačná implikace plyne ze skutečnosti, že pro každé  $\ell \in R^p$  je matice  $W = \ell\ell'$  symetrická a nezáporně definitní; symetrie je zřejmá a pro každé  $v \in R^p$  platí  $v'\ell\ell'v = (v'\ell)(\ell'v) = (v'\ell)^2$ .  $\square$

**Definice 2.4.** Řekneme, že odhad  $\hat{\beta}$  vektoru  $\beta$  dominuje odhadu  $\tilde{\beta}$  ve smyslu maticové MSE, jestliže pro všechna  $\beta \in R^p$  je splněna některá z ekvivalentních podmínek věty 2.3. Je-li  $\mathcal{T}$  nějaká třída odhadů vektoru  $\beta$  a  $\hat{\beta} \in \mathcal{T}$  je odhad, který dominuje ve smyslu maticové MSE každému  $\tilde{\beta} \in \mathcal{T}$ , řekneme, že  $\hat{\beta}$  je nejlepším odhadem v  $\mathcal{T}$ .

**Definice 2.5.** Řekneme, že odhad  $\hat{\beta}$  vektoru  $\beta$  je přípustný (resp. nepřípustný) ve smyslu  $\text{MSE}_W$ , jestliže neexistuje (resp. existuje) odhad vektoru  $\beta$ , který ostře dominuje  $\hat{\beta}$  ve smyslu  $\text{MSE}_W$  (viz definice 2.1).

Hodnotit kvalitu odhadu jen podle hlediska, zda je či není přípustný ve smyslu  $\text{MSE}_W$ , není vhodné. Lze např. ukázat, že je-li  $W$  symetrická pozitivně definitní matice, pak odhad  $\hat{\beta} \equiv 0$  (popř. jiný konstantní odhad) je přípustný ve smyslu  $\text{MSE}_W$ , zatímco (jak uvidíme v odst. 5) odhad  $b$  pořízený pomocí MNČ pro  $p \geq 3$  tuto vlastnost nemá. Důvodem přípustnosti odhadu  $\hat{\beta} \equiv 0$  je skutečnost, že pro  $\beta = 0$  má tento odhad nulovou  $\text{MSE}_W(\beta, \hat{\beta})$  a žádný jiný odhad (zanedbáme-li změny na množině míry 0) tuto vlastnost nemá. Pro  $\|\beta\| \rightarrow \infty$  však  $\text{MSE}_W(\beta, 0)$  roste do nekonečna, zatímco  $\text{MSE}_W(\beta, b)$  je konstantní. Z hlediska velikosti  $\text{MSE}_W(\beta, \cdot)$  pro  $\beta$  „nejméně příznivé“ pro daný odhad je tedy MNČ výhodnější než  $\hat{\beta} \equiv 0$ ; optimální z tohoto hlediska budou tzv. minimax odhady.

**Definice 2.6.** Nechť  $\mathcal{T}$  je nějaká třída odhadů  $\beta$  a  $\beta^* \in \mathcal{T}$  je odhad takový, že pro každé  $\hat{\beta} \in \mathcal{T}$  platí

$$(2.4) \quad \sup_{\beta \in R^p} \text{MSE}_W(\beta, \beta^*) \leq \sup_{\beta \in R^p} \text{MSE}_W(\beta, \hat{\beta}).$$

Pak řekneme, že  $\beta^*$  je minimax odhadem  $\beta$  ve třídě  $\mathcal{T}$  ve smyslu  $\text{MSE}_W$ . Platí-li (2.4) pro všechny odhady  $\hat{\beta}$  včetně tzv. znáhodněných odhadů (pro definici znáhodněných odhadů viz např. [41], str. 37), řekneme, že  $\beta^*$  je minimax odhadem  $\beta$  ve smyslu  $\text{MSE}_W$ .

Je zřejmé, že je-li  $\beta^*$  minimax odhadem  $\beta$  a  $\hat{\beta}$  je odhad, který dominuje  $\beta^*$ , pak  $\hat{\beta}$  je rovněž minimax odhadem  $\beta$  (vše ve smyslu  $\text{MSE}_W$  pro pevnou váhovou matici  $W$ ).

V některých případech se provádí také porovnání odhadů ve smyslu Pitmana, viz např. přehled [19] a citace v něm obsažené, které je definováno následujícím způsobem.

**Definice 2.7.** Řekneme, že odhad  $\hat{\beta}$  vektoru  $\beta$  dominuje odhadu  $\tilde{\beta}$  v Pitmanově smyslu při ztrátové funkci  $L$ , jestliže pro každé  $\beta \in R^p$  platí

$$P(L(\beta, \hat{\beta}) < L(\beta, \tilde{\beta})) \geq P(L(\beta, \hat{\beta}) > L(\beta, \tilde{\beta})).$$

Dominance ve smyslu Pitmana není obecně tranzitivní relací.

### 3. VLASTNOSTI METODY NEJMENŠÍCH ČTVERCŮ

Nejpoužívanější metodou odhadu vektoru  $\beta$  modelu (1.1) je metoda nejmenších čtverců (MNČ). Odhad  $\beta$  spočtený touto metodou budeme značit  $b$ . Jak známo, je

$$(3.1) \quad b = (X'X)^{-1}X'y,$$

a tedy  $b$  je lineární odhad (vzhledem k  $y$ ). Podle známé Gaussovy-Markovovy věty je  $b$  nejlepším odhadem  $\beta$  ve třídě nestranných lineárních odhadů ve smyslu definice 2.4 (odhad  $\hat{\beta}$  vektoru  $\beta$  se nazývá nestranný, jestliže pro každé  $\beta \in R^p$  je  $E\hat{\beta} = \beta$ ). Poměrně jednoduchý přímý důkaz věty lze najít v [26].

V případě, že vektor chyb má normální rozdělení, tj.  $e \sim N_n(0, \sigma^2 I_n)$ , je  $b$  nejlepší dokonce ve třídě všech nestranných (tedy i nelineárních) odhadů vektoru  $\beta$ . Důkaz tohoto tvrzení lze najít např. v [59], str. 152. Dobré vlastnosti MNČ v případě normality vektoru chyb  $e$  souvisí s tím, že v tomto případě odhad MNČ je také odhadem metodou maximální věrohodnosti, viz např. [26], str. 66. Pro jiná rozdělení však (alespoň v typickém případě) metoda maximální věrohodnosti nespĺývá s MNČ; např. pro symetrické exponenciální rozdělení odchylek vede metoda maximální věrohodnosti k minimalizaci součtu absolutních hodnot reziduí ([26] aj.).

Rao dokázal, že dává-li MNČ nejlepší odhad ve třídě nestranných odhadů, pak za jistých velmi obecných předpokladů musí mít náhodný vektor  $e$  normální rozdělení, viz [51], str. 359. Pro nenormální rozdělení chybového vektoru  $e$  může MNČ poskytovat značně horší odhad než jiné metody. Speciálními postupy, které se uplatňují v případě, že máme podezření na nenormalitu vektoru  $e$  nebo na porušení předpokladu  $\text{cov } e = \sigma^2 I_n$ , se v tomto článku zabývat nebudeme a odkazujeme čtenáře zejména na publikace [27], [70], [2] (odst. 2.8, 3.4.3. a 3.5.2.) a dále např. na [1] nebo monografii [54].

I v případě, že klasické předpoklady o vektoru  $e$  a jeho normalita jsou splněny, nemusí být realizace MNČ bez problémů. Na potíže narazíme např. při silné multi-kolinearitě regresorů, kdy sloupce regresní matice  $X$  mají blízko k lineární závislosti, viz [70], [25] a [26]. S vhodnými numerickými algoritmy pro MNČ se čtenář může seznámit v [70], [2] odst. 2.7. a náročný čtenář např. v [40].

