



Marek Karwański

Szkoła Główna Gospodarstwa Wiejskiego
w Warszawie
Wydział Zastosowań Informatyki i Matematyki
Katedra Informatyki
marek_karwanski@sggw.pl

Urszula Grzybowska

Szkoła Główna Gospodarstwa Wiejskiego
w Warszawie
Wydział Zastosowań Informatyki i Matematyki
Katedra Informatyki
urszula_grzybowska@sggw.pl

DOBÓR PROGÓW ŁĄCZENIA STRAT POCHODZĄCYCH Z RÓŻNYCH ŹRÓDEŁ W RYZYKU OPERACYJNYM

Streszczenie: Banki do obliczeń ryzyka operacyjnego wykorzystują metodologię LDA (*Loss Distribution Approach*) polegającą na estymacji rozkładów prawdopodobieństwa strat. Rozkłady hybrydowe są bardzo atrakcyjną alternatywą dla wieloparametrycznych modeli rozkładów stosowanych w praktyce. Tym niemniej rodzą szereg problemów natury technicznej. Na podstawie przykładów przeliczonych w artykule zaprezentowane zostały rozwiązania w dziedzinie budowy estymatorów, które mogą być stosowane w praktyce.

Słowa kluczowe: ryzyko operacyjne, łączenie danych wewnętrznych i zewnętrznych, metody estymacji parametrów rozkładów prawdopodobieństwa, rozkłady hybrydowe.

JEL Classification: C63, C53.

Wprowadzenie

W dzisiejszych czasach banki i inne instytucje finansowe znajdują się pod presją instytucji nadzorczych, wymuszających lepsze zarządzanie ryzykiem operacyjnym oraz alokację kapitału ekonomicznego zgodnie z wymogami regulacyjnymi. W ostatnich latach opublikowane zostały przepisy prawne, takie jak: dyrektywy UE w sprawie wymogów kapitałowych, uchwały KNF transponujących postanowienia Bazylei II/III, dyrektywy Solvency II i warunków wykonywania działalności ubezpieczeniowej w Polsce w ramach Wyplacalności II, Ustawa o finansach publicznych, przepisy dotyczące kontroli wewnętrznej nad sprawozdawczością finansową Sarbanes-Oxley, przepisy o Równym Traktowa-

niu Klientów (*Treating Customers Fairly*), zasady w sprawie kontroli w zakresie informacji i związanych z nimi technologiami COBIT, a także normy ISO 17799 / ISO 27002. Stawia to bardzo restrykcyjne wymagania przed systemami analitycznymi wspierającymi zarządzanie ryzykiem [Deloitte, 2007].

Regulacje prawne, choć nie narzucają konkretnych rozwiązań, przyczyniły się do rozwoju systemów analitycznych wspierających zarządzanie ryzykiem. Cruz [2002] opisuje różne metody analityczne wykorzystywane do obliczenia ryzyka operacyjnego z uwzględnieniem specyfiki ich stosowania w instytucjach finansowych. Porównanie podejść analitycznych w ramach metodologii: IMA (*Internal Measurement Approach*) i LDA (*Loss Distribution Approach*), prezentują [Berger, 2005; Feng i in., 2007; Kilavuka, 2008; Shevchenko, 2009].

Opis przykładowego systemu znaleźć można w artykule [Aue, Kalkbrenner, 2007]. Autorzy pokazują na przykładzie systemu działającego w Bundes Banku wiele elementów modeli matematycznych zaimplementowanych w praktyce i otwierają dyskusję nad metodologią pomiaru ryzyka operacyjnego. Inny artykuł dotyczący funkcjonowania systemu pomiaru ryzyka znaleźć można w pracy [Brown, Wang, 2005].

Do obliczeń wykorzystywane są różne typy modeli. Najpopularniejsze to:

- modele probabilistyczne oparte na rozkładach prawdopodobieństw – tzw. LDA (*Loss Distribution Approach*),
- modele scoringowe oparte na klasyfikowaniu zdarzeń do różnych klas na podstawie analizy atrybutów tych zdarzeń.

Oba typy różnią się zarówno sposobem zbierania danych, jak i metodologią liczenia ryzyka, która musi uwzględniać wymogi prawne zawarte w odpowiednich ustawach.

W ramach niniejszego artykułu omówione zostaną elementy modeli prawdopodobieństwa w ramach podejścia LDA wykorzystujące różne źródła danych. W chwili obecnej jest bardzo mało prac opartych na rzeczywistych danych pozwalających określić czynniki i metody, które mogą być wykorzystane do liczenia ryzyka z uwzględnieniem danych pochodzących z różnych źródeł. Artykuł Samad-Khan [2005] to pionierskie wprowadzenie w problematykę wykorzystania danych zewnętrznych. Dalsze rozwinięcie tych modeli znaleźć można w pracach [Baud, Frachot, Roncalli, 2002; Sataputera Na, 2004; Sataputera Na i in., 2006; Dahan, Dionne, 2007].

