Risk events, statistical point of view

Rizikové události, statistické pojetí

Tomáš Ťoupal [1], František Vávra[2] [3]

Abstract

This paper deals with a discussion of the term risk and we present the processing of model of risk events. The specificity of this model is that the risk event occurs when the value of one random process exceeds the value of the second random process. Therefore, we model two processes: time of the first realization of risk event and the probability of event when the aggregated random component exceeds the aggregated systematic component, i.e. a risk event occurs. Then we focus on the relationship between these two individual processes. We also list important notes for the creation of correct model which have to be fulfilled. Otherwise, the model can produce results that do not necessarily correspond to the reality.

Key words

Risk, risk event, probability modeling, statistical inference.

JEL Classification: C44, C49 [4]

1. Úvod

Pojem rizika, obdobně jako pojem trendu, obdobně jako i jiné popisné pojmy je často používán bez kritického a exaktnějšího pohledu. Je to dáno tím, že mnohé oblasti lidského jednání mají mnoho analogického a na popisné úrovni je danému pojmu rozuměno přijatelně společně v adekvátní míře shody. Tento efekt však mizí, když se dostaneme do oblasti kvantifikace (kolik, kdy, …) a srovnávání (větší nebo menší riziko, …). V odborné literatuře existují pokusy o exaktním zavedení pojmu rizika, např. v [1]. Ty mají velký význam, ale zatím jsou na počátku svého vývoje. Existují i některé dopracované metodiky „měření rizika“ jak je uvedeno např. v [2] a [3]. Ty však v kontextu dalších událostí (tzv. „finanční krize“) týkajících se majitelů práv k těmto metodikám působí poněkud rozpačitě. Zde se budeme zabývat jedním specifickým pojmem a „rizikovou událostí“ při porovnávání dvou náhodných procesů.

1.1 Základní pojmy

Riziková událost je jev, který pokud nastane, lze charakterizovat trojicí následujících údajů. Časem kdy nastal (měřeno od nějakého, vztažného, počátečního okamžiku), efektem (náklady rizikové události, mohou být jako jedno číslo – častější, nebo více čísel = soubor charakteristik – méně časté, obtížnější) a pravděpodobností toho, že nebude detekována (časté

[1] Ing. Tomáš Ťoupal, Katedra matematiky, Fakulta aplikovaných věd Západočeské univerzity v Plzni, ttoupal@kma.zcu.cz.
[2] doc. Ing. František Vávra, CSc., Katedra matematiky, Fakulta aplikovaných věd Západočeské univerzity Plzni, vavra@kma.zcu.cz
[3] Práce byla zpracována v rámci projektu NTIS - Nové technologie pro informační společnost, reg. číslo CZ.1.05/1.1.00/02.0090, v rámci Operačního programu Výzkum a vývoj pro inovace, prioritní osy 1, evropská centra excelence, oblast podpory 1.1 Evropská centra excelence.
[4] Zdroj: http://www.aeaweb.org/jel/jel_class_system.php#C

v technických aplikacích, např. odhalitelnost materiálové vady vedoucí k budoucímu porušení součásti, …) a v netechnických oblastech pravděpodobností toho, že nastane před (po) daným okamžikem s „garantovaným“ efektem.

Uvažovaným modelem rizikové události může být tzv. bilanční vztah, kdy jedna náhodná veličina převýší druhou. V technických aplikacích lze hovořit např. o vztahu, kdy zátěžová síla převýší pevnost materiálu a v ekonomických aplikacích lze zmínit vztah, kdy např. (hypoteticky) výdaje převýší příjmy (tzv. likvidní default) nebo kdy náklady převýší výnosy o známou mez danou vlastním jměním, atd.

1.2 Formulace a příklady modelů

Nechť máme k dispozici dva náhodné procesy označované jako $x(t), y(t)$, pak za rizikovou událost budeme považovat jev, kdy $\{x(t) > y(t)\}$. Nechť každý dílčí náhodný proces má svou systematickou složku („trend“, sezónnost, …) označovanou jako $X(t), Y(t)$ a složku náhodnou (…) s označením X_t, Y_t. Potom pravděpodobnost rizikové události, v našem pojetí této práce, bude vyjádřena jako $P(x(t) > y(t))$.