Všimněme si ještě, za jakých podmínek poskytuje MNČ minimax odhad  $\beta$ .

**Věta 3.1.** *Nechť váhová matice  $W$  určující ztrátovou funkci (2.1) je symetrická pozitivně definitní. Pak odhad MNČ vektoru  $\beta$  je minimax odhadem ve třídě všech lineárních odhadů  $\beta$  ve smyslu  $\text{MSE}_W$ . Má-li vektor chyb  $e$  normální rozdělení, pak odhad MNČ je minimax odhadem (ve třídě všech odhadů  $\beta$ ) ve smyslu  $\text{MSE}_W$ .*

Věta 3.1 je dobře známa pro případ  $W = I_p$ , viz např. [32], str. 19–20. Obecný případ symetrické pozitivně definitní váhové matice lze na tento případ transformovat podle následující poznámky.

**Poznámka 3.2.** *Uvažujme model (1.1) a ztrátovou funkci (2.1). Ukážeme, že vhodnou reparametrizací lze model převést na případ ortonormální regresní matice. Je-li váhová matice  $W$  symetrická pozitivně definitní, lze model transformovat tak, že ztrátová funkce má v reparametrizovaném modelu jednotkovou váhovou matici. Je-li*

$W = X'X$ , lze vhodnou lineární transformací dosáhnout obou vlastností zároveň, tj. ortonormální regresní matice a jednotkové váhové matice.

Nechť  $A$  je nějaká regulární matice typu  $p \times p$ . Označme  $\theta = A\beta$  a  $Z = XA^{-1}$ . Model (1.1) lze pak zapsat ve tvaru

$$(3.2) \quad y = Z\theta + e.$$

Mezi odhady  $\hat{\beta}$  vektoru  $\beta$  a  $\hat{\theta}$  vektoru  $\theta$  definujeme vzájemně jednoznačné přiřazení vztahem  $\hat{\theta} = A\hat{\beta}$  (tj.  $\hat{\beta} = A^{-1}\hat{\theta}$ ). Pak

$$L_W(\beta, \hat{\beta}) = (\hat{\theta} - \theta)' \widetilde{W} (\hat{\theta} - \theta),$$

kde  $\widetilde{W} = (A^{-1})'WA^{-1}$ . Je-li matice  $W$  symetrická nezáporně definitní, platí totéž o matici  $\widetilde{W}$ . Všimněme si, že odhadu MNČ vektoru  $\beta$  odpovídá při výše zmíněném jednoznačném přiřazení odhad MNČ vektoru  $\theta$ , neboť platí

$$\begin{aligned} (X'X)^{-1}X' &= [(ZA)'(ZA)]^{-1}(ZA)' = (A'Z'ZA)^{-1}A'Z' \\ &= A^{-1}(Z'Z)^{-1}(A')^{-1}A'Z' = A^{-1}(Z'Z)^{-1}Z'. \end{aligned}$$

Pro libovolnou symetrickou pozitivně definitní matici  $S$  existuje čtvercová regulární matice  $A$  taková, že  $S = A'A$  (toto tvrzení je důsledkem Věty 2.12 v [70] a skutečnosti, že pozitivně definitní matice je regulární).

- a) Protože matice  $X$  má maximální hodnotu, je matice  $X'X$  symetrická pozitivně definitní a existuje regulární matice  $A$  taková, že  $A'A = X'X$ . Pak pro  $Z = XA^{-1}$  platí

$$Z'Z = (A^{-1})'X'XA^{-1} = (A^{-1})'A'AA^{-1} = (A^{-1})'A' = (AA^{-1})' = I_p.$$

- b) Je-li váhová matice  $W$  symetrická pozitivně definitní, existuje regulární matice  $A$  taková, že  $A'A = W$ . Pak

$$\widetilde{W} = (A^{-1})'WA^{-1} = (A^{-1})'A'AA^{-1} = I_p.$$

V posledních 40 letech byly intenzivně studovány také vychýlené (tj. nikoliv ne-stranné) odhady vektoru  $\beta$ . Některé z nich při počtu regresorů  $p \geq 3$  a normalitě vektoru  $e$  (často i pro obecnější chybový vektor  $e$ ) dominují odhadu MNČ ve smyslu  $MSE_W$  např. pro váhovou matici  $W = X'X$  nebo  $W = I_p$ . Použijeme-li porovnání v Pitmanově smyslu při ztrátové funkci  $L_W$  s pozitivně definitní maticí  $W$ , lze sestavit odhad dominující MNČ již při  $p = 2$ . Těmto odhadům se budeme věnovat v odstavci 5. Ještě předtím se však zmíníme o jednom postupu běžně užívaném v praxi, který rovněž vede k vychýleným odhadům vektoru  $\beta$ .

#### 4. PRE-TEST ODHADY

Součástí regresní analýzy bývá obvykle i vhodný výběr vysvětlujících proměnných. S touto problematikou se čtenář může seznámit např. v [25], [26] odst. 4.3, [27], [70], [47], [33], náročnému čtenáři doporučujeme [45], [7], [44], [68] a literaturu v nich citovanou.

V tomto odstavci se budeme zabývat následujícím v praxi běžným postupem. Regresní matici  $X$  rozdělíme na dvě části,  $X = (U, V)$ . V části  $U$  jsou regresory, které v modelu určitě ponecháme, neboť z nějakého důvodu jsme přesvědčeni o jejich důležitosti bez ohledu na právě analyzovaná data. Regresory umístěné v části  $V$  pak s ohledem na data buď v regresi ponecháme, nebo je všechny z modelu vyřadíme. Obvyklé způsoby rozhodování v praxi sice bývají složitější (vícestupňové), ale je obtížné takové postupy obecně analyzovat.

Úplný model má tedy tvar

$$(4.1) \quad y = X\beta + e = U\alpha + V\gamma + e,$$

kde o vektoru  $e$  budeme v tomto odstavci předpokládat, že má normální rozdělení  $N_n(0, \sigma^2 I_n)$ , a redukovaný model má tvar

$$(4.2) \quad y = U\alpha + e^*,$$

kde  $e^* = V\gamma + e$ . V případě volby modelu (4.1) odhadujeme  $\beta$  pomocí MNČ v úplném modelu, tj. pomocí (3.1), v případě volby modelu (4.2) odhadujeme  $\beta$  pomocí

$$b_0 = (a', 0, \dots, 0)',$$

kde  $a = (U'U)^{-1}U'y$  je odhad  $\alpha$  pomocí MNČ v modelu (4.2). Mezi modely se budeme rozhodovat na základě statistiky  $F$ , která se standardně používá při testu hypotézy  $\gamma = 0$  ([70], [26]). Celkově má tedy odhad  $\hat{\beta}$  vektoru  $\beta$  tvar

$$\hat{\beta} = \begin{cases} b & \text{pro } F \geq F_{\text{krit}} \\ b_0 & \text{pro } F < F_{\text{krit}}, \end{cases}$$

kde  $F_{\text{krit}}$  je kritická hodnota pro zvolenou hladinu významnosti, resp. obecněji  $F_{\text{krit}}$  je nějaké nezáporné číslo nebo  $F_{\text{krit}} = \infty$ . V případě, že matice  $V$  má pouze jeden sloupec, a tedy  $\gamma \in R$ , se obvykle místo statistiky  $F$  použije statistika  $|t|$ , viz např. [70], str. 63.

