SLUČOVÁNÍ INTERNÍ A EXTERNÍ DATOVÉ ZÁKLADNY PŘI KALKULACI OPERAČNÍHO RIZIKA

Havlický Jiří

ABSTRACT

The aim of this paper is to describe and verify a statistical method of pooling of internal and external data for operational risk calculation within Loss Distribution Approach. The paper deals with three model situations where different concepts of external data collection threshold (constant and known, constant but unknown and stochastic) are considered.

ABSTRAKT

Cílem tohoto článku je popsat a ověřit statistickou metodu slučování interní a externí datové základy pro výpočet operačního rizika v rámci přístupu distribuce ztrát. Článek se zabývá třemi modelovými situacemi, které se liší pojetím prahu sběru externích dat, který je postupně považován za konstantní známý, konstantní neznámý a stochastický.

Úvod

Výpočet výše operačního rizika v rámci přístupu distribuce ztrát pomocí interní datové základny s nízkým počtem pozorování může pro vybrané typy událostí vést k výsledkům, které nedostatečným způsobem postihují podstupované riziko. Je proto potřeba posílit interní datovou základnu relevantními externími daty, která však bývají zpravidla významně zešikmená směrem k velkým ztrátám.

Cílem tohoto článku je popsat a ověřit statistickou metodu, která slučuje interní a externí datovou základnu tak, aby při slučování dat nedošlo ke zkreslení získaných výsledků.

1. Modelování operačního rizika v rámci přístupu distribuce ztrát

V rámci přístupu distribuce ztrát (LDA) je celková ztráta z operačního rizika obchodní linie *i* a typu události *j* za zvolený časový interval dána následujícím vztahem

$$x(i,j) = \sum_{n=1}^{N(i,j)} \zeta_n(i,j), \tag{1}$$

kde ζ představuje náhodnou veličinu popisující výši ztráty z jedné události operačního rizika a *N* představuje náhodnou veličinu popisující počet událostí (ztrát) za zvolené časové období.

Cílem LDA přístupu je získat distribuční funkci $F_{i,j}$ náhodné veličiny $x(i,j)$, která je dána následujícím vztahem, viz Panjer (2006), kde byly pro větší čitelnost zápisu vynechány indexy *i* a *j*

$$F(x)=\begin{cases} p(0)+\sum_{n=1}^{\infty} p(n)S^{n*}(x) & x>0 \\ p(0) & \text{pro} \quad x=0 \end{cases}, \tag{2}$$

kde p představuje pravděpodobnostní funkci náhodné veličiny N, S představuje distribuční funkci náhodné veličiny ζ a $*$ značí operátor konvoluce distribučních funkcí, kdy S^{n*} představuje n-tou konvoluci S se sebou sama, tedy $S^{n*}=S^{n-1}*S$.

Pro výpočet distribuční funkce $F(x)$ je potřeba zvolit odpovídající rozdělení pravděpodobnosti náhodných veličin ζ a N a provést odhad jejich parametrů na základě interních a případně také externích dat popisujících reálně napozorované události operačního rizika.

V dalším textu se předpokládá, že je znám typ skutečného rozdělení pravděpodobnosti náhodných veličin ζ a N a slučování interní a externí datové základny je prováděno za účelem posílení a zpřesnění odhadů parametrů rozdělení pravděpodobnosti náhodné veličiny ζ.

2. Odhad parametrů rozdělení pravděpodobnosti výše ztráty při sloučení interní a externí datové základny

Při aplikaci níže popsané metodologie slučování interní a externí datové základny popisující ztráty z událostí operačního rizika se předpokládá, že

1) externí a interní ztráty vycházejí ze shodného rozdělení pravděpodobnosti,
2) externí ztráty jsou zaznamenány pouze v případě, že výše ztráty překročí stanovený práh *H*, přičemž stanovený práh výše ztráty externích dat může být pro uživatele databáze:
   a) deterministický (konstantní) a známý,
   b) deterministický a neznámý,
   c) stochastický,
3) interní ztráty jsou zaznamenávány bez aplikace prahu *H.*

Dále se zavádí nad rámec kapitoly 1 následující značení:

1) interní ztráty jsou značeny $\zeta_{1,\dots,n}$, kde n představuje počet zaznamenaných ztrát a externí ztráty jsou analogicky značeny $\zeta^*{}_{1,\dots,n^*}$,
2) funkce hustoty zaznamenávaných interních ztrát je značena $s(\zeta;\theta)$ a analogicky funkce hustoty zaznamenávaných externích ztrát je značena $s^*(\zeta^*;\theta)$, kde θ představuje vektor parametrů rozdělení pravděpodobnosti.

