



Tomasz Szkutnik

Uniwersytet Ekonomiczny w Katowicach
Wydział Zarządzania
Katedra Ekonometrii
tomasz.szkutnik@uekat.pl

ANALIZA ŹRÓDEŁ NIEPEWNOŚCI W MODELOWANIU RYZYKA OPERACYJNEGO

Streszczenie: W artykule poruszony został problem związany z identyfikacją źródeł niepewności w modelowaniu ryzyka operacyjnego w oparciu o metodę LDA. Modelowanie ryzyka operacyjnego w bankowości opiera się najczęściej na informacjach zbieranych powyżej określonego progu. Fakt ucinania obserwacji jako jeden ze szczególnych rodzajów cenzurowania danych w wielu praktycznych przypadkach nie jest brany po uwagę przy estymacji parametrów rozkładu dotkliwości strat operacyjnych w metodzie LDA. W artykule zostały porównane różne formy cenzurowania danych wraz z oceną stopnia oddziaływania utraty informacji na zmienność zarówno wartości szacowanych parametrów rozkładów, jak i wielkości wyznaczonego ryzyka. Proponowane w artykule wykorzystanie estymacji cenzurowanej, które nie jest szeroko stosowane w praktyce bankowej, pozwala na poprawę jakości estymacji bez konieczności pełnego raportowania strat banku. Ponadto cenzurowanie informacji stanowi cenne źródło informacji nie tylko w procesie dotkliwości strat, ale także zapewnia dodatkowe informacje przy modelowaniu procesu częstości strat. Za pomocą symulacji pokazane jest oddziaływanie różnych sposobów zbierania informacji na jakość estymatorów wybranego rozkładu dotkliwości strat.

Słowa kluczowe: ryzyko operacyjne, LDA, cenzurowanie danych, ucinanie danych.

Wprowadzenie

Analiza źródeł niepewności w odniesieniu do ryzyka operacyjnego, będąca tematem przewodnim przeprowadzonych badań, skupia uwagę na zasadniczej kwestii, jaką jest charakter źródła informacji. Docelowym punktem zainteresowania każdej analizy ryzyka jest wyznaczenie wielkości ryzyka, która w przypadku ryzyka operacyjnego najczęściej określana jest na podstawie modelu rozkładu strat LDA. Metoda LDA polega na wyznaczeniu zagregowanego rozkładu

strat z dwóch oddzielnych procesów, jakimi są proces częstości i proces dotkliwości strat. Szeroki opis metody rozkładu strat można znaleźć w takich pracach, jak [Shevchenko, 2011], [Klugman i in., 2008] oraz [Szkutnik, 2013].

Straty operacyjne są postrzegane jako zmienne losowe należące do przestrzeni \mathbb{R}^+ . W wielu przypadkach charakter danych dotyczących wielkości strat operacyjnych właściwych dla danej instytucji jest pewnym podzbiorem dodatniej półosi. Wynika to bezpośrednio ze sposobu zbierania informacji dotyczących zdarzeń i start operacyjnych. Ogólnie przyjętym standardem jest raportowanie i rejestrowanie w systemach wielkości strat powyżej określonego progu, przy czym wyróżnia się dwie szczególne sytuacje, tj. ucinanie danych oraz ich cenzurowanie. Występujące w praktyce dwie sytuacje dotyczące ucinania lub cenzurowania danych mogą prowadzić do różnych sposobów kwantyfikacji ryzyka. W przypadku danych uciętych dostępny zakres informacji w stosunku do rzeczywistego zjawiska jest ograniczony jednostronnie lub obustronnie. Wyróżnia się dane ucięte lewostronnie (rejestrowane są wtedy tylko obserwacje powyżej określonego progu), ucięte prawostronnie (rejestrowanie danych poniżej określonego progu) i ucięte obustronnie (dane rejestrowane są w przedziale pomiędzy wartością dolną i górną ograniczenia).