Systemy pomiaru Ryzyka Operacyjnego muszą brać pod uwagę źródła danych: dane wewnętrzne i zewnętrzne, analizy scenariuszy wariantowych RCSA (*Risk Control Self Assessment*), czynniki odzwierciedlające otoczenie gospodar-

cze KRI (*Key Risk Indicator*) oraz systemy kontroli wewnętrznej [Uchwała nr 1/2007 KNB].

Różne źródła danych zawierają różne profile zdarzeń i dlatego nie mogą być łączone bezpośrednio. Potrzebne są procedury odpowiednio skalujące dane. Łączenie informacji uzyskanej np. ze scenariuszy i danych wewnętrznych oraz zewnętrznych w ramach modeli LDA można przeprowadzić na dwa sposoby:

- *ex post* – czyli przeprowadzić oddzielne modelowanie i połączyć wyniki na poziomie zagregowanym,
- *ex ante* – włączyć informacje o parametrach rozkładów do modeli częstości i dotkliwości zdarzeń na etapie procedur analitycznych.

Celem niniejszego artykułu jest analiza różnych podejść stosowanych przy łączeniu danych wewnętrznych i zewnętrznych w ramach procedur *ex ante*. Wyniki zostaną zaprezentowane na przykładzie danych jednego z polskich banków, który wykorzystuje dane zewnętrzne z bazy konsorcjalnej ZORO. Z uwagi na poufność danych zostały one poddane pewnym modyfikacjom zapewniającym odpowiedni poziom poufności, które pozwalają jednak na rzeczywiste porównanie różnych podejść.

1. Podstawy teoretyczne liczenia ryzyka operacyjnego

Ryzyko operacyjne według definicji Komitetu Bazylejskiego definiowane jest w oparciu o straty w zadanym przedziale czasowym, np. jednego roku. Załóżmy, że wartości strat są reprezentowane przez zmienną losową X , a ich liczba przez zmienną losową N oraz że są od siebie niezależne. Oznacza to, że straty agregowane w skali jednego roku można zapisać w postaci [Gilli, Källezi, 2006]:

$$Strata_{agregat} = \sum_{i=1}^n x_i \quad (1)$$

gdzie:

- n – liczba zdarzeń w ciągu 1 roku (realizacja zmiennej N);
- x_i – dotkliwości kolejnych strat (realizacje zmiennej X).

Postać rozkładu dla agregatu strat wyraża się wzorem:

$$F_S(x) = \begin{cases} p(0) + \sum_{n=1}^{\infty} p_N(n) F_X^{n*}(x) & \text{dla } x > 0 \\ p_N(0) & \text{dla } x \leq 0 \end{cases} \quad (2)$$

gdzie:

$p_N(n)$ – funkcja gęstości rozkładu zmiennej N (liczby zdarzeń w ciągu 1 roku);

F_S – funkcja rozkładu agregatu S ;

F_x^{n*} – n -krotny spłot rozkładu F_x .

Wartość ryzyka oblicza się na podstawie wartości straty oczekiwanej EL (obliczonej dla rozkładu F_S) oraz straty nieoczekiwanej UL (obliczonej jako 99,95% percentyl rozkładu F_S).

Z uwagi na problem dostępności danych estymację rozkładu F_S przeprowadza się w oparciu o parametryczne modele rozkładów p_N i F_X . [Klugman, Panjer, Willmot, 2008] to jeden z pełniejszych wykładów o rodzinach parametrycznych rozkładów używanych w ryzyku operacyjnym. Obejmuje kilkadziesiąt rodzin rozkładów od jedno- do czteroparametrycznych.

W wielu przypadkach, spotykanych w praktyce do modelowania parametrycznych rozkładów prawdopodobieństwa, wykorzystuje się tzw. modele hybrydowe, które można zdefiniować następująco:

$$f(x) = \begin{cases} a_1 f_1(x), & 0 < x \leq x_b \\ a_2 f_2(x) & x_b < x < \infty \end{cases} \quad (3)$$

gdzie a_i są wagami, a f_i rozkładami obcięzonymi (*truncated*).

Wykorzystuje się uproszczony sposób zapisu rozkładu hybrydowego:

$$f(x) = \begin{cases} r f_1^*(x), & 0 < x \leq x_b \\ (1-r) f_2^*(x) & x_b < x < \infty \end{cases} \quad (4)$$

gdzie $f_1^*(x) = \frac{f_1(x)}{F_1(x_b)}$ i $f_2^*(x) = \frac{f_2(x)}{1-F_2(x_b)}$.

Na rozkłady narzucany jest jeden z warunków:

- funkcja dystrybuanty F jest klasy C^0 (warunek ciągłości),
- rozkład $f \in Z^0$ (funkcja ciągła), czyli funkcja dystrybuanty F jest klasy C^1 (warunek różniczkowalności).