**Definice 4.1.** *Výše popsaný odhad  $\hat{\beta}$  budeme nazývat pre-test odhadem, zkráceně PTE (z angl. preliminary test estimator, pre-test estimator nebo pretest estimator, výjimečně se používá i název testimator).*

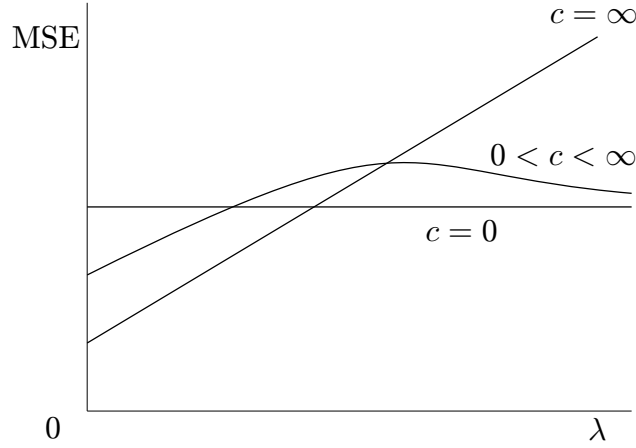
Prvním, kdo analyzoval pre-test odhady důkladněji z teoretického hlediska, byl patrně Bancroft [3]. Později jim byla věnována pozornost v desítkách prací, viz např. [32], [21], [43], [52] a citace v nich obsažené.

Nechť  $W$  je nějaká symetrická pozitivně definitní matice. Na obr. 1 je znázorněna závislost  $\text{MSE}_W(\beta, \hat{\beta})$  pre-test odhadu  $\hat{\beta}$  příslušejícího kritické hodnotě  $F_{\text{krit}} = c$  na tzv. parametru necentrality  $\lambda$ , který je kvadratickou formou  $\gamma$  ([32]), při kritických hodnotách  $c = 0$ ,  $c = \infty$  a  $c \in (0, \infty)$ .

Problémem volby kritické hodnoty  $F_{\text{krit}}$  tak, aby příslušný PTE byl v jistém smyslu optimální, se zabývala řada autorů ([55], [64], [32], [63], [53], [69], [52] aj.). Neexistuje totiž žádné  $c \in [0, \infty]$  takové, aby PTE s kritickou hodnotou  $F_{\text{krit}} = c$  dominoval (ve smyslu  $\text{MSE}_W$ ) ostatní PTE.

Popišme stručně přístup, který k volbě „optimálního“ PTE zaujali Reif a Vlček v nedávné publikaci [52]. Tito autoři se omezili na případ, kdy matice  $V$  má jeden sloupec, takže  $\gamma \in R$ . Zvažuje se tedy vyřazení jednoho konkrétního regresoru. Protože  $\text{MSE}_W(\beta, \hat{\beta})$  závisí na  $\beta$  pouze prostřednictvím složky  $\gamma$ , integrovali autoři tuto rizikovou funkci vzhledem ke  $\gamma$  s využitím nějaké váhové integrovatelné funkce  $f(\gamma)$ , kterou může být např. apriorní rozdělení  $\gamma$ . Lze tedy říci, že kvalita PTE byla měřena pomocí bayesovského rizika, které reprezentuje průměrnou hodnotu  $\text{MSE}_W$  vzhledem k určitému rozdělení složky  $\gamma$ . Přitom byly uvažovány *zobecněné* PTE, které kombinovaly úplný model a redukovaný model v závislosti na tom, zda  $t \in S$  nebo  $t \notin S$ ,

$$\hat{\beta} = \begin{cases} b & \text{pro } t \in S \\ b_0 & \text{pro } t \notin S, \end{cases}$$



OBRÁZEK 1. Závislost  $\text{MSE}_W(\beta, \hat{\beta})$  pre-test odhadu  $\hat{\beta}$  na parametru necentrality  $\lambda$  při různých volbách kritické hodnoty  $F_{\text{krit}} = c$

kde  $S$  je nějaká obecná měřitelná podmnožina  $R$  určující daný PTE. Při takto postavené úloze vždy existuje nejlepší zobecněný PTE.

Autory zajímalo, pro která rozdělení  $f(\gamma)$  se optimální zobecněný PTE chová jako tradiční PTE v tom smyslu, že množina  $S$  odpovídající tomuto zobecněnému PTE obsahuje okolí  $\pm\infty$  a neobsahuje okolí 0. Výsledky lze shrnout následovně.

1. Existují kladná čísla  $\sigma_1^2 > \sigma_2^2$  taková, že má-li  $f(\gamma)$  pro  $\gamma \rightarrow +\infty$  chvost těžší než normální rozdělení s rozptylem  $\sigma_1^2$  (resp. lehčí než normální rozdělení s rozptylem  $\sigma_2^2$ ), pak pro velký počet stupňů volnosti je optimální zobecněný PTE pro velká  $t$  roven odhadu MNČ v úplném modelu (resp. v redukovaném modelu), tedy množina  $S$  charakterizující optimální zobecněný PTE obsahuje (resp. neobsahuje) okolí  $+\infty$ . Čísla  $\sigma_1^2$  a  $\sigma_2^2$  závisí pouze na rozptylu poslední složky  $b$ .
2. Je-li apriorní rozdělení  $\gamma$  symetrické kolem 0 a má velký (resp. malý) rozptyl, pak optimální zobecněný PTE je pro malá  $|t|$  roven odhadu MNČ v úplném modelu (resp. v redukovaném modelu), tedy množina  $S$  příslušná tomuto zobecněnému PTE neobsahuje (resp. obsahuje) okolí 0.

Kdybychom např. použili pro  $\gamma$  rovnoměrné rozdělení (nebo symetrické useknuté normální rozdělení) se střední hodnotou 0 a velkým rozptylem, mělo by optimální rozhodování (z hlediska minimalizace  $\text{MSE}_W$ ) být pro velký počet stupňů volnosti opačné, než je v praxi zvykem.

Pro nevlastní neurčitou apriorní hustotu  $f(\gamma) = 1$ ,  $\gamma \in R$ , byly tradiční PTE příslušné kritickým hodnotám  $c \in [0, \infty]$  porovnávány v práci [64] a bylo zjištěno, že optimální hodnotou je pak  $c = 0$  odpovídající použití jen odhadu MNČ v úplném modelu. Neurčité apriorní rozdělení parametru  $\gamma$  se však nehodí pro častou intuitivní představu, že parametr  $\gamma$  by mohl být „se značnou pravděpodobností“ blízký k nule, která právě inspirovala praktiky k používání PTE.

Pre-test odhady, byť by byly voleny optimálně, nedávají tak dobré výsledky z hlediska  $\text{MSE}_{X'X}$  jako zkrácené odhady, kterými se budeme zabývat v dalším odstavci. Výhodou PTE však je, že připouští použití neúplného modelu, takže vzorec může mít menší počet členů než při použití odhadů z následujícího odstavce.

## 5. ODHADY STEINOVA TYPU

V tomto odstavci se budeme zabývat dosti obecnou třídou odhadů s výhodnými vlastnostmi, které budeme nazývat zkrácenými či sraženými odhady nebo také odhady Steinova typu. V jistém ohledu jsou příbuzné tzv. hřebenovým odhadům, se kterými se čtenář může seznámit např. v [25] odst. 19.5 a 19.6 nebo v [70] odst. 9.3.

Pokud nebude řečeno jinak, budeme předpokládat, že chybový vektor  $e$  modelu (1.1) má normální rozdělení  $N_n(0, \sigma^2 I_n)$ .