Proces cenzurowania danych jest podobny do procesu ucinania danych, z tą różnicą, że dla danych cenzurowanych obserwacja niespełniająca relacji mniejszości lub większości w stosunku do określonych progów lewo lub prawostronnych jest zliczana bez rejestrowania wartości zmiennej losowej. Dzięki procesowi cenzurowania danych utrata informacji jest znacznie mniejsza niż w przypadku ich ucinania. Podobnie jak dla danych uciętych, w przypadku danych cenzurowanych wyróżnia się cenzurowanie lewostronne, prawostronne lub cenzurowanie obustronne. Dla procesu ryzyka operacyjnego właściwe będą mechanizmy ucinania lub cenzurowania lewostronnego. Wynika to bezpośrednio z systematyki ryzyka operacyjnego występującego w bankowości. Od procesu raportowania i rejestracji danych oczekuje się zbierania informacji o wielkościach przekroczeń powyżej ustalonych progów. Należy podkreślić dodatkowo, że to proces ucinania danych jest najbardziej powszechny i w praktyce rzadko spotyka się sytuacje, w których dane rejestrowane są w postaci danych cenzurowanych (częściej zamiast operacji cenzurowania niektóre instytucje wybierają rejestrowanie wszystkich danych, jednak w praktyce korzystają tylko z ich pewnego uciętego podzbioru).

1. Wpływ ucinania i cenzurowania obserwacji na rozkład dotkliwości strat operacyjnych

Celem niniejszej części pracy jest zweryfikowanie za pomocą badania symulacyjnego wpływu ucinania oraz cenzurowania danych na wielkość parametrów wybranego rozkładu prawdopodobieństwa dotkliwości strat oraz na wielkości miary Value-at-Risk. Podkreślone są tu mocne i słabe strony możliwych do zastosowania rozwiązań, w zależności od przyjętej koncepcji modelowej.

Badanie ma na celu przedstawienie problemu ze względu na stosowane powszechnie w praktyce rozwiązania polegające na modelowaniu danych uciętych rozkładami dla pełnej dziedziny zmiennej losowej oraz poddanie tych rozwiązań krytycznej weryfikacji w kontekście innych, równie praktycznych rozwiązań.

2. Opis badania symulacyjnego

Wyróżniono cztery możliwe scenariusze, jakie mogą wystąpić w przypadku modelowania procesów dotkliwości strat. Każdy z nich oparty jest na dwóch założeniach. Pierwsze założenie wiąże się z wyborem dziedziny zmiennej losowej $x \in (h, +\infty)$, w zależności od progu h , gdzie $h = 0$ – dane pełne, oraz $h > 0$ – dane ucięte lub dane cenzurowane (w zależności od sposobu postępowania z obserwacjami poniżej progu h). Drugie założenie związane jest z przyjętym do modelowania ciągłym rozkładem prawdopodobieństwa (rozkład standardowy, rozkład ucięty lewostronnie, rozkład cenzurowany lewostronnie), którego parametry estymowane są w zależności od przyjętego wariantu dla danych z pierwszego założenia. Określenie *rozkład standardowy* oznacza, że został wybrany do modelowania określony rozkład prawdopodobieństwa. Natomiast rozkłady ucięte i cenzurowane lewostronnie są tymi samym rozkładami co rozkład standardowy, jednak fakt ograniczonej informacji uwzględniony jest na etapie estymacji (tzn. zostały wybrane odpowiednie procedury estymacji parametrów dla rozkładów uciętych oraz cenzurowanych).

Postać funkcji wiarygodności odpowiednia dla przypadku danych uciętych i cenzurowanych jest identyczna jak w pracy [Klein, 2003, s. 74], dlatego nie będzie tu powtarzana. Odpowiednie podstawienia do podanej tam funkcji wiarygodności pozwalają na rozpatrzenie zarówno przypadku danych cenzurowanych, jak i uciętych. W celu estymacji wymienionych typów rozkładów zostało wykorzystane oprogramowanie R, wraz z pakietami *fitdistrplus* oraz *truncdist*.