W pierwszej części artykułu zostanie przedyskutowany warunek różniczkowalności, ponieważ jego stosowanie w istotny sposób modyfikuje uzyskane wyniki.

2. Warunek różniczkowalności

Liczbę parametrów można zmniejszyć, wykorzystując więzy, np. założenie o odpowiedniej gładkości sklejaną f_1 i f_2 . Niestety okazuje się, że wprowadzenie więzów pozwala co prawda zmniejszyć liczbę estymowanych parametrów, ale

jednocześnie wprowadza inne zaburzenia, które są równie dotkliwe. Można to prześledzić na poniższym symulowanym przykładzie.

Weźmy pod uwagę straty X , których dotkliwość będzie modelowana dwoma parametrycznymi rozkładami. Pierwszy rozkład będzie używany do analizy zdarzeń typowych, f_1 – „ciało”, czyli występujących stosunkowo często, drugi do zdarzeń rzadkich, ekstremalnych, f_2 – „ogon”. Jako rodzinę dla zdarzeń typowych można zaproponować 2-parametryczną- (μ, σ) rodzinę rozkładów log-normalnych, z funkcją gęstości:

$$f_{\log_normal}(x; \mu, \sigma) = \frac{1}{x\sigma\sqrt{2\pi}} \exp\left(-\frac{(\ln(x) - \mu)^2}{2\sigma^2}\right) \quad (5)$$

Jako rodzinę dla zdarzeń rzadkich wybrano 2-parametryczny- (β, ξ) rozkład z rodziny Uogólnionych Rozkładów Pareto (GDP) [Buch-Kromann, 2009] o funkcji gęstości:

$$f_{GDP}(x) = \begin{cases} \frac{1}{\beta} \left(1 + \xi \frac{(x - x_b)}{\beta}\right)^{-1 - (1/\xi)} & \xi \neq 0 \\ \frac{1}{\beta} \exp\left(-\frac{(x - x_b)}{\beta}\right) & \xi = 0 \end{cases} \quad (6)$$

Dodatkowym parametrem rozkładu hybrydowego jest punkt podziału x_b . Przyjmowane będzie założenie, że wartości typowe to wartości, gdy ($strata < x_b$) natomiast wartości rzadkie to takie, gdy ($strata \geq x_b$).

Rozkład hybrydy log-normal/GPD będzie modelem prawdopodobieństwa wystąpienia straty. W tym przypadku trzeba estymować 5 parametrów $(\mu, \sigma, \xi, \beta, x_b)$. Bardzo często parametr x_b estymowany jest w oddzielnej procedurze – na przykład uwarunkowanej procesem zbierania danych. Dlatego w dalszych rozważaniach będzie on traktowany jako parametr o krytycznym wpływie na wynik estymacji, podlegający oddzielnym szacowaniom.

Wygenerujmy dane pochodzące z rozkładu mieszaniny dwóch rozkładów parametrycznych:

$$F(x) = \begin{cases} \frac{p_n}{G(x_b)} G(x), & \text{jeżeli } x \leq x_b \\ p_n + (1 - p_n)H(x - x_b), & \text{jeżeli } x > x_b \end{cases} \quad (7)$$

gdzie:

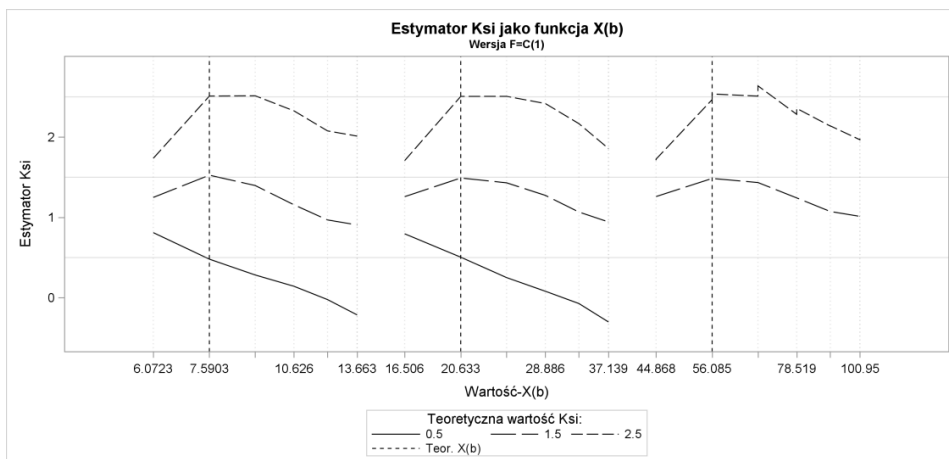
p_n – waga mieszania;

rozkład $G()$ – log-normalny;

rozkład $H()$ – rozkład GDP.

Estymatory szacowane są metodą największej wiarygodności. Warto zwrócić uwagę na estymatory parametrów rozkładu „ogona”, gdyż od niego głównie zależy wartość ryzyka UL.