Protože v literatuře byly intenzivně studovány odhady vektoru středních hodnot vícerozměrného normálního rozdělení, objasníme nejprve v následujících poznámkách, jakou mají souvislost s odhady vektoru  $\beta$  modelu (1.1).

**Poznámka 5.1.** Všimněme si, že speciálním případem naší úlohy je odhad střední hodnoty  $\mu \in R^p$  normálního rozdělení  $N_p(\mu, \sigma^2 I_p)$  s neznámou hodnotou  $\sigma^2$ ; máme-li k dispozici náhodný výběr rozsahu  $k > 1$ , můžeme tuto úlohu interpretovat jako úlohu odhadu parametrů modelu (1.1) s neznámým vektorem  $\beta = \mu$  a vhodnou ortogonální maticí  $X$  typu  $(kp) \times p$  takovou, že  $X'X = kI_p$ . Odhad  $\mu$  metodou maximální věrohodnosti přitom odpovídá odhadu vektoru  $\beta$  pomocí MNČ v odpovídajícím modelu (1.1).

**Poznámka 5.2.** Ukážeme, jakým způsobem lze pro získání některých odhadů vektoru  $\beta$  modelu (1.1) využít úlohy odhadu  $\mu$  rozdělení  $N_p(\mu, \sigma^2 I_p)$ .

Nechť  $A$  je konstantní regulární matice typu  $p \times p$ , která nezávisí na neznámých parametrech. Označme

$$x = Ab,$$

kde  $b$  je odhad (3.1). Platí  $x \sim N_p(\mu, \Sigma)$ , kde  $\mu = A\beta$  a  $\Sigma = \sigma^2 A(X'X)^{-1}A'$ . Je-li  $\hat{\mu}$  nějaký odhad  $\mu$  založený na pozorování  $x$ , pak pro odhad  $\hat{\beta}$  vektoru  $\beta$  definovaný vztahem  $\hat{\beta} = A^{-1}\hat{\mu}$  platí

$$\text{MSE}_W(\beta, \hat{\beta}) = \text{MSE}_{\widetilde{W}}(\mu, \hat{\mu}), \quad \text{kde} \quad \widetilde{W} = (A^{-1})'WA^{-1}.$$

Speciálně,  $\text{MSE}_W(\beta, b) = \text{MSE}_{\widetilde{W}}(\mu, x)$ .

Např. pro váhovou matici  $W = X'X$  můžeme volit regulární matici  $A$  tak, že  $A'A = X'X$  (její existenci jsme odůvodnili již v pozn. 3.2). Pak podle části b) pozn. 3.2 je

$$\widetilde{W} = I_p$$

a dále  $A(X'X)^{-1}A' = A(A'A)^{-1}A' = AA^{-1}(A')^{-1}A' = I_p$ , tedy

$$x \sim N_p(\sigma^2 I_p).$$

Výše popsaná transformace umožňuje přenášet výsledky o odhadech vektoru  $\mu$  založené na  $x$  na výsledky (speciálních) odhadů vektoru  $\beta$  modelu (1.1) založených na  $b$ . Nalezneme-li odhad  $\hat{\mu}$  vektoru  $\mu$  založený na pozorování  $x$ , který ve smyslu  $\text{MSE}_{\widetilde{W}}$  (ostře) dominuje odhadu  $x$  (což je maximálně věrohodný odhad  $\mu$  založený na jednom pozorování  $x$ ), pak  $\hat{\beta} = A^{-1}\hat{\mu}$  je odhad  $\beta$ , který ve smyslu  $\text{MSE}_W$  (ostře) dominuje  $b$ . Podobně, je-li  $\hat{\mu} = \hat{\mu}(x)$  nepřipustným odhadem  $\mu$  ve smyslu  $\text{MSE}_{\widetilde{W}}$ , pak  $\hat{\beta} = A^{-1}\hat{\mu}$  je nepřipustným odhadem  $\beta$  ve smyslu  $\text{MSE}_W$ .

Nevíme však, zda přípustný odhad  $\hat{\mu}$  se popsáním způsobem transformuje na přípustný odhad  $\hat{\beta}$ ; víme pouze, že pak  $\hat{\beta}$  není ostře dominován ve smyslu  $\text{MSE}_W$  žádným odhadem ve tvaru  $A^{-1}(\tilde{\mu}(Ab))$ , tedy žádným odhadem ve tvaru  $f(b)$ , kde  $\tilde{\mu}$  a  $f$  jsou funkce z  $R^p$  do  $R^p$ .

Platí však, že je-li matice  $W$  (a tedy i  $\widetilde{W}$ ) symetrická pozitivně definitní a  $\widehat{\mu}$  je minimax odhadem  $\mu$  ve smyslu  $\text{MSE}_{\widetilde{W}}$ , pak  $\widehat{\beta} = A^{-1}\widehat{\mu}$  je minimax odhadem  $\beta$  ve smyslu  $\text{MSE}_W$ , neboť pro libovolný odhad  $\widetilde{\beta}$  vektoru  $\beta$  s využitím věty 3.1 a výše uvedených transformačních vzorců dostáváme

$$\sup_{\beta} L_W(\beta, \widetilde{\beta}) \geq \sup_{\beta} L_W(\beta, b) = \sup_{\mu} L_{\widetilde{W}}(\mu, x) \geq \sup_{\mu} L_{\widetilde{W}}(\mu, \widehat{\mu}) = \sup_{\beta} L_{\widetilde{W}}(\beta, \widehat{\beta}).$$

**Definice 5.3.** (Zjednodušená terminologie za účelem stručnosti dalšího popisu.)

Nechť  $\widetilde{\beta}$  je nějaký odhad  $\beta$  v modelu (1.1). Řekneme, že odhad  $\widehat{\beta}$  vektoru  $\beta$  je lepší než  $\widetilde{\beta}$ , jestliže  $\widehat{\beta}$  ostře dominuje  $\widetilde{\beta}$  ve smyslu  $\text{MSE}_{X'X}$ . Řekneme, že odhad  $\widehat{\beta}$  je přípustný (resp. nepřípustný), je-li přípustný (resp. nepřípustný) ve smyslu  $\text{MSE}_{X'X}$ . Řekneme, že odhad  $\widehat{\beta}$  je minimax odhadem, je-li minimax odhadem ve smyslu  $\text{MSE}_{X'X}$ .

Analogickou terminologii budeme používat pro odhady vektoru  $\mu$  normálního rozdělení  $N_p(\mu, \Sigma)$  s tím, že odhady budou porovnávány ve smyslu  $\text{MSE}_{I_p}$ .

Protože podle věty 3.1 je odhad MNČ vektoru  $\beta$  modelu (1.1) minimax odhadem, bude totéž platit i o každém lepším odhadu.

Co se týče přípustnosti odhadu MNČ vektoru  $\beta$ , není odpověď jednoznačná. V této souvislosti je vhodné si připomenout, že podle poznámky 3.2 se stačí omezit na případ ortogonální regresní matice a jednotkové váhové matice. V literatuře se běžně uvádí tvrzení, že odhad vektoru středních hodnot  $p$ -rozměrného normálního rozdělení metodou maximální věrohodnosti je pro  $p \leq 2$  přípustný (např. [5], str. 256), avšak odtud bezprostředně neplyne přípustnost odhadu MNČ vektoru  $\beta$ , viz poznámka 5.2. Tvrzení o přípustnosti odhadu MNČ v modelu (1.1) pro  $p \leq 2$  je uvedeno bez důkazu v [32], str. 240.