Wybór progów h z założenia pierwszego, dotyczącego scenariuszy symulacyjnych, nie może być wielkością przyjętą z góry, jednak kryteria jego wyboru muszą być w stanie obiektywnie porównać różnego rodzaju rozkłady i ich zdolność do opisu zjawiska wraz z postępowaniem procesów utraty informacji. Zapewnienie porównywalności dla różnych rodzajów rozkładów parametryzowanych w odmienny sposób wymaga przyjęcia wartości progów h w oparciu o wartość kwantyla odpowiedniego rzędu. Tak zdefiniowany próg h będzie gwarantował, że odpowiedni procent zmiennych losowych będzie odpowiednio odrzucany – w przypadku danych uciętych – lub zliczany w przypadku danych cenzurowanych.

3. Przyjęte scenariusze symulacyjne

Na przyjęte scenariusze symulacyjne składają się:

1. Scenariusz bazowy – dane pełne, rozkład standardowy.
2. Scenariusz naiwny – dane ucięte, rozkład standardowy.
3. Scenariusz ucięty – dane ucięte, rozkład ucięty.
4. Scenariusz cenzurowany – dane cenzurowane, rozkład cenzurowany.

4. Ogólny schemat badania symulacyjnego

Badanie symulacyjne zostało podzielone na następujące kroki:

Krok 1: Dla wybranego rozkładu standardowego oraz przyjętych dla niego arbitralnie parametrów generowane są pseudolosowe zmienne, co stanowi jedną realizację teoretycznego rozkładu prawdopodobieństwa. Zbiór ten nazwany jest zbiorem bazowym i oznaczony jest jako XB . Liczba wygenerowanych obserwacji oznaczona jest przez NB .

Krok 2: Należy wyznaczyć wartości progów h określającego poziom, od którego obserwacje ze zbioru XB będą ucinane lub cenzurowane.

Wartość progów h_i , jako kwantyl rzędu i , gdzie: $i \in \{0; 0, 1; 0, 2; 0, 3; 0, 4; 0, 5; 0, 6; 0, 7; 0, 8\}$ dla wybranego rozkładu teoretycznego i określonych parametrów.

Krok 3: Dla każdej z wartości progów h_i z kroku 1 i należy utworzyć zbiory danych uciętych XU oraz cenzurowanych XC na podstawie danych XB z kroku 1.

Krok 4: Do utworzonych zbiorów należy dopasować zdefiniowaną funkcję prawdopodobieństwa, przyjmując odpowiednie kryteria wynikające ze scenariuszy od 1 do 4:

- Do danych ze zbioru XB należy dopasować rozkład standardowy.
- Do danych ze zbioru XU należy dopasować rozkład standardowy.

- Do danych ze zbioru XU należy dopasować rozkład ucięty.
- Do danych ze zbioru XC należy dopasować rozkład cenzurowany.

Krok 5: Kroki od 1 do 4 należy powtórzyć określoną liczbę iteracji NI dla wybranego rzędu kwantyla i .

Krok 6: Dysponując wszystkimi NI oszacowanymi zbiorami wartości parametrów dla modeli pochodzących z rozważanej rodziny rozkładu teoretycznego, należy wyznaczyć oddzielnie dla każdego z oszacowanych parametrów określone charakterystyki liczbowe, takie jak: wartość średnia, odchylenie standardowe czy współczynnik zmienności w celu identyfikacji głównych tendencji, jakie będą charakteryzowały rozpatrywane scenariusze symulacyjne.