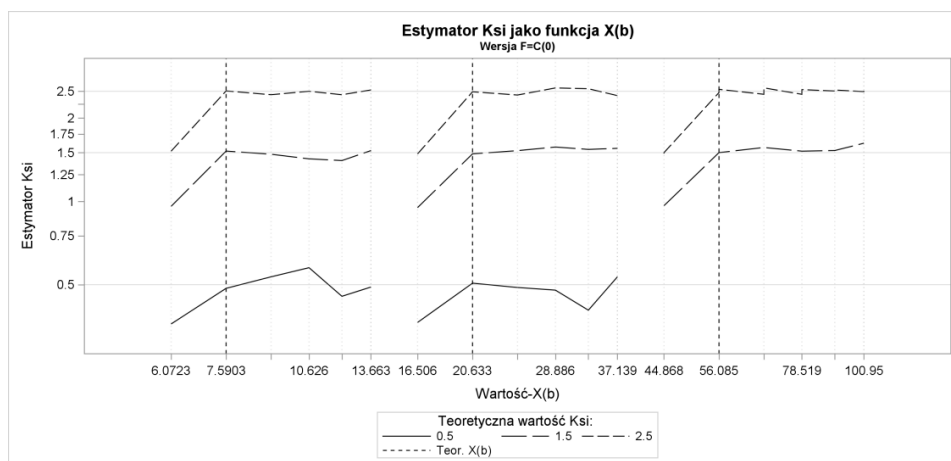
W przypadku, gdy wymagany jest warunek ciągłości, estymator $\hat{\xi}$ (w rozkładzie GPD) silnie zależy od doboru progę łączenia x_b – parametr ten ma krytyczne znaczenie przy estymacji.



Rys. 1. Estymator parametru $\hat{\xi}$ w funkcji wyboru parametru x_b dla $F \in C^1$

Źródło: Opracowanie własne.

Inaczej wygląda sytuacja w przypadku, gdy odrzucimy warunek ciągłości. Wówczas estymator $\hat{\xi}$ jest znacznie mniej podatny na wpływ x_b – prawidłową wartość uzyskać można w szerokim zakresie zmian x_b . Zależność estymatora $\hat{\xi}$ (w rozkładzie GPD) w obu przypadkach zależy silnie od wyboru progę łączenia x_b . Jako wniosek z tego przykładu można przyjąć, że warunek różniczkowalności nie jest niezbędny, aczkolwiek wymaga to dodatkowych badań.



Rys. 2. Estymator parametru ξ w funkcji wyboru parametru x_b dla $F \in C^0$

Źródło: Opracowanie własne.

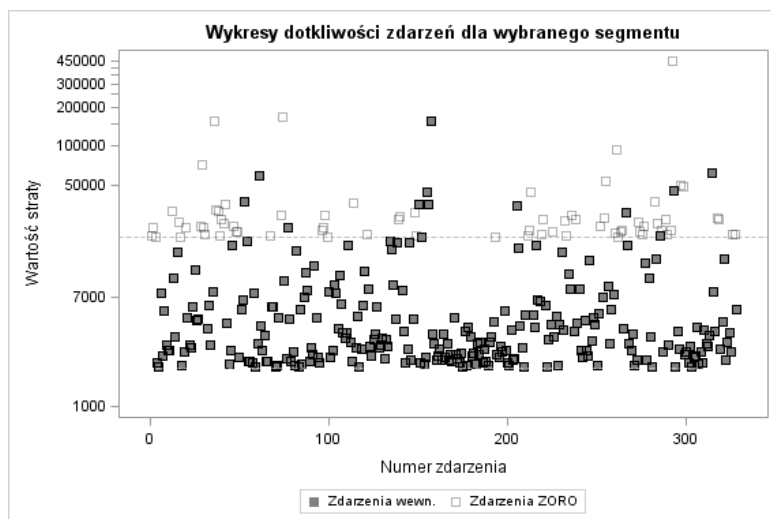
W dalszej części artykułu obowiązywać będzie założenie, że rozkłady hybrydowe będą budowane w oparciu o warunek ciągłości, czyli $F \in C^0$.

3. Dane wewnętrzne i zewnętrzne

Zgodnie z Rekomendacją M/KNF dotyczącą zarządzania ryzykiem operacyjnym w bankach banki obok rejestracji strat wewnętrznych powinny w miarę możliwości dokonywać wymiany informacji z innymi bankami na temat przypadków wystąpienia strat oraz gromadzić te informacje i je analizować. Analiza takich danych zewnętrznych powinna obejmować weryfikację przyczyn i poziomu strat.

