ROBUST'2002, 315 – 320 © JČMF 2002

VYUŽITÍ STATISTICKÝCH METOD V BANKOVNICTVÍ

FILIP TROJAN

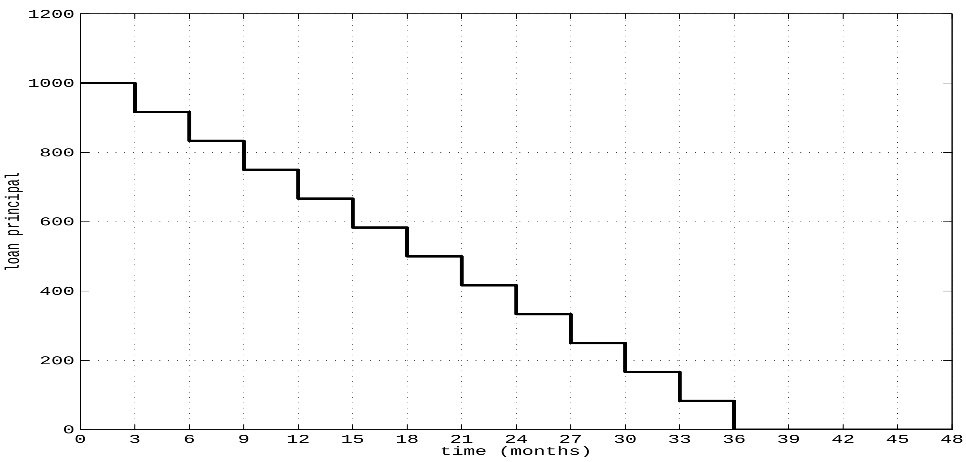
Abstract. Goal of this contribution is to show how statistical methods and concepts are used in Komerční banka, Prague.
Резюме. Цель этой приложительной работы есть показать как в Комерционном банке в Праге используются методы математической статистики.

1. Úvod

Záměrem tohoto příspěvku není rozvíjet teorii matematické statistiky, ale ukázat, jak se jednotlivé metody a koncepty dají uplatnit v oblasti bankovnictví. Tyto poznatky vycházejí z 3 leté praxe v oddělení skóringu a správy portfolia v Komeční bance, a.s.

2. Odhad výše rizikových nákladů

t	spojitý čas, 0 = poskytnutí úvěru.
$\xi(t)$	nesplacená jistina v čase t (náh. fce).
$i(t)$	úroková sazba pro výpočet PV.
$C(t)$	tržní cena zajištění úvěru (náh. fce).
T	doba do defaultu ($T \geq 0$ náh. veličina).



Obr. 1 *Typický tvar funkce* $\xi(t)$.

Definujme $j(t) = \int_0^t i(s)\,\mathrm{d}\,s$. Ztráta banky při defaultu klienta je $e^{-j(T)} \max(\xi(T) - C(T), 0)$. Výběr z rizikové přirážky ρ je $\int_0^T e^{-j(t)} \rho\, \xi(t)\,\mathrm{d}\,t$. Celková bilance

2000 *Mathematics Subject Classification.* Primary 62P20.
Klíčová slova. Bankovnictví, riziko, kredit skóring, správa portfolia.

$$B = \int_0^T e^{-j(t)} \rho\, \xi(t)\, \mathrm{d}\, t - e^{-j(T)} \max(\xi(T) - C(T), 0). \tag{1}$$

Úkolem je stanovit číslo ρ tak, aby

(a) $\mathrm{E}\, B = 0$
(b) $\mathrm{P}[B < 0] \leq \alpha$
(c) $\mathrm{E}\, B - \beta \operatorname{var} B = 0$
(d) $\mathrm{E}\, B - \gamma\sqrt{\operatorname{var} B} = 0$

Dále uvažujeme pouze variantu (a). První zjednodušení je, položíme-li $i(t) = 0 \Rightarrow j(t) = 0$, což je konzervativní odhad. Potom

$$\mathrm{E}\, B = \mathrm{E} \int_0^T \rho\, \xi(t)\, \mathrm{d}\, t - \mathrm{E} \max(\xi(T) - C(T), 0) = 0. \tag{2}$$

Chceme-li oddělit vliv zajištění, položíme $C(T) \approx \alpha\xi(0)$, kde α bude odhadnuto později. Zároveň zavedeme $\rho_\alpha(k)$ – k-leté rizikové náklady při míře zajištění α:

$$\rho_\alpha(k) = \frac{\mathrm{E} \max(\xi_k(T) - \alpha\xi_k(0), 0)}{\mathrm{E} \int_0^T \xi_k(t)\, \mathrm{d}\, t}, \tag{3}$$

kde

$$\xi_k(t) = \begin{cases} \xi(t) & 0 \leq t < k \\ 0 & t \geq k. \end{cases} \tag{4}$$