V roce 1956 Stein [61] ukázal, že odhad střední hodnoty normálního rozdělení metodou maximální věrohodnosti je při počtu složek  $p \geq 3$  nepřípustný. Explicitní lepší odhad zkonstruovali James a Stein [31] jednak pro případ nezávislých složek se známým rozptylem, dále pro případ společného neznámého rozptylu nezávislých složek a rovněž pro případ závislých složek, kde rozptyl je odhadnut Wishartovou maticí.

Pokud jde o model (1.1), po využití poznámky 5.2 můžeme pro  $p \geq 3$  psát Jamesův-Steinův odhad (JS odhad) vektoru  $\beta$  ve tvaru

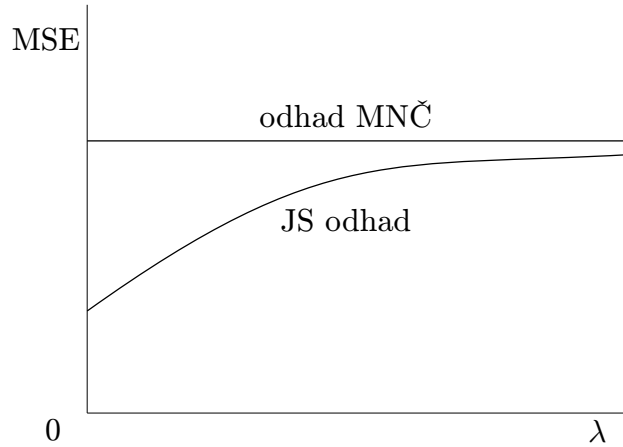
$$(5.1) \quad b^{(JS)} = \left(1 - \frac{c(n-p)s^2}{b'X'Xb}\right)b, \quad \text{kde } 0 < c < \frac{2(p-2)}{n-p+2}$$

a  $s^2$  je standardní nestranný odhad  $\sigma^2$ . JS odhad je tedy vhodným násobkem odhadu MNČ. Optimální konstantou  $c$  (z hlediska  $\text{MSE}_{X'X}$ ) je  $c = (p-2)/(n-p+2)$ . I při optimálním  $c$  je JS odhad nepřípustný, lepším odhadem je např.

$$(5.2) \quad b^{(PPJS)} = \left(1 - \frac{c(n-p)s^2}{b'X'Xb}\right)^+ b,$$

kde  $\lambda^+$  značí  $\max\{\lambda, 0\}$ . Pro tento odhad, který se v anglickém jazyce nazývá *positive-part JS estimator*, použijeme označení PPJS odhad. Rovněž PPJS odhad je však nepřípustný při libovolném  $c$  ([32], [60]) a navíc nelze určit optimální  $c$ , protože rizikové funkce se protínají ([32], str. 211).

Výhodnost použití JS odhadu v porovnání s MNČ roste s počtem regresorů  $p$  a také se snižující se velikostí vektoru  $\beta$ . Pro  $p = 10$  a malé  $\beta$  může být střední kvadratická chyba JS odhadu až 5krát menší než tatáž chyba odhadu MNČ ([13]).



OBRÁZEK 2. Závislost  $\text{MSE}_{X'X}(\beta, \hat{\beta})$  na  $\lambda = \beta'X'X\beta/\sigma^2$  pro odhad metodou nejmenších čtverců a Jamesův-Steinův odhad.

Odhady Steinova typu patří mezi tzv. zkrácené odhady (v angl. *shrinkage estimators*, český název *zkrácené odhady* byl použit v [25]). Lze totiž ukázat, že délka vektoru (5.1) je s velkou pravděpodobností (u PPJS odhadu dokonce vždy) nejvýše rovna délce vektoru  $b$ . Podrobné informace o vývoji odhadů tohoto typu můžeme nalézt např. v monografii [23] a přehledových článcích [29] a [38]. Uveďme v přibližně chronologicky seřazené posloupnosti některé práce věnované zkráceným odhadům. Jde o stručný výběr; např. monografie [23] obsahuje více než 400 referencí, značná část z nich se však vztahuje k případu známého rozptylu nebo k  $p$ -rozměrnému normálnímu rozdělení s určitým předpokladem o kovarianční matici. Na druhé straně mnoho prací o zkrácených odhadech není v seznamu literatury [23] uvedeno.

Ortogonální lineární model s alespoň třemi regresory vyšetřoval Sclove [56] a uvedl tvar JS odhadu pro tento případ. Navrhnul kombinovat odhad MNČ s JS odhadem podle významnosti jednotlivých složek odhadu MNČ. V článku je také zmíněn PPJS odhad.

Baranchik [4] (viz např. Judge a Bock [32]) ukázal, že PPJS odhad je lepší než JS odhad. Efron a Morris [14] a [15] upozorňují, že ačkoli má JS odhad menší střední kvadratickou chybu než odhad metodou maximální věrohodnosti, jednotlivé složky často tuto vlastnost nemají. Navrhují rovněž kombinovat JS odhad s odhadem MNČ.

Sclove, Morris a Radhakrishnan v r. 1972 zkonstruovali pre-test odhad Steinova typu, který je lepší než standardní pre-test odhad (avšak není lepší než MNČ), stručný komentář k tomuto viz [32], str. 190. Zmíněný odhad v obecné podobě lze charakterizovat jako „srážení“ odhadu MNČ směrem k určitému pevnému hypotetickému vektoru regresních koeficientů a obecně tedy nemusí jít o zkracování odhadu MNČ (základní informace o principu srážení k nenulovému vektoru lze najít také v monografiích [41], [33] a [59]). V konkrétním příkladě, který byl vyšetřován simulací, byl zisk pre-test odhadu Steinova typu v porovnání s tradičním pre-test odhadem pro běžně používané hladiny významnosti malý ([32], str. 214).

Baranchik se zabýval JS odhady v kontextu lineární regrese se stochastickou maticí  $X$  a dokázal, že maximálně věrohodný odhad za předpokladu sdružené normality

lze pro  $p \geq 3$  zlepšit, viz [32], kap. 12. Efron [13] udává zkrácený odhad střední hodnoty  $\mu$  jednorozměrného normálního rozdělení  $N(\mu, 1)$  na základě jednoho pozorování  $x$  a tvrdí, že je v Pitmanově smyslu lepší než maximálně věrohodný odhad (tedy než  $x$ ), přičemž za ztrátovou funkci je možno volit libovolnou rostoucí funkci argumentu  $|\hat{\mu} - \mu|$ . Efron své tvrzení nedokazuje. Jeho odhad je uváděn též v [58] a porovnáván s jinými odhady v kontextu lineárního regresního modelu. Efronův vzorec má tvar

$$(5.3) \quad \hat{\mu}^{(E)} = x - h(x),$$

kde  $h(x)$  je lichá funkce, která je pro  $x \geq 0$  definována vztahem

$$h(x) = \frac{1}{2} \min\{x, \Phi(-x)\} \quad (\Phi \text{ je distribuční funkce } N(0, 1)).$$

Bock ukázal, že je-li stopa matice  $(X'X)^{-1}$  menší nebo rovna dvojnásobku největšího vlastního čísla téže matice, pak pro žádnou volbu konstanty  $c$  ve vzorci (5.1) není JS odhad minimax odhadem ve smyslu  $MSE_{I_p}$  a tedy JS odhad nedominuje odhad MNČ ve smyslu  $MSE_{I_p}$ , viz [32], str. 243, další informace k tomuto jsou v [33], str. 923–924.