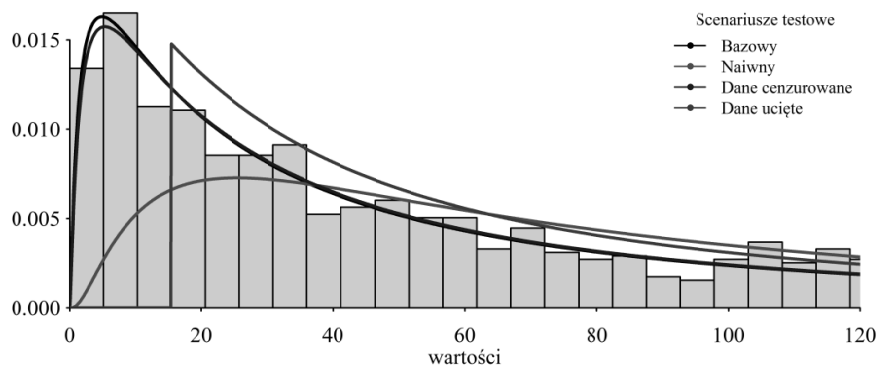
Krok 7: Dysponując uśrednionymi wynikami symulacji z kroku 6, należy procedurę powtórzyć, począwszy od kroku 1 do 6, dla wszystkich wartości przyjętych kwantyli oznaczonych jako elementu zbioru i .

Realizacja scenariuszy testowych w oparciu o przyjęty schemat symulacji pozwala na określenie tempa zmian w zależności od przyjętego podziału wartości zmiennych losowych według wartości kwantyla odpowiedniego rzędu.

W celu dokładniejszej analizy zjawiska można zagęścić szereg wartości h_i , przyjmując większą liczbę wartości z zadanego przedziału (np. dla danych z przedziału $[0; 0, 8]$, przyjmując krok rzędu 1%, można wyznaczyć 81 elementowy wektor). Jednak w przypadku chęci identyfikacji jedynie głównej tendencji, wystarczający jest przyjęty podział wektora na 9 elementów. Ogólny schemat badania symulacyjnego został zrealizowany dla rozkładu logarytmiczno-normalnego oznaczanego dalej jako LN.

5. Porównanie wpływu ucinania oraz cenzurowania danych na wielkości miary VaR na przykładzie rozkładu logarytmiczno-normalnego

Za hipotetyczny rozkład przyjęto rozkład LN(4; 1, 5). Rysunek 1 przedstawia sytuację, w której dla wybranego rozkładu wygenerowano 1000 realizacji zmiennych losowych i oszacowano jego parametry przy różnych kryteriach. Badanie przebiegało według scenariuszy testowych zdefiniowanych w ogólnym schemacie badania symulacyjnego. Histogram (rys. 1) przedstawia dane pełne, tj. z przedziału $(0, +\infty)$. W celu demonstracji wpływu ucinania oraz cenzurowania informacji, na podstawie wygenerowanych danych przedstawiono poniżej jeden scenariusz symulacyjny opisany we wcześniejszej procedurze symulacyjnej wraz z dopasowanymi funkcjami gęstości dla różnych wariantów. Przyjęto, że dane będą lewostronnie ucięte dla progów $h \approx 15.44$. Taka wielkość stanowi w przybliżeniu wartość kwantyla rzędu 0,2, co oznacza, że procesowi ucinania lub cenzurowania danych poddanych zostało około 20% obserwacji.



Rys. 1. Porównanie dopasowania rozkładu logarytmiczno-normalnego przy różnych wariantach estymacji parametrów

Źródło: Opracowanie własne.