Například pro jednorázově splatné úvěry ($\xi(t) = I(t \in [0, \kappa))\xi(0)$) platí

$$\rho_\alpha(k) = (1 - \alpha)^2 \rho_0(k). \tag{5}$$

Stačí proto hledat pouze $\rho_0(k)$. Označíme-li $\Xi_k(t) = \int_0^t \xi_k(\tau)\, \mathrm{d}\, \tau$, pak

$$\rho_0(k) = \frac{\mathrm{E}\, \xi_k(T)}{\mathrm{E} \int_0^T \xi_k(t)\, \mathrm{d}\, t} = \frac{\int_0^\infty f(t)\xi_k(t)\, \mathrm{d}\, t}{\int_0^\infty f(t)\Xi_k(t)\, \mathrm{d}\, t} = \frac{\int_0^k f(t)\xi(t)\, \mathrm{d}\, t}{\int_0^k f(t)\Xi(t)\, \mathrm{d}\, t}. \tag{6}$$

Zavedeme-li „survival function“ $\Lambda(t) = \mathrm{P}[T > t] = 1 - F(t)$ a „hazard function“

$$\lambda(t) = \lim_{\Delta \to 0} \frac{\mathrm{P}[t < T < t + \Delta | T > t]}{\Delta} = \frac{f(t)}{\Lambda(t)}, \tag{7}$$

známé z teorie spolehlivosti. Platí $\Lambda(t) = e^{-\int_0^t \lambda(\tau)\, \mathrm{d}\, \tau}$. Pak

$$\begin{aligned} \rho_0(k) &= \frac{\int_0^k \lambda(t)\Lambda(t)\xi(t)\, \mathrm{d}\, t}{\int_0^\infty f(t)\Xi_k(t)\, \mathrm{d}\, t} = \frac{\int_0^k \lambda(t)\Lambda(t)\xi(t)\, \mathrm{d}\, t}{F(t)\Xi_k(t)|_0^\infty - \int_0^\infty F(t)\xi_k(t)\, \mathrm{d}\, t} = \\ &= \frac{\int_0^k \lambda(t)\Lambda(t)\xi(t)\, \mathrm{d}\, t}{\Xi_k(\infty) - \int_0^k F(t)\xi(t)\, \mathrm{d}\, t} = \frac{\int_0^k \lambda(t)\Lambda(t)\xi(t)\, \mathrm{d}\, t}{\int_0^k \xi(t)\, \mathrm{d}\, t - \int_0^k F(t)\xi(t)\, \mathrm{d}\, t} = \\ &= \frac{\int_0^k \lambda(t)\xi(t)\Lambda(t)\, \mathrm{d}\, t}{\int_0^k \xi(t)\Lambda(t)\, \mathrm{d}\, t}. \end{aligned}$$

Rizikové náklady tedy závisí na rozdělení náhodné veličiny T a na tvaru funkce ξ. Bez újmy na obecnosti můžeme předpokládat $\xi(0) = 1$. Potom je počet funkcí ξ konečný a pro každou z nich spočteme rizikové náklady zvlášť. Zbývá odhadnout rozdělení T. Pro odhad jsou k dispozici pouze cenzorovaná data (omezená délka časových řad, období bez úvěrového zatížení klienta). Pro velká k nejsou data vůbec. Modely rozdělení náhodné veličiny T:

- Exponenciální rozdělení ($\rho_0(k) = \lambda$)
- Gamma
- Logaritmicko normální
- Weibull
- Zobecněné gamma
- Neparametrické (Kaplan-Meierův odhad)

Nevýhoda této metody – není postihnuta stochastická závislost mezi T a ξ. Jiný přístup k odhadu rizikových nákladů tuto nevýhodu nemá. Spočívá v následující úpravě:

$$\rho_0(k) = \frac{\mathrm{P}[T < k]\,\mathrm{E}[\xi_k(T)|T < k]}{\mathrm{E}\int_0^T \xi_k(t)\,\mathrm{d}\,t} = \mathrm{P}[T < k]\frac{\mathrm{E}[\xi(T)|T < k]}{\mathrm{E}\int_0^{\min(T,k)} \xi(t)\,\mathrm{d}\,t} = p\frac{\mathrm{E}\,L}{\mathrm{E}\,I}, \tag{8}$$

nebo

$$\rho_0(k) = \mathrm{P}[T < k, \xi(T) > 0]\frac{\mathrm{E}[\xi(T)|T < k, \xi(T) > 0]}{\mathrm{E}\int_0^{\min(T,k)} \xi(t)\,\mathrm{d}\,t} = p'\frac{\mathrm{E}\,L'}{\mathrm{E}\,I'}. \tag{9}$$