Následující kapitoly popisují metodologii sloučení interní a externí datové základny pro výše uvedené tři varianty prahu *H* (konstantní a známý, konstantní a neznámý, stochastický) s praktickým ověřením, že provedené sloučení datových základen nezpůsobuje zkreslení získaných výsledků.

2.1. Konstantní a známý práh sběru externích dat

Předpokládejme nyní, že externí ztráty jsou zaznamenány pouze v případě, že překročí stanovený práh výše ztráty *H*, který je považován za konstantní a jeho hodnota je známa. Z důvodu vyloučení externích ztrát menších než *H* lze konstatovat, že $s(\zeta;\theta)$ se liší od $s^*(\zeta^*;\theta)$ a platí následující vztah, blíže viz Frachot (2002):

$$s^*(\zeta^*;\theta)=1\{\zeta\geq H\}\cdot\frac{s(\zeta;\theta)}{\int_H^\infty s(\zeta;\theta)}=1\{\zeta\geq H\}\cdot\frac{s(\zeta;\theta)}{1-S(H;\theta)}, \tag{3}$$

kde $1\{\zeta\geq H\}$ zobrazuje pomocnou proměnnou, která nabývá hodnoty 1 v případě, že platí, že $\zeta\geq H$, v opačném případě je rovna nule. Platnost uvedeného vztahu vyplývá mimo jiné i z té skutečnosti, že $s(\zeta;\theta)$ dle uvedených předpokladů odpovídá také podkladové funkci hustoty externích dat před aplikací prahu sběru dat.

Odhad parametrů $\hat{\theta}$ rozdělení pravděpodobnosti výše ztráty se současným využitím interních a externích dat lze pak provést metodou maximální věrohodnosti, kdy věrohodnostní funkce má následující tvar

$$l(\theta)=\sum_{i=1}^{n}\ln s(\zeta_i;\theta)+\sum_{i=1}^{n^*}\ln s^*(\zeta_i^*;\theta). \tag{4}$$

Pro ověření uvedeného vztahu a ukázku vlivu prahu sběru externích dat *H* na výši odhadovaných parametrů rozdělení $\hat{\theta}$ byla provedena simulace Monte Carlo. Z lognormálního rozdělení $LN(9,2)$[1] byly náhodně generovány dvě množiny dat o 1000 hodnotách. První množina reprezentuje databázi interně napozorovaných dat $\zeta_{1,\dots,1000}$ a druhá množina reprezentuje databázi externě napozorovaných dat $\zeta^*_{1,\dots,n^*}$, kde byl navíc aplikován práh sběru dat *H* ve výši 2000 p.j. (ztráty nižší než 2000 byly z této množiny vyloučeny, $n^*\leq 1000$). Následně byl metodou maximální věrohodnosti proveden výpočet odhadu parametru rozdělení $\hat{\theta}=(\hat{\mu},\hat{\sigma})$ pro následující čtyři varianty:

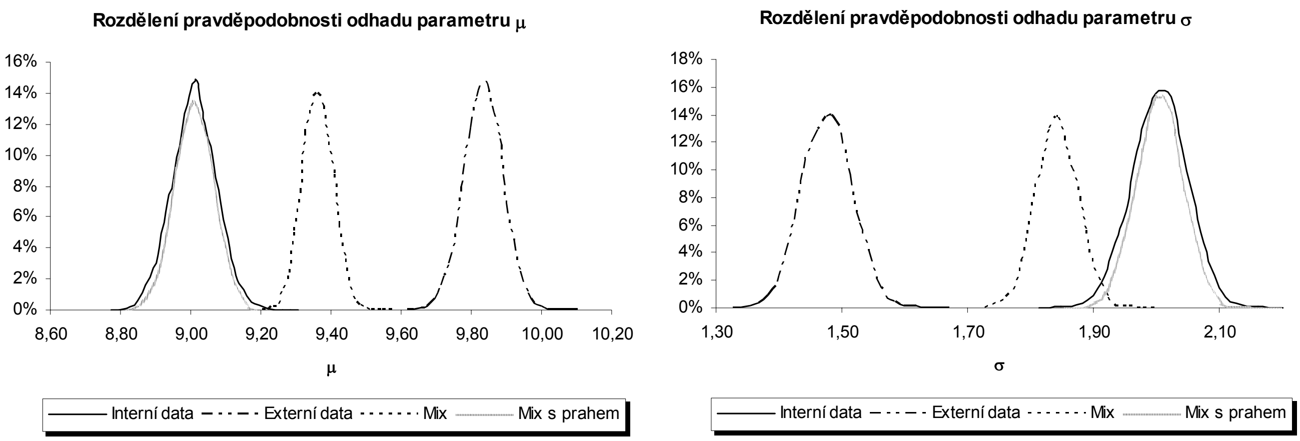
1) „Interní data“ – odhady $\hat{\mu}$ a $\hat{\sigma}$ jsou provedeny pouze na základě interních dat,
2) „Externí data“ – odhady $\hat{\mu}$ a $\hat{\sigma}$ jsou provedeny pouze na základě externích dat,
3) „Mix“ – odhady $\hat{\mu}$ a $\hat{\sigma}$ jsou provedeny na základě sloučení interních a externích dat bez zohlednění prahu sběru dat *H,*
4) „Mix s prahem“ – odhady $\hat{\mu}$ a $\hat{\sigma}$ jsou provedeny na základě sloučení interních a externích dat se zohledněním prahu sběru dat *H* (aplikace rovnic 3 a 4).

Výše uvedený algoritmus byl mnohonásobně opakován (5000 pokusů), čímž bylo získáno rozdělení pravděpodobností odhadů parametrů rozdělení $\hat{\theta}=(\hat{\mu},\hat{\sigma})$, které zobrazuje graf 1.

Z uvedeného grafu vyplývá, že pouze varianty č.1 „Interní data“ a č.4 „Mix s prahem“ vedou k nezkreslenému odhadu parametrů $\hat{\theta}$ (odpovídají parametrům rozdělení Monte Carlo simulace). V případě aplikace variant č.2 nebo č.3 by došlo

[1] Lognormální rozdělení s parametry μ = 9 a σ=2 odpovídá standardnímu rozdělení používanému při modelování ztrát z událostí operačního rizika

k výraznému zkreslení odhadovaných parametrů $\hat{\theta}$ a tím i k neadekvátnímu navýšení odhadu potřebného kapitálu ke krytí podstupovaného rizika.



Graf č.1 Rozdělení pravděpodobnosti odhadů parametrů μ a σ

2.2. Konstantní a neznámý práh sběru externích dat

Předpokládejme nyní, že externí ztráty jsou nadále zaznamenány pouze v případě, že překročí stanovený práh výše ztráty *H*, který je v čase konstantní avšak jeho hodnota není uživateli databáze známá.

V tomto případě dochází k rozšíření množiny parametrů o práh sběru dat *H*, který je potřeba současně s ostatními parametry rozdělení odhadnout metodou maximální věrohodnosti. Věrohodnostní funkce má shodný tvar jako v předchozí kapitole s tím rozdílem, že *H* je nyní považováno za jeden z hledaných parametrů, blíže viz Frachot (2002).

$$l(\theta, H) = \sum_{i=1}^{n} \ln s(\zeta_i; \theta) + \sum_{i=1}^{n^*} \ln s^*(\zeta_i^*; \theta, H). \tag{5}$$