Odhad, který ostře dominuje odhadu MNČ ve smyslu  $MSE_W$ , kde  $W$  je libovolná (pevně zvolená) symetrická pozitivně definitní matice, lze odvodit z výsledků Bergera. Bergerův odhad, který byl původně odvozen pro normální rozdělení se známým rozptylem, byl upraven do kontextu regresního modelu (1.1) s neznámým  $\sigma^2$  v monografii [32], str. 250–251, a jeho dominance odhadu MNČ byla dokázána tamtéž pro kvadratickou ztrátovou funkci s jednotkovou váhovou maticí. Použijeme-li lineární transformaci podle poznámky 3.2, můžeme pro  $p \geq 3$  sestavit analogický odhad dominující MNČ i pro případ obecné symetrické pozitivně definitní váhové matice  $W$ . Výsledný odhad Steinova typu, který budeme nazývat BS odhadem, má pak tvar

$$(5.4) \quad b^{(BS)} = \left( I_p - \frac{c(n-p)s^2}{b'X'XW^{-1}X'Xb} W^{-1}X'X \right) b, \quad \text{kde } 0 < c < \frac{2(p-2)}{n-p+2}.$$

Všimněme si, že JS odhad obdržíme jako zvláštní případ tohoto odhadu, zvolíme-li váhovou matici  $W = X'X$ . Optimální konstantou  $c$  (z hlediska  $MSE_W$ ) je  $c = (p-2)/(n-p+2)$ . Je zde tedy analogie s JS odhadem. Dále uvažujme BS odhad (5.4) jen pro jednotkovou váhovou matici  $W = I_p$ . Všimněme si, že je-li matice  $X$  ortogonální, pak BS odhad je tvaru  $b^{(BS)} = Db$  pro jistou diagonální matici  $D = \text{diag}\{d_1, \dots, d_p\}$ . „Positive-part variantu“ BS odhadu (PPBS) lze pak definovat jako  $D^+b$ , kde  $D^+ = \text{diag}\{d_1^+, \dots, d_p^+\}$ . V monografii [32] bylo metodou Monte-Carlo porovnáno 9 variant odhadu Steinova typu pro dvě různé ortogonální matice  $X$  s  $p = 5$  regresory. V souladu s teoretickými výsledky všechny uvažované odhady výrazně předčily odhad MNČ, avšak nebylo možné určit jednoznačného vítěze, neboť rizikové funkce (2.2) se v oboru parametrů  $\beta$  často křížily. Odhad PPJS byl jen o málo lepší než JS, avšak odhad PPBS byl podstatně lepší než BS. Zatímco pro případ neznámého rozptylu a nepřiliš velkého  $n$  nebyl mezi odhady PPJS a PPBS velký rozdíl, pro případ známé rozptylové matice bylo možno z těchto dvou odhadů považovat PPBS za značně lepší pro malé  $\|\beta\|$  a jen nepatrně horší pro velká  $\|\beta\|$ .

Pokračujme ve stručném přehledu dalších prací. Efron a Morris [16] prezentují intuitivní výklad JS odhadu. Zkrácené odhady z bayesovského pohledu popisuje Vinod [66] v přehledném článku s rozsáhlou bibliografií. Draper a Van Nostrand [12] poskytují přehled literatury o zkrácených odhadech dostupné v té době, citující 77 odkazů.

Odhad regresních koeficientů Steinova typu byl uplatněn v USA na data o příjmech při neúplném statistickém šetření v r. 1970 ([17]) a bylo ukázáno, že výsledné odhady mají menší průměrnou chybu než tradiční odhady.

Ullah, Srivastava a Chandra [65] uvažují obecnou třídu zkrácených odhadů zahrnující také JS odhad a odvozují podmínky, za kterých jsou lepší než MNČ, když u rozdělení chyb se předpokládá jen konečnost čtvrtého momentu. Mittelhammer [46] a později též Ohtani [48] ukazují, že v případě špatně určeného regresního modelu JS odhad nedominuje MNČ. Judge a kol. [33] objasňují pro odhady Steinova typu řadu různých souvislostí s aktuálními referencemi. Ghosh a kol. [20] v pre-test odhadu kombinují JS odhad a odhad MNČ, viz též [33].

Některé práce se zabývají hodnocením zkrácených odhadů pomocí Pitmanova kritéria (viz def. 2.7). Keating a Czitrom [34] ukázali, že i při porovnání pomocí Pitmanova kritéria se ztrátovou funkcí  $L_{X'X}$  je JS odhad lepší než odhad MNČ. Rovněž Sen, Kubokawa a Saleh [58] se zabývají Steinovým principem v kontextu Pitmanova kritéria a používají kvadratickou ztrátovou funkci s obecnou pozitivně definitní váhovou maticí  $W$ . Z jejich výsledků plyne, že je-li  $p \geq 2$ , pak odhad

$$(5.5) \quad \left( I_p - \frac{as^2}{b'X'XW^{-1}X'Xb} W^{-1}X'X \right) b, \quad \text{kde} \quad 0 \leq a \leq \frac{(p-1)(3p+1)}{2p},$$

dominuje odhadu MNČ v Pitmanově smyslu při ztrátové funkci (2.1). Pozoruhodné je, že výsledek platí již pro  $p = 2$ , takže umožňuje zlepšit odhad MNČ (v Pitmanově smyslu) již pro regresní přímku (připomínáme, že Efronův odhad (5.3) umožňuje při známém parametru  $\sigma^2$  zlepšit v Pitmanově smyslu metodu maximální věrohodnosti také pro  $p = 1$ ). Odhad (5.5) lze ještě dále v uvedeném smyslu zlepšit jeho „positive-part verzi“ ([58]), která má tvar

$$(5.6) \quad \left( I_p - \frac{\min\{as^2, b'X'Xb\}}{b'X'XW^{-1}X'Xb} W^{-1}X'X \right) b.$$

Čtenář si může povšimnout, že pro  $a = c(n-p)$  splývá odhad (5.5) s BS odhadem (5.4) a je-li navíc  $W = X'X$ , splývá odhad (5.5) také s JS odhadem a odhad (5.6) s PPJS odhadem, pokud ovšem konstanty  $a, c$  vyhovují příslušným rozmezím.

Miller [45] se zabývá JS odhady v regresním modelu z pohledu volby vysvětlujících proměnných. Zdůrazňuje, že MNČ často dává příliš velké odhady, s čímž se lze vypořádat pomocí zkrácených odhadů. Carter a kol. [8] docházejí k závěru, že MNČ je užitečnější pro posuzování jednotlivých regresních koeficientů, avšak pro celý vektor je při  $p \geq 3$  užitečnější JS odhad. Fomby a Samanta [18] navrhuji dvě metody předpovědi založené na zkrácených odhadech a metodou Monte-Carlo je porovnávají s jinými metodami. Brandwein a Strawderman [6] zobecňují teorii zkrácených odhadů pro sféricky symetrická rozdělení a uvažují kvadratické a konkávní ztrátové funkce.

Ki [35] se blíže zabývá odhadem Steinova typu, který navrhnul v r. 1986 Georg a který sráží odhad směrem k několika „apriorním podprostorům“ zároveň. Na reálných i simulovaných datech ukazuje, že tento odhad dává obvykle lepší předpovědi než kroková regrese. Guo a Pal [24] definují posloupnost odhadů (níže se na ně budeme odkazovat jako na odhady  $GP_i$ ) se snižující se střední kvadratickou chybou, které jsou lepší než JS odhad a na rozdíl od PPJS odhadu jsou hladké, a dále nehladký odhad, který je ještě lepší. Kritické zhodnocení zkrácených odhadů podává Hill [28]. Hwang a Ullah [30] porovnávají konfidenční množiny pro parametry lineárního regresního modelu se středy rovnými JS odhadu a odhadu MNČ. První varianta vychází z porovnání lépe.