Poniżej przedstawione zostały obliczenia dla dużego polskiego banku uwzględniające dane wewnętrzne i zewnętrzne. Pod uwagę wzięto tylko dotkliwość strat (zmienna losowa X), przyjmując, że liczba strat $N = 1$. Bank został podzielony na kilka segmentów związanych z liniami biznesowymi (dyskusja podziału banku na segmenty wykracza poza zakres artykułu i nie będzie tutaj przytaczana). Do analizy przyjęto jako dane wewnętrzne informacje z jednego segmentu, zbierane w okresie czterech lat: 2013-2016. Z uwagi na zmiany w środowisku biznesowym banku dane zostały poddane transformacjom skalującym z wykorzystaniem kilku wskaźników KRI. Jako dane zewnętrzne wykorzystane zostały dane zbierane w bazie konsorcjalnej ZORO. Poddano je transformacjom ze względu na kategorię banku, w którym powstały. Dyskusja

związana ze skalowaniem wartości strat w obu bazach również wykracza poza zakres artykułu. W celu zapewnienia poufności danych wszystkie analizy, których wyniki prezentowane są poniżej, zostały przeprowadzone na danych zmodyfikowanych. Modyfikacje zachowały relacje pomiędzy estymowanymi parametrami prezentowane we wnioskach.



Rys. 3. Prezentacja danych wewnętrznych i zewnętrznych

Źródło: Obliczenia własne.

4. Analiza i wyniki

Łączenie danych wewnętrznych i zewnętrznych rozpatrywane jest najczęściej przy następujących założeniach [Baud, Frachot, Roncalli, 2002].

Założenie 1

Dane wewnętrzne i zewnętrzne pochodzą z tego samego rozkładu za wyjątkiem tego, że próg obserwowalności danych wewnętrznych wynosi H_0 , a próg obserwowalności danych zewnętrznych wynosi H_1 , przy czym: $H_1 > H_0$.

Założenie 2a (znana wartość progu)

Próg H jest *non-stochastic*, jego wartość jest znana.

Założenie 2b (nieznana wartość progu)

Próg H jest *non-stochastic*, jego wartość jest nieznana.

Założenie 2c (próg stochastyczny)

Próg H jest *stochastic*.

W niniejszym artykule rozpatrywany jest przypadek 2a – Próg H jest *non-stochastic*, jego wartość jest znana, parametry rozkładów estymowane są na danych wewnętrznych i zewnętrznych.

W literaturze znaleźć można różne schematy estymowania rozkładu hybrydowego. Według teorii zdarzeń ekstremalnych próg łączenia rozkładów f_1 i f_2 powinien być odpowiednio wysoki, tak aby można było założyć, że zbieżność f_2 do GDP jest dostateczna. Problem polega na tym, że wybór wysokiego progu w praktyce oznacza, że decydujący wpływ na wartość ryzyka będą miały dane zewnętrzne przy minimalnym udziale danych wewnętrznych. Dodatkowo trzeba pamiętać, że próg zbierania danych w bazie ZORO wynosi 20 000 PLN. Stąd ustawienie wartości x_b na innym poziomie spowoduje konieczność estymacji trzech funkcji rozkładu prawdopodobieństwa: f_{1a} dla strat poniżej 20 000 PLN – w oparciu o dane wewnętrzne, f_{1b} dla strat (20 000; x_b) w oparciu o dane wewnętrzne i zewnętrzne oraz f_2 dla strat powyżej x_b , również w oparciu o dane wewnętrzne i zewnętrzne.

Z powyższych względów wygodnie jest przyjąć, że parametr x_b jest równy 20 000 PLN. Dzięki temu f_1 estymowane jest na danych wewnętrznych, a f_2 na mieszaninie danych wewnętrznych i zewnętrznych.

Do analizy przyjęto trzy warianty estymacji rozkładu hybrydowego:

1. (W1) Rozkłady f_1 i f_2 są takie same z dokładnością do parametrów i różnią się jedynie progiem odcięcia (*truncation*).
2. (W2) Rozkłady f_1 i f_2 są takie same, ale estymowane na różnych danych mają różne parametry i różne progi odcięcia.
3. (W3) Rozkłady f_1 i f_2 są różne, dobierane pod kątem miary dopasowania AIC.

Wariant W1

Za przyjęciem wariantu W1 przemawia rozumowanie, w którym rozkłady dla danych wewnętrznych i zewnętrznych nie powinny się istotnie różnić. W procedurze estymacyjnej należy przyjąć inne progi odcięcia, gdyż dane zbierane są w innych procedurach. Dzięki takiemu podejściu dane wewnętrzne mają bardzo duży wpływ na wartość ryzyka, a dane zewnętrzne służą do uzupełnienia danych wewnętrznych.