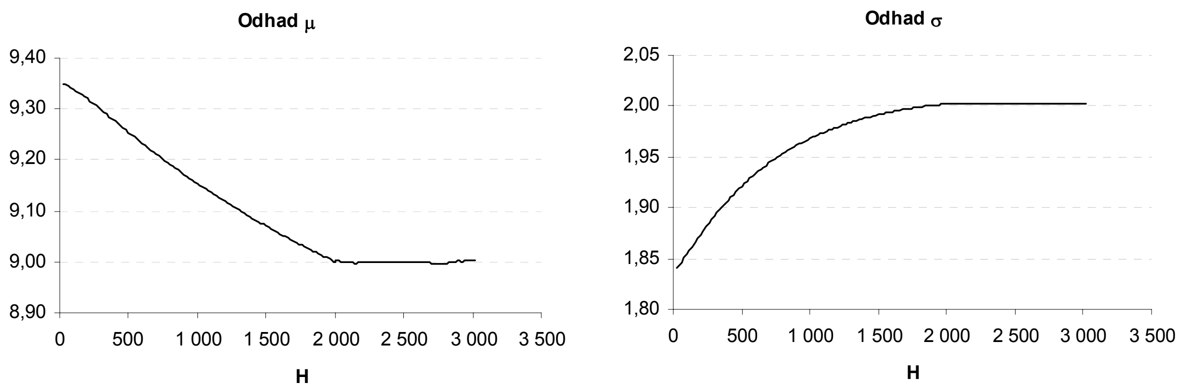
Lze konstatovat, že odhad $\hat{H}$ lze v souladu s metodou maximální věrohodnosti získat pomocí následujícího vztahu

$$\hat{H} = \min_i \zeta_i^* . \tag{6}$$

Vzhledem k možné horší kvalitě databáze externích dat (nejnižší hodnota může tvořit odlehlou hodnotu a být tedy dramaticky nižší než většina ostatních dat) je zpravidla vhodné odhad $\hat{H}$ prověřit pomocí procedury obsahující následující kroky:

1. ocenění parametrů $\hat{\theta}$ pro každé *H* z intervalu od 0 do ∞,
2. grafické zobrazení $\hat{\theta}(H)$,
3. nalezení hodnoty *H,* kde začíná stabilita odhadu $\hat{\theta}(H)$.

Takto nalezená hodnota *H* pak představuje hledaný odhad $\hat{H}$. Pro ověření uvedeného algoritmu byla opět provedena Monte Carlo simulace v rámci které byly generovány obdobně jako v předchozí kapitole dvě množiny náhodných čísel z lognormálního rozdělení $LN(9,2)$ reprezentující interní a externí databáze (pro externí databázi opět aplikován práh sběru dat ve výši 2000). Následně byl prováděn odhad parametrů $\hat{\theta}$ pro jednotlivé hodnoty *H* z intervalu 0 až 3000 pomocí metody maximální verohodnosti (věrohodnostní funkce je dána rovnicí č.4), kdy byly z externí databáze dodatečně vyloučeny ztráty nižší než zvolená hodnota *H.* Výsledky zobrazuje následující graf.



Graf č.2 Odhady parametrů μ a σ pro různé hodnoty prahu sběru dat H

Z grafu je patrné, že výše popsaný algoritmus odhalil stanovený práh sběru dat externí databáze ve výši 2000 při odhadu obou parametrů μ a σ.

Dále lze ukázat, že v případě slučování více externích databází s různými prahy sběru dat *H* dochází ke stabilizaci odhadů $\hat{\theta}$ pro nejvyšší hodnotu *H* ze všech databází, což má za důsledek pouze částečné využití externě napozorovaných dat (externí data nižší než nejvyšší hodnota *H* nejsou brány při odhadu parametrů $\hat{\theta}$ v úvahu).

Závěry získané v předchozí kapitole platí také v tomto případě (sloučení interních a externích dat bez zohlednění prahu sběru dat by vedlo k vychýleným odhadům parametrů rozdělení pravděpodobnosti $\hat{\theta}$).