Odpowiednie krzywe przedstawiają dopasowanie rozkładów LN przy założeniach scenariuszy testowych z kroku 4 (opisu schematu symulacyjnego). Scenariusz bazowy przedstawia dopasowanie rozkładu LN do danych pełnych (kolor czarny), scenariusz naiwny przedstawia dopasowanie rozkładu LN do danych powyżej progu h (w sytuacji, gdy około 20% danych poniżej progu h będzie niedostępnych, co zaznaczono kolorem czerwonym), scenariusz dla danych cenzurowanych ukazuje dopasowanie krzywej gęstości dla lewostronnego cenzurowania (około 20% obserwacji poniżej progu jest zliczonych i informacja ta jest wykorzystana w procesie estymacji, natomiast prawdziwe wartości nie są dostępne, co zaznaczono kolorem niebieskim), natomiast scenariusz dla danych uciętych prezentuje dopasowanie rozkładu uciętego lewostronnie (lewostronne odrzucenie danych poniżej progu h). Na przedstawionym rysunku 1 widać, jak proces rejestrowania danych tylko powyżej określonego progu wpływa na dopasowywany rozkład prawdopodobieństwa. W tym przypadku pominięcie faktu ucinania danych w znaczący sposób wpłynęło na postać dopasowanego rozkładu.

Proces ucinania obserwacji powoduje, że brak jest informacji o skali zjawiska poniżej progu h , co może stanowić w praktyce duży problem, bowiem dopasowywany rozkład prawdopodobieństwa nie musi być rozkładem, z którego pełne dane będą pochodzić. Fakt wykorzystania rozkładów dla pełnej dziedziny bez uwzględniania lub badania procesu ucinania danych w znaczący sposób obniża efektywne modelowanie dotkliwości strat operacyjnych przy przyjętych w hipotetycznym przykładzie założeniach.

Nieodpowiedni wybór procedur estymacji, jakie powinny być wykorzystane przy wnioskowaniu o postaci rozkładu dotkliwości strat, można zaliczyć do głównego źródła niepewności przy modelowaniu ryzyka operacyjnego. Brak jasno zdefi-

niowanego problemu lub nawet celowe pominięcie specyfiki danych powoduje, że modelowane zjawisko, tj. wyznaczenie wielkości miary ryzyka, samo w sobie będzie obciążone ryzykiem operacyjnym, w tym przypadku ryzykiem modelu.

W celu dodatkowego pokazania różnic pomiędzy otrzymanymi rozwiązaniami wyznaczone zostały wartości kwantyli rzędu 0,99 oraz 0,999, mające przedstawiać wartości miary Value-at-Risk. Dla ryzyka operacyjnego miara VaR nie jest wyznaczana z rozkładów dotkliwości, tylko np. za pomocą metody LDA ze zagregowanego rozkładu strat, jednak wpływ rozkładu dotkliwości, jako jeden z 2 procesów składających się na metodę LDA, ma silny wpływ na późniejsze wartości VaR pochodzące z rozkładu zagregowanego.

W artykule podane są tylko wartości miary VaR dla scenariusza bazowego, scenariusza naiwnego oraz scenariusza dla danych uciętych (wartości dla scenariusza dla danych cenzurowanych nie są prezentowane, ponieważ w rozważanym przykładzie są to wielkości średnie, które są prawie identyczne ze scenariuszem bazowym).

Wyniki zawarte w tabeli 1 pokazują, że w przypadku, gdy rzeczywistym procesem byłby rozkład LN o zadanych parametrach i przy założeniu, że około 20% obserwacji zostanie odrzuconych, to różnice w prawym ogonie wyrażone przez wartości miary VaR są znaczne. Wartości VaR przy scenariuszu bazowym są mniejsze o około 30% i 50%, odpowiednio dla poziomów ufności 0,99 i 0,999.

Tabela 1. Porównanie wartości miary VaR dla wybranych scenariuszy testowych

Scenariusz testowy	VaR 0,99	VaR 0,999
Bazowy	1846	5969
Naiwny	1288	3065
Dane ucięte	1938	5836

Źródło: Opracowanie własne.