Zgodnie z założeniem f_1 i f_2 są równe f oraz próg dla danych zewnętrznych wynosi H , czyli:

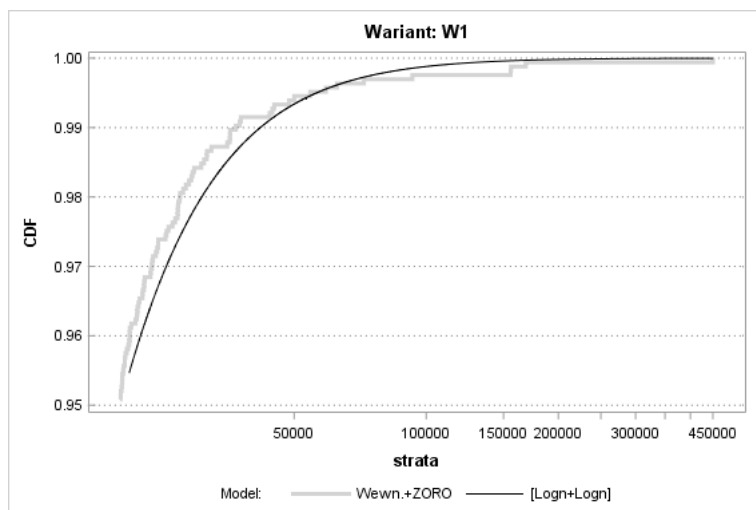
$$X_{wew} \sim f(x; \theta)$$

$$X_{zew} \sim f^*(x; \theta) = \frac{f(x; \theta)}{1 - F(H; \theta)}$$

Stąd funkcja wiarygodności użyta do estymacji parametrów ma postać:

$$\ell(\theta) = \sum \ln f(x; \theta) + \sum 1\{x \geq H\} \cdot \ln \frac{f(x; \theta)}{1 - F(H; \theta)} \quad (7)$$

Jako rodzinę f rozkładów do modelowania ryzyka przyjęto rozkład log-normalny. Zgodnie z oczekiwaniami dopasowanie „ciała” nie było najlepsze, natomiast dopasowanie „ogona” było bardzo dobre, najlepsze ze wszystkich modeli testowanych w tym badaniu. Statystyka AIC dla „ogona” wyniosła 1092.7.



Rys. 4. Rozkład prawdopodobieństwa straty – wariant W1

Źródło: Obliczenia własne.

Wartości statystyk: średnia i percentyle dla pojedynczej straty zostały przedstawione w tabeli 1.

Tabela 1. Statystyki rozkładu pojedynczej straty w wariantcie W1

Model W1	Średnia	Perc. 50,00%	Perc. 99,90%	Perc. 99,99%
[Logn + Logn]	3 518	54	126 417	250 103

Źródło: Obliczenia własne.

Wariant W2

Wariant W2 został wprowadzony jako benchmark dla wariantu W1. W procedurze estymacyjnej należało przyjąć inne progi odcięcia oraz inne parametry, gdyż dane zbierane były w innych procedurach. Niestety w tym podejściu dane wewnętrzne miały stosunkowo mały wpływ na wartość ryzyka, a dane zewnętrzne były głównym źródłem informacji o ryzyku.

Zgodnie z założeniem f_1 i f_2 są równe f , przy założeniu, że próg dla danych zewnętrznych wynosi H , ale mają inne parametry:

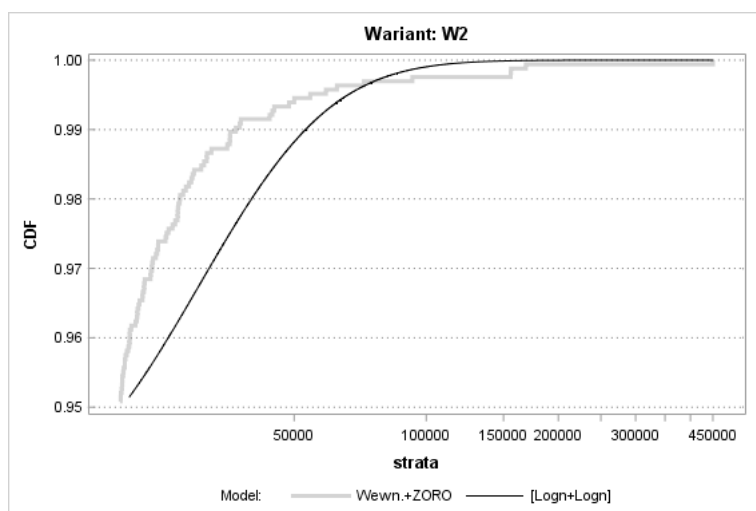
$$X_{wew} \sim f(x; \theta_1)$$

$$X_{zew} \sim f^*(x; \theta_2) = \frac{f(x; \theta_2)}{1 - F(H; \theta_2)}$$

Stąd funkcja wiarygodności użyta do estymacji parametrów ma postać:

$$\ell(\theta_1, \theta_2) = \sum \ln f(x; \theta_1) + \sum 1\{x \geq H\} \cdot \ln \frac{f(x; \theta_2)}{1 - F(H; \theta_2)} \quad (8)$$

Jako rodzinę f rozkładów do modelowania ryzyka przyjęto rozkład log-normalny. Zgodnie z oczekiwaniami dopasowanie „ciała” było wyraźnie lepsze niż w przypadku wariantu W1. Dopasowanie „ogona” było bardzo przeciętne wśród modeli testowanych w tym badaniu. Statystyka AIC dla „ogona” wynosi 1160.7.