Kim a Hill [36] uvádějí verzi PPJS odhadu v nelineárním modelu s menší střední kvadratickou chybou, než má maximálně věrohodný odhad a pre-test odhad při kvadratické ztrátové funkci, jejíž váhová matice je rovna informační matici. Uvažují též jednotkovou váhovou matici a některé další. Vinod a Srivastava [67] vyšetřují jistou třídu odhadů (double k class estimators) a uvádějí podmínky, kdy jsou lepší než JS odhad vzhledem k  $MSE_{I_p}$  a  $MSE_{X'X}$ ; chyby nemusejí mít normální rozdělení. Výsledky se částečně opírají o simulační porovnání. Giri v knize [22] popisuje odhady Steinova typu a má řadu referencí.

Cellier, Fourdrinier a Robert [9] a Cellier a Fourdrinier [10] prezentují robustní odhad Steinova typu, který je lepší než maximálně věrohodný odhad jak v případě normálního rozdělení, tak i pro širokou třídu nenormálních rozdělení. Obecné vzorce pro momenty parametrických funkcí JS odhadu a PPJS odhadu odvozují Ohtani a Kozumi [49]. Uvádějí postačující podmínku pro to, aby parametrické funkce JS odhadu dominovaly funkcím odvozeným z odhadu MNČ (obecně takové dominance neplatí).

Kubokawa [37] navrhl speciální odhad střední hodnoty  $p$ -rozměrného normálního rozdělení a ukázal, že je lepší než JS odhad. Navíc dokazoval, že jeho odhad je přípustný, důkaz však obsahuje chybu, kterou opravili až Lin a Pal [42]. Pal a Chano [50] srovnávají tento Kubokawův odhad (K odhad) s odhady JS, PPJS, s výše zmíněnými odhady  $GP_i$  a se Strawdermanovým odhadem jak pro normální rozdělení, tak pro vícerozměrné  $t$ -rozdělení chyb. Ukazují, že i pro  $t$ -rozdělení jsou JS odhad a K odhad lepší než metoda maximální věrohodnosti. Pro normální rozdělení vyjadřují rizikové funkce ve tvaru řad a ukazují, že mezi odhady PPJS,  $GP_1$  a K není jasný vítěz, všechny jsou však lepší než JS odhad.

Shao a Strawderman [60] uvedli pro případ známé rozptylové matice odhad střední hodnoty normálního rozdělení, který je lepší než PPJS odhad, a Sugiura a Takagi [62] rozšířili jejich výsledek na případ, kdy rozptylová matice není známa. DasGupta a Sinha [11] pracovali s obecnými ztrátovými funkcemi a navrhli pro střední hodnotu vícerozměrného normálního rozdělení odhady Steinova typu. Např. pro zkrácený odhad navržený pro ztrátovou funkci rovnající se součtu absolutních hodnot odchylek dokázali jeho dominanci nad JS odhadem a maximálně věrohodným odhadem.

Závěrem je možno říci, že odhady Steinova typu prokázaly v mnoha studiích své výhodné vlastnosti a proto nelze pochybovat o tom, že si v budoucnu najdou svoje místo i v praxi.

## LITERATURA

1. ANTOCH, J., VÍŠEK, J. Á., *Robust estimation in linear model and its computational aspects*. In: Computational aspects of model choice (Prague, 1991), str. 39–104, Contrib. Statist., Physica, Heidelberg, 1993.
2. ANTOCH, J., VORLÍČKOVÁ, D., *Vybrané metody statistické analýzy dat*. Academia, Praha, 1992.
3. BANCROFT, T. A., *On biases in estimation due to the use of preliminary tests of significance*. Ann. Math. Statistics 15, (1944), 190–204.
4. BARANCHIK, A. J., *Multiple regression and estimation of the mean of a multivariate normal distribution*. Technical Report 51, Department of Statistics, Stanford University, 1964.
5. BERGER, J. O., *Statistical decision theory and Bayesian analysis*. Second edition, Springer, New York, 1985.
6. BRANDWEIN, A. C., STRAWDERMAN, W. E., *Generalizations of James-Stein estimators under spherical symmetry*. Ann. Statist. 19 (1991), no. 3, 1639–1650.
7. BURNHAM, K. P., ANDERSON, D. R., *Model selection and inference*. Springer, New York-Berlin, 1998.

8. CARTER, R. A. L., SRIVASTAVA, M. S., SRIVASTAVA, V. K., ULLAH, A., *Unbiased estimation of the MSE matrix of Stein-rule estimators, confidence ellipsoids, and hypothesis testing*. Econometric Theory 6 (1990), no. 1, 63–74.
9. CELLIER, D., FOURDRINIER, D., ROBERT, C., *Robust shrinkage estimators of the location parameter for elliptically symmetric distributions*. J. Multivariate Anal. 29 (1989), no. 1, 39–52.
10. CELLIER, D., FOURDRINIER, D., *Shrinkage estimators under spherical symmetry for the general linear model*. J. Multivariate Anal. 52 (1995), no. 2, 338–351.
11. DASGUPTA, A., SINHA, B. K., *A new general interpretation of the Stein estimate and how it adapts: applications*. J. Statist. Plann. Inference 75 (1999), no. 2, 247–268.
12. DRAPER, N. R., VAN NOSTRAND, R. C., *Ridge regression and James-Stein estimation: review and comments*. Technometrics 21 (1979), no. 4, 451–466.
13. EFRON, B., *Biased vs. unbiased estimation*. Adv. Math. 16 (1975), 259–277.
14. EFRON, B., MORRIS, C., *Limiting the risk of Bayes and empirical Bayes estimators. I. The Bayes case*. J. Amer. Statist. Assoc. 66 (1971), 807–815.
15. EFRON, B., MORRIS, C., *Empirical Bayes on vector observations: an extension of Stein's method*. Biometrika 59 (1972), 335–347.
16. EFRON, B., MORRIS, C., *Stein's Paradox in statistics*. Scientific American 236 (1977), 119–127.
17. FAY, R. E. III, HERRIOT, R. A., *Estimates of income for small places: an application of James-Stein procedures to census data*. J. Amer. Statist. Assoc. 74 (1979), no. 366, part 1, 269–277.
18. FOMBY, T. B., SMANTA, S. K., *Application of Stein rules to combination forecasting*. Journal of Business and Economics Statistics 9 (1991), 391–407.
19. FRIEŠL, M., REIF, J., *Pitmanovo porovnání odhadů*. Robust 2002, sborník dvanácté zimní školy JČMF (J. Antoch, G. Dohnal a J. Klaschka eds.), JČMF, Praha, 2002, str. 117–125.
20. GHOSH, M., SALEH, A. K. MD. E., SEN, P. K., *Empirical Bayes subset estimation in regression models*. Statist. Decisions 7 (1989), no. 1-2, 15–35.
21. GILES, J. A., D. E. A. GILES, *Pre-test estimation and testing in econometrics: recent developments*. Journal of Economic Surveys 7 (1993), 145–197.
22. GIRI, N. C., *Multivariate statistical analysis*. Marcel Dekker, Inc., New York, 1996.
23. GRUBER, M. H. J., *Improving efficiency by shrinkage*. Marcel Dekker, New York, 1998.
24. GUO, Y. Y., PAL, N., *A sequence of improvements over the James-Stein estimator*. J. Multivariate Anal. 42 (1992), no. 2, 302–317.
25. HEBÁK, P., HUSTOPECKÝ, J., *Vícerozměrné statistické metody s aplikacemi*. SNTL/Alfa, Praha, 1987.
26. HEBÁK, P., *Regrese. I. část*, VŠE, Praha, 1998.
27. HEBÁK, P., SVOBODOVÁ, A., *Regrese. II. část*, VŠE, Praha, 2001.
28. HILL, B. M., *On Steinian shrinkage estimators: the finite/infinite problem and formalism in probability and statistics*. Aspects of uncertainty (Freeman, R. P., Smith, A. F. M., eds.) Wiley, Chichester, 1994, pp. 223–260.
29. HOFFMANN, K., *Stein estimation—a review*. Statist. Papers 41 (2000), no. 2, 127–158.
30. HWANG, J. T. G., ULLAH, A., *Confidence sets centered at James-Stein estimators: a surprise concerning the unknown-variance case*. J. Econometrics 60 (1994), no. 1-2, 145–156.
31. JAMES, W., STEIN, C., *Estimation with quadratic loss*. Proc. 4th Berkeley Sympos. Math. Statist. and Prob., Vol. I, pp. 361–379, Univ. California Press, Berkeley, Calif., 1961.
32. JUDGE, G. G., BOCK, M. E., *The statistical implications of pre-test and Stein-rule estimators in econometrics*. North-Holland Publishing Co., Amsterdam, 1978.
33. JUDGE, G. G., GRIFFITHS, W. E., HILL, R. C., LEE, T. C., *The theory and practice of econometrics*. Wiley, New York, 2. vyd., 1985.
34. KEATING, J. P., CZITROM, V., *A comparison of James-Stein regression with least squares in the Pitman nearness sense*. The Journal of Computation and Simulation 34 (1989), 1–9.
35. KI, Y. C. F., *Multiple shrinkage estimators in multiple linear regression*. Comm. Statist. Theory Methods 21 (1992), no. 1, 111–136.
36. KIM, M., HILL, R. C., *Shrinkage estimation in nonlinear regression: the Box-Cox transformation*. J. Econometrics 66 (1995), no. 1-2, 1–33.
37. KUBOKAWA, T., *An approach to improving the James-Stein estimator*. J. Multivariate Anal. 36 (1991), no. 1, 121–126.
38. KUBOKAWA, T., *The Stein phenomenon in simultaneous estimation: a review*. In: Applied statistical science, III, pp. 143–173, Nova Sci. Publ., Commack, NY, 1998.
39. KUBOKAWA, T., *Shrinkage and modification techniques in estimation of variance and the related problems: a review*. Comm. Statist. Theory Methods 28 (1999), no. 3-4, 613–650.