Samotné propojení obou složek může mít následující povahu:

- aditivní

$$x(t) = X(t) + X_t\,;\;\; y(t) = Y(t) + Y_t$$
$$P(x(t) > y(t)) = P(X_t - Y_t > Y(t) - X(t))$$

- multiplikativní

$$x(t) = X(t)(1 + X_t);\;\; y(t) = Y(t)(1 + Y_t);\, X(t), Y(t) > 0;\, X_t, Y_t > -1$$
$$P(x(t) > y(t)) = P\left(\frac{1 + X_t}{1 + Y_t} > \frac{Y(t)}{X(t)}\right) = P\left(\lg\left(\frac{1 + X_t}{1 + Y_t}\right) > \lg\left(\frac{Y(t)}{X(t)}\right)\right)$$

- smíšené, např.

$$x(t) = X(t)(1 + X_t);\;\; y(t) = Y(t) + Y_t\,;\, X(t) > 0;\, X_t > -1$$
$$P(x(t) > y(t)) = P(X(t)X_t - Y_t > Y(t) - X(t)).$$

Všechny zde uvedené (ale i jiné) lze převést do pravděpodobnostního vyjádření ve tvaru $P(Z_t > Z(t))$, kde:

$Z_t = X_t - Y_t\,,\; Z(t) = Y(t) - X(t))$ u aditivního,

$Z_t = \lg\left(\frac{1 + X_t}{1 + Y_t}\right),\; Z(t) = \lg\left(\frac{Y(t)}{X(t)}\right)$ u multiplikativního,

$Z_t = X(t)X_t - Y_t\,,\; Z(t) = Y(t) - X(t))$ u uvedeného příkladu smíšeného modelu.

Za zmínku stojí, že provedení takového „oddělení“ u náhodné a systematické složky není jediné možné.

2. Vazba času do první rizikové události a vztahu mezi náhodnou a systematickou složkou

Nadále budeme pracovat s předchozím modelem pravděpodobnosti výskytu rizikové události ve tvaru

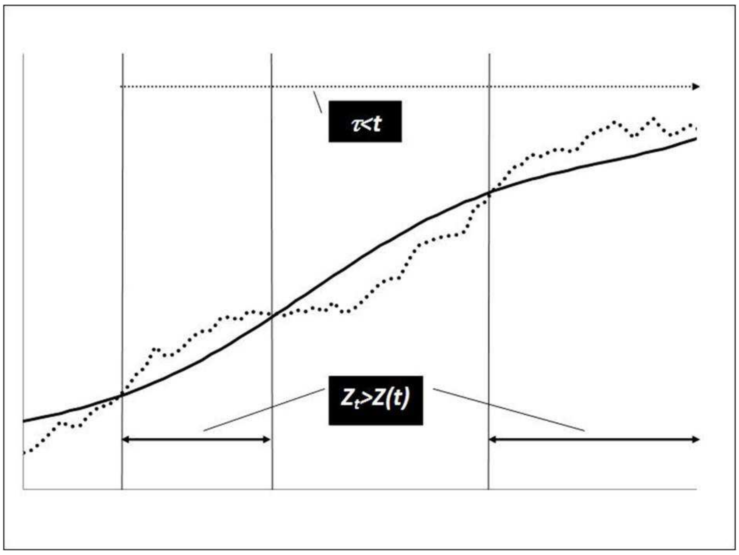
$$P(Z_t > Z(t)). \tag{2.1}$$

Dále označíme τ jako čas prvního dosažení rizikové oblasti $Z_t > Z(t)$ za předpokladu, že ve vztažném, počátečním čase t_0 platila nerovnost $Z_{t_0} < Z(t_0)$. Tj. soustava je na počátku sledování v nerizikovém stavu. Samozřejmě analogicky lze modelovat situaci, kdy je riziková událost definována opačným vztahem $Z_t < Z(t)$. Nadále proto budeme pracovat s konvencí rizikové události $Z_t > Z(t)$. Z uvedeného je zřejmé, že platí

$$\{Z_t > Z(t)\} \subset \{\tau < t\}, \tag{2.2}$$

což je demonstrováno následujícím obrázkem.