Podle přístupu (8) se p odhadne pomocí Kaplan-Meierova odhadu a $\mathrm{E}\,L, \mathrm{E}\,I$ se odhadly některou z robustních metod (trimmed mean), nebo se odhadlo celé rozdělení L a I (LogNormal, Gamma, Zobecněné Gamma, Pareto, Gompertz) pomocí některé robustní metody (D-odhady s Kolmogorov-Smirnov, Hellinger, M-odhady nebo L-odhady).

Podle přístupu (9) je to totéž, pouze p stačí odhadnout jednoduše jako relativní četnost defaultů v cenzorovaných datech.

Poslední přístup je založen opět na bilanční rovnici

$$B = \rho_0(k)V \;- Z. \tag{10}$$

Jsou-li $i = 1, \dots, n$ indexy úvěrů a $j = 1, \dots, p$ indexy časových oken délky k (mohou se překrývat), pak rizikové náklady stanovíme jako

$$\rho_0(k) = \arg\min_{\rho} \sum_{j=1}^{p} w_j \sum_{i=1}^{n} (\rho V_{ij} - Z_{ij})^2. \tag{11}$$

Řešení této kvadratické optimalizace lze spočíst exaktně a doprovodit manažerskými grafy výběrů a ztrát.

3. Odhady skóringových funkcí

Skóringová funkce je zobrazení $f : X \to [0, 1]$ z množiny charakteristik klienta, které v jistém smyslu co nejlépe separuje dobré a špatné klienty. Prvek z množiny X může být

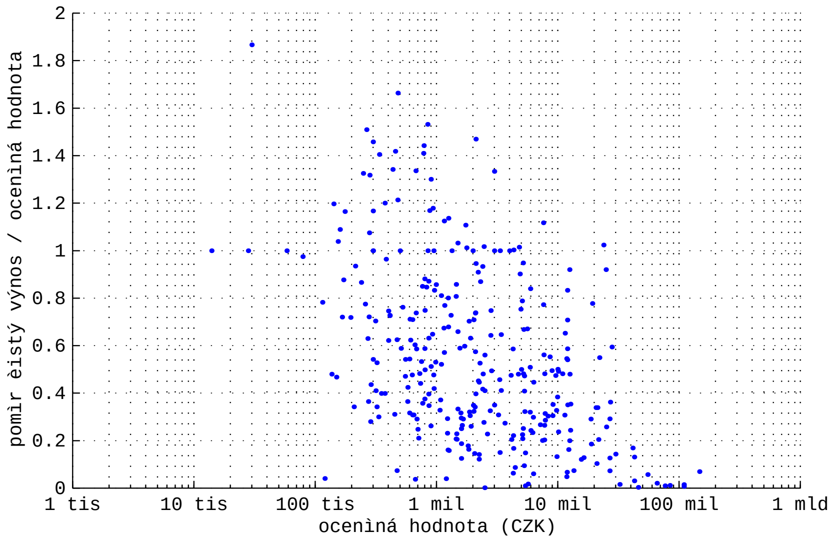
- Účetních výkaz (podniky)
- Osobních údaje z vyplněného formuláře (fyzické osoby)
- Transakční historie (behavioral scoring)

Data pro vývoj funkce jsou pozorování náhodného vektoru $x = \phi(a), s \;\in X$ a náhodné veličiny $y \in \{0, 1\}$ (dobrý/špatný klient). Používané metody jsou lineární diskriminační analýza, logistická regrese, rozhodovací (klasifikační) stromy, neuronové sítě atd.

4. Zajištění úvěru

Zajištění úvěru jsou věci dané do zástavy bance pro případ defaultu. Bývají to nemovitosti, věci movité, ručení (fyzickou osobou) nebo záruka (právnické osoby), cenné papíry (akcie, směnky), pohledávky z obchodního styku nebo tzv. jistota (zástava hotových peněz). Nechť X je odhadní cena zajištění v okamžiku uzavření úvěru a Y je tržní cena zajištění typu v době prodeje zajištění někdy v budoucnosti, nastane-li default (případně diskontovaná hodnota). Při zaúčtování se vedle oceněné hodnoty eviduje ještě tzv. výpůjční hodnota zajištění μX, kde μ je tzv. míra realizace. Úkolem je stanovit míru realizace μ tak, aby platilo

$$\mathrm{E}\,\mu X = \mathrm{E}\,Y \Leftrightarrow \mu = \frac{\mathrm{E}\,Y}{\mathrm{E}\,X}. \tag{12}$$