Szczegółowe wyniki dotyczące prezentowanego przykładu symulacyjnego są w tabeli 2 (załącznik). Z zamieszczonych tam informacji wynikają dwa istotne wnioski. Po pierwsze, pominięcie faktu ucinania danych i wykorzystania rozkładów z dziedziny \mathbb{R}^+ powoduje, że powstające różnice pomiędzy szacunkami parametrów z dwóch scenariuszy tj. naiwnego i uciętego, będą miały tendencje do zwiększania się wraz z przesuwaniem się progu ucinania w prawą stronę. Przesunięcie progu ucinania równoznaczne będzie w tym przypadku z utratą informacji. Jak pokazano w tabeli 1, nawet 20% utrata informacji wpływa na zmniejszenie się szacunków wielkości VaR dla rozkładu dotkliwości strat o około 50% (co całkowicie będzie zmieniać wyniki uzyskane z metody LDA). Spowodowane jest to faktem, że dopasowując rozkład do tej części dziedziny, z któ-

rej brak jest jakichkolwiek obserwacji, funkcja gęstości może starać się rozłożyć swoją masę także w tym zakresie. W przypadku rozkładów o dużej liczbie parametrów, charakteryzujących się znaczną elastycznością, zjawisko to będzie mniej widoczne niż w przypadku takich rozkładów, jak np. rozkład logarytmiczno-normalny. Ponadto sam proces ucinania danych jest kompromisem pomiędzy utratą informacji a kosztami ich zbierania. Duża liczba strat o niskiej dotkliwości może łącznie przełożyć się na znaczne straty dla banku. Pominięcie w tym zakresie specyfiki danych będzie prowadziło nieuchronnie do błędnych szacunków parametrów ciągłych procesów dotkliwości strat. Skala tych błędów będzie uzależniona od przyjętego progu ucinania oraz od rozkładu analizowanych danych.

Po drugie, należy podkreślić fakt, że proces ucinania danych nie tylko wpływa na proces dotkliwości strat. Brak informacji o wielkości straty oznacza równocześnie brak informacji, że taka strata zaistniała. W związku z tym proces ucinania danych wpływa także na wspomniany we wstępnej części proces częstości występowania strat (jako jedna ze składowych metody LDA). W praktyce będzie to powodowało w pewnym stopniu niedoszacowanie wielkości ryzyka wyznaczanego ze zagregowanego rozkładu strat.

Wspominany aspekt ucinania danych motywowany jest zainteresowaniem stratami o największej dotkliwości oraz redukcją kosztów koniecznych do poniesienia, które wynikałyby z faktu rejestrowania wszystkich zdarzeń. Wybór pomiędzy rejestrowaniem wszystkich danych a ich ucinaniem poniżej zadanego progu nie jest jedynym rozwiązaniem, jakie można zaproponować w tym zakresie. Proces cenzurowania danych nie jest powszechny w praktyce bankowej, ale jego wpływ na jakość estymacji nie może być pominięty.

Brak jest znaczących rozwiązań w zakresie ryzyka operacyjnego, które by wskazywały na proces cenzurowania danych jako jednej z możliwości przy kwantyfikacji ryzyka. W przypadku cenzurowania danych realizowane są dwa zasadnicze elementy. Po pierwsze, proces dotkliwości jest estymowany z uwzględnieniem częściowych informacji o obserwacjach poniżej zadanego progu. Po drugie, zachowana jest pełna informacja o częstości zdarzeń, jakie zaszły w analizowanym obszarze, z którego pochodzą obserwacje. Proces cenzurowania danych daje jeszcze jedną możliwość, bowiem liczba obserwacji cenzurowanych zliczonych poniżej danego progu nie musi pochodzić z systemu sprawozdawczości w banku. Wielkość ta w praktyce może być zakładana arbitralnie (np. przyjęcie, że 20% danych jest cenzurowanych) lub może wynikać z innych modeli, zarówno tych eksperckich o kwalitatywnym charakterze, jak również tych o charakterze ilościowym, będąc tym samym przedmiotem np. analiz statystycznych.