Obrázek 1: Demonstrace vztahu $\{Z_t > Z(t)\} \subset \{\tau < t\}$



Tento triviální vztah v sobě skrývá vazbu mezi časem do výskytu první rizikové události a chováním obou srovnávaných komponent, přesněji vazbu mezi jejich agregovanou náhodnou složkou Z_t a agregovanou systematickou složkou $Z(t)$. Potom pravděpodobnost toho, že se náhodná složka bude nacházet v rizikové „zóně“ (bude setrvávat nebo se bude opakovat riziková událost) pokud se tam alespoň jednou dostala, lze vyjádřit jako:

$$P(Z_t > Z(t) / \tau < t) = \frac{P(Z_t > Z(t); \tau < t)}{P(\tau < t)} = \frac{P(Z_t > Z(t))}{P(\tau < t)} . \tag{2.3}$$

Dále se budeme zabývat jen takovými soustavami, u nichž platí (motivováno např. v [4], „náhodná procházka“), že

$$\frac{P(Z_t > Z(t))}{P(\tau < t)} = \rho, \qquad 0 < \rho < 1 . \tag{2.4}$$

Pravděpodobnost ρ je vlastně pravděpodobnost toho, že náhodná složka Z_t převýší systematickou složku $Z(t)$ (riziková „zóna“, tj. nastala nebo nastala opakovaně riziková událost), pokud se alespoň jednou (předtím) již realizovala riziková událost. Obvykle bývá uváděna (učebnicově) hodnota $\rho = 1/2$. V takovém případě se jedná o „symetrickou“ náhodnou složku kolem systematické složky a znamenalo by to, že je stejná pravděpodobnost výskytu v rizikové a nerizikové zóně (za podmínky, že již bylo alespoň jednou do rizikové zóny $Z_t > Z(t)$ vstoupeno). Podmínku (2.4) splňuje poměrně široká skupina modelů náhodné složky Z_t ve vztahu ke složce systematické (procesy s nezávislými přírůstky, jejich limitní verze, …), nejedná se tedy o významné omezení.

Ze vztahu (2.4) bezprostředně plyne následující rovnost

$$P(\tau < t) = \frac{1}{\rho} P(Z_t > Z(t)), \qquad 0 < \rho < 1. \tag{2.5}$$

Nyní pro další úpravy označíme distribuční funkce:

$F_\tau(t) = P(\tau < t)$ jako distribuční funkci náhodné proměnné prvního času dosažení rizikové oblasti a

$F_{Z_t}(t; x) = P(Z_t < x)$ jako distribuční funkce náhodné proměnné Z_t v čase t.

Potom pro tyto zavedené distribuční funkce dostáváme:

$$F_\tau(t) = \frac{1}{\rho}\left(1 - F_{Z_t}(t; Z(t))\right) \tag{2.6}$$

a duálně k tomu:

$$F_{Z_t}(t; Z(t)) = 1 - \rho F_\tau(t). \tag{2.7}$$

Nepřímým důsledkem vztahu (2.4) je skutečnost, že budeme pracovat s náhodnou složkou, jejíž pravděpodobnostní model je „translačně invariantní“, tj.

$$F_{Z_t - Z_0}(t; x) = F_{Z_t}(t - t_0; x - Z_0); \quad Z_0 < Z(t_0). \tag{2.8}$$

Stručně řečeno, pro měření (zjišťování) jsou podstatné časové a hodnotové přírůstky. To znamená, že sledujeme a vyhodnocujeme běžící, nikoliv jednorázová dění. Opět toto „omezení“ je podstatným pouze v teorii, nikoliv v realitě (jak již bylo uvedeno dříve, jedná se o alespoň potenciálně běžící děje), neboť nemáme k dispozici absolutní čas. Za daných výchozích podmínek jsou vztahy (2.6) a ekvivalentně (2.7) tuhou vazbou mezi časem dosažení „rizikové oblasti“. Popis vztahu náhodné složky ke složce systematické jednoznačně určuje čas prvního dosažení „rizikové oblasti“ a naopak.

Příklad: Pokud pro modelování času do prvního výskytu rizikové události užijeme často zmiňované a populární exponenciální rozdělení, dostaneme vztah:

$$F_\tau(t) = 1 - e^{-\frac{t}{T}}$$

a

$$P(Z_t < Z(t)) = 1 - \rho + \rho e^{-\frac{t}{T}}.$$

Zjednodušeně řečeno v takovém případě pravděpodobnost toho, že soustava zůstane v nerizikové oblasti, exponenciálně klesá k hodnotě $1-\rho$, viz vztah (2.4) a střední doba (od počátku pozorování) do rizikové události je pak T.