My se však nespokojujeme s pouhým odhadem $\mathrm{E}\,Y$ a $\mathrm{E}\,X$, ale zkoumáme podmíněné rozdělení $f(y|x)$. Uvažujeme regresní model

$$\mathrm{E}[Y|X] = \nu X^{\gamma}, \tag{13}$$

$$\mathrm{var}[Y|X] = \sigma^2 \max(c + X^{\theta}, \varepsilon). \tag{14}$$

Pro známé $\gamma, c, \theta, \varepsilon$ se jedná o lineární model, který odhadujeme metodou zobecněných nejmenších čtverců. Pro neznámé γ jde o nelineární model, který odhadujeme numerickými metodami. Označme standardizovaná rezidua

$$U = \frac{Y - \nu x^{\gamma}}{\max(c + x^{\theta}, \varepsilon)} \tag{15}$$

Je-li model správně specifikován, pak platí

$$\mathrm{E}\,U = 0 \tag{16}$$

$$\mathrm{var}\,U = \sigma^2. \tag{17}$$

Vztah (16) ověřujeme pomocí t-testu, vztah (17) pomocí Goldfeldt-Quandtova testu. Problém. Existuje c, θ, ε tak, aby oba testy byly splněny? Odpověď: záleží na datech. Pro některá data toho nelze dosáhnout. Někdy pomůže vyjmutí odlehlých pozorování. Zkoušeli jsme i některé robustní metody odhadu, jako například regresní kvantily. Ty se ale nemají používat pro modely bez absolutního členu.

5. Odhad tržní hodnoty úvěrového portfolia

$k = 1, \dots, K$ .... klienti
$\xi_k(t)$ .... nesplacená jistina klienta k
T_k .... doba do defaultu klienta k
U_k .... výtěžnost z dluhu při defaultu
$r = 1, \dots, R$ .... riziková kategorie
$R_k(t)$ .... riziková kategorie klienta k v čase t
$\rho_k(t)$ .... riziková přirážka klienta k v rizikové kategorii R_k stanovená v čase H pro splátku v čase t.
$m = 1, \dots, M$ .... měny, CZK přiřadíme $m = 1$
$t > 0$ .... spojitý čas
$v_m(t)$ .... bezriziková výnosová křivka měny m, tj. úroková sazba, za kterou je možno uložit měnu m na období $[0, t]$.
H .... predikční horizont (1 rok)

Forwardová úroková sazba na období $[\tau_1, \tau_2]$:

$$v_m(\tau_1, \tau_2) = \frac{(1 + v_m(\tau_2))^{\tau_2/(\tau_2-\tau_1)}}{(1 + v_m(\tau_1))^{\tau_1/(\tau_2-\tau_1)}} - 1$$

Pro simulace $i = 1, \dots, N$:
Riziková kategorie klienta v čase H:
Podle modelu Markovského řetězce s korelacemi (viz dále).
Výtěžnost při defaultu:
Podle beta rozdělení s parametry závisejícími na parametrech zajištění.
Hodnota částek splacených v intervalu $[0, H]$ v čase H:

$$V_i^{(1)} = \sum_{k=1}^{K} \int_0^{\min(T_k, H)} -(1 + v_1(t, H))^{H-t} \, \mathrm{d}\, \xi(t)$$

Hodnota realizovaných zástav v čase H:

$$V_i^{(2)} = \sum_{T_k \le H} U_k \xi_k(T_k)(1 + v_1(T_k, H))^{H-T_k}$$

Hodnota částek splacených v období $[H, \infty]$:

$$V_i^{(3)} = \sum_{T_k > H} \int_H^{T_k} -(1 + v_1(H, t) + \rho_k(t))^{H-t} \, \mathrm{d}\, \xi(t)$$

Celková hodnota portfolia v čase H:

$$V_i = V_i^{(1)} + V_i^{(2)} + V_i^{(3)}.$$

Markovský model:

$$\mathrm{P}[R_k(t + H) = j | R_k(\tau) = f(\tau), \tau \le t] = \\ = \mathrm{P}[R_k(t + H) = j | R_k(t) = f(t)].$$