40. LAWSON, C. L., HANSON, R. J., *Solving least squares problems*. Prentice-Hall, Inc., Englewood Cliffs, New Jersey, 1974.
41. LEHMANN, E. L., *Theory of point estimation*. Wiley, New York, 1983.
42. LIN, J. J., PAL, N., *An admissibility result on estimation of a multivariate normal mean vector*. Tamsui Oxf. J. Math. Sci. 17 (2001), no. 1, 23–33.
43. MAGNUS, J. R., *The traditional pretest estimator*. Theory of Probability and Its Applications 44 (1999), 293–308.
44. MCQUARRIE, A. D. R., TSAI, C.-L., *Regression and time series model selection*. World Scientific Publishing Co., Inc., River Edge, NJ, 1998.
45. MILLER, A. J., *Subset selection in regression*. Chapman and Hall, London, 1990.
46. MITTELHAMMER, R. C., *Restricted least squares, pretest, OLS and Stein rule estimators: risk comparisons under model misspecification*. J. Econometrics 25 (1984), no. 1-2, 151–164.
47. MYERS, R. M., *Classical and modern regression with application*. 2. vyd., PWS-Kent Publish. Co., Boston, 1990.
48. OHTANI, K., *A comparison of the Stein-rule and positive-part Stein-rule estimators in a misspecified linear regression model*. Econometric Theory 9 (1993), no. 4, 668–679.
49. OHTANI, K., KOZUMI, H., *The exact general formulae for the moments and the MSE dominance of the Stein-rule and positive-part Stein-rule estimators*. J. Econometrics 74 (1996), no. 2, 273–287.
50. PAL, N., CHANO, C. H., *Risk analysis and robustness of four shrinkage estimators*. Calcutta Statist. Assoc. Bull. 46 (1996), 35–62.
51. RAO, C. R., *Lineární metody statistické indukce a její aplikace*. Academia, Praha, 1978.
52. REIF, J., VLČEK, K., *Optimal pre-test estimators in regression*, J. Econometrics 110 (2002), 91–102.
53. ROEHRIG, C. S., *Optimal critical regions for pre-test estimators using a Bayes risk criterion*. Journal of Econometrics 25 (1984), 3–14.
54. ROUSSEEUW, P. J., LEROY, A. M., *Robust regression and outlier detection*. Applied Probability and Statistics. John Wiley, New York, 1987.
55. SAWA, T., HIROMATSU, T., *Minimax regret significance points for a preliminary test in regression analysis*. Econometrica 41 (1973), 1093–1101.
56. SCLOVE, S. L., *Improved estimators for coefficients in linear regression*. J. Amer. Statist. Assoc. 63 (1968), 596–606.
57. SCLOVE, S. L., MORRIS, C., RADHAKRISHNAN, R., *Non-optimality of preliminary-test estimators for the mean of a multivariate normal distribution*. Ann. Math. Statist. 43 (1972), 1481–1490.
58. SEN, P. K., KUBOKAWA, T., SALEH, A. K. MD. E., *The Stein paradox in the sense of the Pitman measure of closeness*. Ann. Statist. 17 (1989), no. 3, 1375–1386.
59. SHAO, J., *Mathematical statistics*. Springer, New York - Berlin, 1998.
60. SHAO, P. Y.-S., STRAWDERMAN, W. E., *Improving on the James-Stein positive-part estimator*. Ann. Statist. 22 (1994), no. 3, 1517–1538.
61. STEIN, C., *Inadmissibility of the usual estimator for the mean of a multivariate normal distribution*. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1954–1955, vol. I, pp. 197–206. Univ. of California Press, Berkeley, 1956.
62. SUGIURA, N., TAKAGI, Y., *Dominating James-Stein positive-part estimator for normal mean with unknown covariance matrix*. Comm. Statist. Theory Methods 25 (1996), no. 12, 2875–2900.
63. TOYODA, T., OHTANI, K., *Optimal pre-testing procedure in regression*. Keizai Kenkyu, 29 (1978), 39–43.
64. TOYODA, T., WALLACE, T. D., *Optimal critical values for pre-testing in regression*. Econometrica 44 (1976), 365–375.
65. ULLAH, A., SRIVASTAVA, V. K., CHANDRA, R., *Properties of shrinkage estimators in linear regression when disturbances are not normal*. J. Econometrics 21 (1983), no. 3, 389–402.
66. VINOD, H. D., *A survey of ridge regression and related techniques for improvements over ordinary least squares*. Rev. Econom. Statist. 60 (1978), no. 1, 121–131.
67. VINOD, H. D., SRIVASTAVA, V. K., *Large sample asymptotic properties of the double k-class estimators in linear regression models*. Econometric Rev. 14 (1995), no. 1, 75–100.
68. VÍŠEK, J. Á., *Selecting regression model*. Probab. Math. Statist. 21 (2001), no. 2, 467–492.
69. WAN, A. T. K., *The non-optimality of interval restricted and pre-test estimators under squared error loss*. Communications in Statistics – Theory and Methods 23 no. 8 (1994), 2231–2252.
70. ZVÁRA, K., *Regresní analýza*. Academia, Praha, 1989.