2.3. Stochastický práh sběru externích dat

Předpokládejme nyní, že externí ztráty jsou nadále zaznamenány pouze v případě, že překročí stanovený práh výše ztráty *H*, který se však považuje za stochastický a lze jej popsat rozdělením pravděpodobnosti dle následující funkce

$$H \sim g(h;\delta), \qquad (7)$$

kde δ představuje množinu parametrů charakterizujících rozdělení pravděpodobnosti *g*. Funkce hustoty zaznamenávaných externích ztrát $s^*(\zeta^*;\theta,\delta)$ má pak následující tvar, viz Frachot (2002),

$$s^*(\zeta^*;\theta,\delta) = \int_0^\infty s^*(\zeta^*;\theta | H = h) g(h;\delta) dh, \qquad (8)$$

kde podmíněná funkce hustoty $s^*(\zeta^*;\theta|H=h)$ odpovídá rovnici č.3. Věrohodnostní funkce pro odhad parametrů $\hat{\theta}$ je v tomto případě zobecněním předchozí věrohodnostní funkce (rovnice 5) a má následující tvar

$$l(\theta,\delta)=\sum_{i=1}^{n}\ln s(\zeta_i;\theta)+\sum_{i=1}^{n^*}\ln s^*(\zeta_i^*;\theta,\delta). \tag{9}$$

Po dosazení rovnice 8 do rovnice 9 se stává věrohodnostní funkce více komplexní a její ohodnocení je možno provést například numerickými metodami. Platnost uvedeného vztahu lze opět ověřit například pomocí Monte Carlo simulace obdobně jako v předchozích kapitolách.

Výhodou tohoto přístupu ve srovnání s přístupem popsaným v předchozí kapitole je, že vede k maximálnímu využití napozorovaných externích dat při slučování různých databází s odlišnými hodnotami prahu *H*.

Praktický problém však spočívá ve volbě vhodného rozdělení pravděpodobnosti prahu sběru dat *H*, které významným způsobem ovlivňuje rozdělení pravděpodobnosti externích ztrát $s^*(\zeta^*;\theta,\delta)$.

Lze dokázat, že také v tomto případě platí závěr uvedený v předchozích kapitolách, který říká, že sloučení interních a externích dat bez zohlednění prahu sběru dat by vedlo k vychýleným odhadům parametrů rozdělení pravděpodobnosti $\hat{\theta}$.

Závěr

Článek byl věnován popisu a ověření statistické metody, která vede ke sloučení interní a externí datové základny tak, aby nedošlo k vychýlení odhadovaných parametrů rozdělení pravděpodobnosti výše ztráty z události operačního rizika. Byly popsány tři modelové případy, které se liší pojetím prahu sběru dat externí databáze. První případ odpovídal situaci, kdy práh sběru dat je konstantní a uživateli databáze známý. Druhý případ předpokládal konstantní, avšak neznámý práh sběru dat a třetí případ považoval práh sběru dat za stochastický.

Pomocí modelových propočtů bylo prokázáno, že sloučení interní a externí datové základny bez odpovídajícího zohlednění prahu sběru dat externí databáze vede k vychýleným odhadům parametrů rozdělení pravděpodobnosti ztrát. Negativním důsledkem pak může být neadekvátní navýšení kapitálového krytí podstupovaného rizika.

LITERATURA

1. CRUZ, M. G., Modeling, Measuring and Hedging Operational Risk, John Wiley & Sons Ltd., 2002
2. FRACHOT, A., RONCALLI, T., BAUD, N. Internal Data, External Data and Consortium Data for Operational Risk Management: How to pool data properly, Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France, 2002
3. FRACHOT, A., RONCALLI, T., Mixing internal and external data for managing operational risk, Working Paper, Groupe de Recherche Opérationnelle, Crédit Lyonnais, France, 2002

4. HUŠEK, R. a LAUBER, J. Simulační modely. STNL – Nakladatelství technické literatury Praha, 1987
5. PANJER, H., J., Operational Risk – Modeling Analytics, John Wiley & Sons Ltd., 2006
6. ZMEŠKAL, Z. a KOLEKTIV, Finanční modely, Ekopress, 2004

KONTAKT

Ing. Jiří Havlický
Českomoravská stavební spořitelna, a.s.
Vinohradská 3218/169
100 17, Praha 10
Email : jiri_havlicky@cmss.cz