Matice pravděpodobností přechodu:

$$\mathrm{P}[R_k(t + H) = j | R_k(t) = i] = P_{ij}.$$

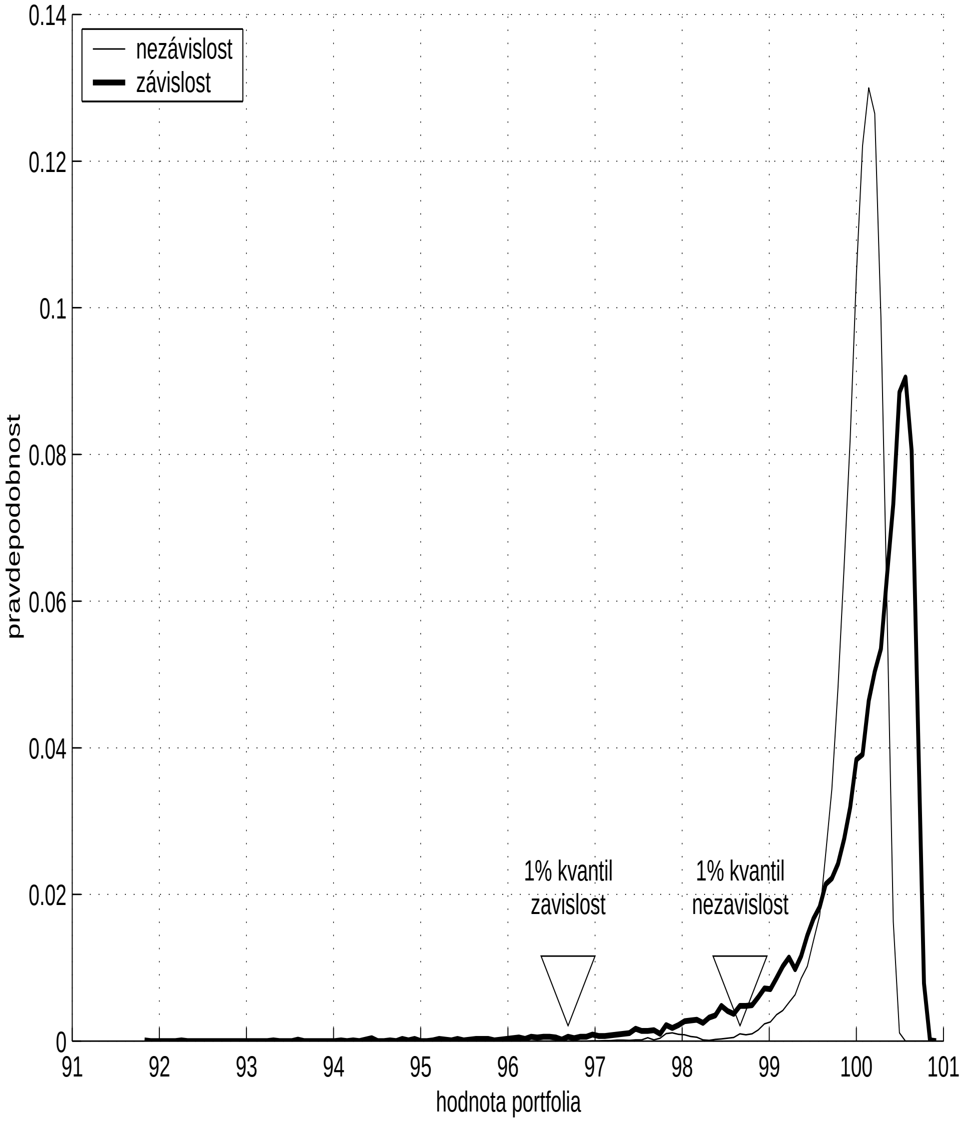
Korelace mezi klienty:

- Korelace mezi regiony
- Korelace mezi odvětvími
- Korelace mezi ESSK

$w_{kz}^{(1)}$ .... příslušnost klienta k do regionu z
$w_{ks}^{(2)}$ .... příslušnost klienta k do odvětví s
$w_{ki}^{(3)}$ .... příslušnost klienta k do ESSK i
$\rho^{(1)}(z_1, z_2)$ .... korelace mezi regiony z_1 a z_2
$\rho^{(2)}(s_1, s_2)$ .... korelace mezi odvětvími s_1 a s_2
$\rho^{(3)}(i_1, i_2)$ .... korelace mezi ESSK i_1 a i_2

Tímto postupem obdržíme náhodný výběr V_i,
$i = 1, \ldots, N$. Z něho spočítáme průměr $\bar{V}$ a kvantily $V(\gamma)$. Neočekávané riziko $\mathsf{NR}(\gamma)$ na hladině $100\gamma\%$ je definováno jako

$$\mathsf{NR}(\gamma) = \bar{V} - V(1-\gamma).$$



Oddělení skóringu a správy portfolia, Komerční banka a.s., Praha