3. Závěry, náměty a diskuze

Srovnávané procesy $x(t), y(t)$ mohou mít povahu v čase aktuálních hodnot, kumulací od počátku pozorování (náhodné procházky, procesy s nezávislými přírůstky, jejich limitní verze, …) nebo kumulací za uvažované období. Tomu musí odpovídat model $F_{Z_t}(t;x) = P(Z_t < x)$. Z modelované reality dále vyplyne, zda se jedná o (alespoň potenciálně) pokračující procesy (po dosažení rizikové zóny) nebo se jedná o procesy, které po dosažení rizikové zóny končí (úlohy o „stopping time“, úlohy obnovy, uvedeno např. v [5]) a následně dochází k „restartu“ (oprava, náprava, …) nového běhu. V případě takových procesů se bude parametr ρ zjišťovat jinou metodikou (maximální věrohodnost), než u soustav, kde jsou pozorovatelné „pobyty“ v rizikové zóně a mimo ni. Veškerá statistická inference bude této skutečnosti podřízena.

Obecně by měla být statistická inference uskutečňována na úrovni dílčích srovnávaných procesů $x(t), y(t)$, protože odhady z jejich rozdílu $x(t) - y(t)$ budou ve většině případů zatíženy velkou chybou. Pokud se bude jednat o aplikační úlohy, měly by být zvažovány tři základní fakty. První, aby model co „nejlépe“ popisoval modelovanou realitu. Druhý, aby zvolený model měl použitelné a interpretovatelné parametry a výstupy. A třetí, aby model respektoval potřebnou „hloubku“ detailu. Příliš detailní model vede na široká chybová pásma a případně i rozsáhlé datové zdroje pro „statistické zpracování“, naopak výsledky příliš agregovaného modelu bývají obvykle nepoužitelné pro aplikace.

Vztah (2.4) vlastně určuje, že nenáhodná složka $Y(t) - X(t)$ nebo $\lg\left(\frac{Y(t)}{X(t)}\right)$ musí být odhadována jako kvantilová čára (přesněji dvě kvantilové čáry pro procesy $x(t), y(t)$) z pozorování po prvním dosažení rizikové oblasti (nebo alespoň potenciálně). Jedna z možných metodik pro takové postupy je uvedena např. v [6].

Obecně je v modelu rizikové události $\{x(t) > y(t)\}$ zahrnuta i situace „s posunutím“ tj. $\{x(t) > y(t) + c\}$, kde c je nějaká konstanta. V konkrétním případě je však zapotřebí respektovat její specifika. Tento případ je poměrně častý a lze se s ním setkat např. v situaci, kde $x(t)$ jsou náklady, $y(t)$ jsou výnosy a c jsou disponibilní, dodatečné, zdroje (vlastní kapitál, …).

Literatura

[1] Artzner, Ph., Delbaen F., Eber, J.-M., and Heath, D., 1997. *Thinking Coherently*, RISK 10, November, 68-71.

[2] Introduction to RiskMetrics™, November 21, 1995, *Fourth edition*. New York: Morgan Guaranty Trust Company, Risk Management Services, riskmetrics@jpmorgan.com.

[3] CreditMetrics™, April 2, 1997. *Technical Document*, New York: J.P. Morgan.

[4] Rozanov, J.A., 1979. *Slučajnye processy*. Moskva: Nauka.

[5] Aven, T., Jensen U., 1998. *Stochastic Models in Reliability*. New York: Springer Verlag.

[6] Ťoupal, T., 2012. *Odhad trendové složky*, 6th International Scientific Conference Managing and Modelling of Financial Risk, VŠB-TU Ostrava, Faculty of Economics, Finance Department 10th – 11th September 2012.

Shrnutí

Práce vychází z diskuze pojmu riziko a je v ní předloženo zpracování jednoho modelu rizikové události. Model je specifický tím, že riziková událost nastává, pokud jeden náhodný proces překročí hodnotu druhého. Jsou modelovány dva jevy. Čas prvního uskutečnění rizikové události a pravděpodobnost jevu kdy agregovaná náhodná složka překročí agregovanou systematickou a tím nastane riziková událost. Práce se soustřeďuje na vzájemný vztah obou dílčích problémů. V práci jsou uvedeny poznámky, bez jejichž respektování, při tvorbě modelu dochází (nebo může docházet) k paradoxním, realitě neodpovídajícím výsledkům.