Rys. 5. Rozkład prawdopodobieństwa straty – wariantu W2

Źródło: Obliczenia własne.

Wartości statystyk: średnia i percentyle dla pojedynczej straty zostały przedstawione w tabeli 2.

Tabela 2. Statystyki rozkładu pojedynczej straty w wariancie W2

Model W2	Średnia	Perc. 50,00%	Perc. 99,90%	Perc. 99,99%
[Logn + Logn]	5 932	2 239	98 840	153 481

Źródło: Obliczenia własne.

Wariant W3

Za przyjęciem wariantu W3 przemawia rozumowanie, w którym zakładamy, że dane wewnętrzne i zewnętrzne mają różne rozkłady. W procedurze estymacyjnej należy przyjąć zarówno inne progi odcięcia, jak i rodziny rozkładów. W takim podejściu dane wewnętrzne mają wpływ na wartość ryzyka w zależności od wysokości progu H , który jest zależny od proporcji danych.

Zgodnie z założeniem f_1 i f_2 są różne:

$$X_{wew} \sim f_1(x; \theta_1)$$

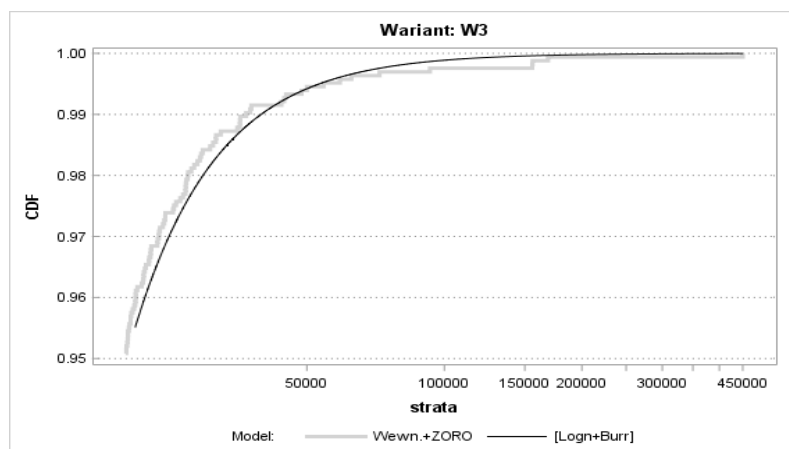
$$X_{zew} \sim f_2^*(x; \theta_2) = \frac{f_2(x; \theta_2)}{1 - F_2(H; \theta_2)}$$

Stąd funkcja wiarygodności użyta do estymacji parametrów ma postać:

$$\tilde{\ell}(\theta_1, \theta_2) = \sum \ln f_1(x; \theta_1) + \sum 1\{x \geq H\} \cdot \ln \frac{f_2(x; \theta_2)}{1 - F_2(H; \theta_2)} \quad (9)$$

Jako rodziny f_1, f_2 rozkładów do modelowania ryzyka przyjęto rozkłady: GPD, IGauss, InvBurr, Burr, Logn, GDP, Weibull, Gamma, Loglog, Frechet, Gumbel [Klugman, Panjer, Willmot, 2008].

Dopasowanie „ciała” oraz „ogona” zależało od przyjętej rodziny rozkładów. Jako miarę dopasowania rozkładu przyjęto kryterium AIC. Najlepiej dopasowany model składał się z $f_1 = \text{log-normal}$ oraz $f_2 = \text{Burr}$. Statystyka AIC dla tego modelu wynosi 1091.9.



Rys. 6. Rozkład prawdopodobieństwa straty – wariantu W3

Źródło: Obliczenia własne.

Wartości statystyk: średnia i percentyle dla pojedynczej straty zostały przedstawione w tabeli 3.

Tabela 3. Statystyki rozkładu pojedynczej straty w wariantcie W3

Model W3	Średnia	Perc. 50,00%	Perc. 99,90%	Perc. 99,99%
[Logn + Burr]	5 284	2 239	104 614	278 226

Źródło: Obliczenia własne.

Podsumowanie

Estymatory hybrydowe są bardzo atrakcyjną alternatywą dla rozkładów wieloparametrycznych. Tym niemniej rodzą szereg problemów natury technicznej. Na podstawie przykładów zaprezentowanych w artykule można zauważyć, że bardzo naturalne wymagania, aby funkcja gęstości była ciągła, mogą być powodem gorszych własności estymatorów. Wydaje się, że bardziej naturalnym kandydatem dla modeli ryzyka operacyjnego jest dystrybuanta.

Prezentowana analiza ograniczona została do jednego zdarzenia, dlatego w tym przypadku ryzyko nie pokrywało się z definicjami bazylejskimi. Tym niemniej na uwagę zasługują wartości 99,99% percentyla jednej straty. Mniejsze wartości tej statystyki nie zawsze oznaczają lepsze oszacowanie, gdyż częstym błędem jest niedoszacowanie ryzyka spowodowane obciążeniem. Regulacyjne Standardy Techniczne zalecają stosowanie, jako miary dobroci dopasowania, narzędzi, które są bardziej wrażliwe na odstępstwa w „ogonie” niż „ciele” roz-

kładu i które biorą pod uwagę zarówno wielkość próby, jak i liczbę szacowanych parametrów – dlatego do oceny użyte zostało kryterium informacyjne AIC. Bardzo interesujące są wyniki wariantów W1 i W3. W obu przypadkach wartości statystyki AIC były bardzo zbliżone, ale również podobne były wartości wysokich percentyli. Niższe percentyle były zdominowane przez dopasowanie w obszarze „ciała” i z punktu widzenia pomiaru ryzyka są mniej interesujące.

Wyniki sugerują stosowanie wariantu W1. Zgodnie z nim dane zewnętrzne mogą być postrzegane jako „ukryte dane wewnętrzne”, co oznacza, że dane zewnętrzne i dane wewnętrzne można łączyć ze sobą w celu uzyskania bazy danych z większą liczbą obserwacji¹.

Literatura

- Aue F., Kalkbrener M. (2007), *LDA at Work: Deutsche Bank's Approach to Quantifying Operational Risk*, "Journal of Operational Risk", Vol. 1, s. 49-93.
- Baud N., Frachot A., Roncalli T. (2002), *Internal Data, External Data and Consortium Data for Operational Risk Measurement: How to Pool Data Properly?* Technical report, Groupe de Recherche Op'erationnelle, Cr'edit Lyonnais, France.
- Berger J.O. (2005), *Statistical Decision Theory and Bayesian Analysis*, Springer Series in Statistics, New York.
- Brown D.J., Wang J.T. (2005), *Implementing an AMA for Operational Risk Implementing an AMA for Operational Risk*, Presentation Federal Reserve Bank of Boston, May 19.
- Buch-Kromann T. (2009), *Comparison of Tail Performance of the Transformed Kernel Density Estimator, the Generalized Pareto Distribution and the G-and H-distribution*, "Journal of Operational Risk", Vol. 4, No. 2, s. 43-67.
- Cruz M. (2002), *Modelling, Measuring and Hedging Operational Risk*, John Wiley & Sons, New York.
- Dahen H., Dionne G. (2007), *Scaling Models for the Severity and Frequency of External Operational Loss Data*, CRCiRM Working Paper 07-01, Canada.
- Deloitte (2007), *Global Risk Management Survey*, 5th Edition.
- Feng Ch., Nitin J., Wanli M., Ramachandran B., Gamarnik D. (2007), *Modeling Operational Risk in Business Processes*, "Journal of Operational Risk", Vol. 2, No. 2, s. 73-98.
- Gilli M., Käellezi E. (2006), *An Application of Extreme Value Theory for Measuring Financial Risk*, "Computational Economics", Vol. 27, No. 1, s. 207-228.

¹ Oczywiście dane muszą być odpowiednio przeskalowane, w opisywanym przykładzie dane zostały poddane transformacji ze względu na wskaźniki KRI i kategorie banków.

- Kilavuka M.I. (2008), *Managing Operational Risk Capital in Financial Institutions*, "Journal of Operational Risk", Vol. 3, No. 1, s. 67-83.
- Klugman S.A., Panjer H., Willmot G.E. (2008), *Loss Models: From Data to Decision*, 3rd Edition, John Wiley & Sons, New York.
- Samad-Khan A. (2005), *Assessing & Measuring Operational Risk. Why COSO is Inappropriate*, Presentation in London, January 18.
- Sataputera Na H. (2004), *Operational Risk. Analysing and Scaling Operational Risk Loss Data*, Erasmus University Rotterdam, Netherlands.
- Sataputera Na H., van den Berg J., Lourenco C.M., Marc L. (2006), *An Econometric Model to Scale Operational Losses*, "Journal of Operational Risk", Vol. 1, s. 11-31.
- Shevchenko P.V. (2009), *Implementing Loss Distribution Approach for Operational Risk*, CSIRO "Mathematical and Information Sciences", Sydney.
- Uchwała nr 1/2007 Komisji Nadzoru Bankowego z dnia 13 marca 2007 r., Dz. Urz. NBP nr 2/2007, poz. 3.

THE STRATEGIES OF COMBINING LOSS DATA FROM DIFFERENT SOURCES IN OPERATIONAL RISK

Summary: Banks use LDA (*Loss Distribution Approach*) method to estimate loss probability distributions. Mixed distributions are a very attractive alternative to multi-parameter distribution models currently used. However, this approach gives rise to a number of technical problems. Based on the real life examples the suitable solutions are presented.

Keywords: operational risk, internal and external data, probability mixed distributions.