Z informacji zawartych w tabeli 2 wynika, że w hipotetycznym przypadku utrata informacji nawet przy wysokim poziomie cenzurowania nie była znaczna. Oznacza to, że cenzurowanie nie miało zasadniczego wpływu na oszacowania estymatorów (co można zauważyć, porównując uśrednione wielkości). Wzrasta tylko w tym przypadku rozrzut wartości estymowanych wielkości, co widać w kolumnach odpowiednio $V_z(\mu)$ oraz $V_z(\sigma)$, natomiast ich średni poziom dla przeprowadzonych symulacji pozostaje zbliżony do wartości scenariusza bazowego. W odniesieniu do tych samych wielkości z tabeli 2 dla procesu ucinania (elementy *Fit.trunc* z kolumny *Przypadek*), wielkości uśrednione $E(\mu)$ oraz $E(\sigma)$ są zbieżne z rozwiązaniem bazowym, nawet gdy proces ucinania danych dotyczy 60-70% obserwacji. Natomiast nie można tego powiedzieć w odniesieniu do poziomu zróżnicowania, tj. wielkości $V_z(\mu)$ oraz $V_z(\sigma)$, bowiem w tym przypadku widać postępujący wraz z procesem ucinania danych wzrost poziomu zmienności (czego w tak znacznym stopniu nie można zaobserwować w przypadku estymacji cenzurowanej).

Podsumowanie

W kwantyfikacji ryzyka operacyjnego proces cenzurowania stanowi znaczącą alternatywę dla obecnie używanych rozwiązań. Koszty, jakie mogą być związane z rejestrowaniem danych poniżej progu cenzurowania, będą znacznie mniejsze niż przy pełnym raportowaniu wielkości straty. Dane te mogą być rejestrowane poza właściwą bazą danych dotyczącą strat operacyjnych banku. Interesujące informacje związane z cenzurowanym zdarzeniem podawane mogą być zbiorczo w ustalonych podokresach, np. miesięcznie lub tygodniowo.

Zważając na fakt, że proces kwantyfikacji ryzyka operacyjnego nie musi być wykonywany z dużą częstością, to dane dotyczące zdarzeń cenzurowanych mogą być uaktualniane cyklicznie. Proces cenzurowania danych może być także rozumiany jako kolejny krok w rozwoju metod kwantyfikacji ryzyka operacyjnego. Eliminowanie źródeł związanych z niepewnością i wyrażanie nieznanych procesów w kategoriach mierzalnych powinno być dostrzegane w kategoriach efektywności procesu kwantyfikacji i zarządzania ryzykiem, co powinno znaleźć odpowiednie pokrycie w badaniach i późniejszych rekomendacjach Komitetu Bazylejskiego.

Aneks

Tabela 2. Wyniki symulacji przy różnych wariantach estymacji*

Przypadek	Poziom cenzurowania lub ucinania obserwacji	$E(\mu)$	$E(\sigma)$	$V_z(\mu)$	$V_z(\sigma)$
Fit.base	xxx	4,00	1,50	0,01	0,02
Fit.naive	0	4,00	1,50	0,01	0,02
Fit.naive	0,1	4,29	1,26	0,01	0,02
Fit.naive	0,2	4,52	1,14	0,01	0,03
Fit.naive	0,3	4,75	1,05	0,01	0,03
Fit.naive	0,4	4,97	0,97	0,01	0,03
Fit.naive	0,5	5,20	0,90	0,01	0,04
Fit.naive	0,6	5,45	0,83	0,01	0,04
Fit.naive	0,7	5,74	0,77	0,01	0,05
Fit.naive	0,8	6,10	0,70	0,01	0,07
Fit.cens	0	4,00	1,50	0,01	0,02
Fit.cens	0,1	4,00	1,50	0,01	0,02
Fit.cens	0,2	4,00	1,50	0,01	0,03
Fit.cens	0,3	4,00	1,50	0,01	0,03
Fit.cens	0,4	4,00	1,50	0,01	0,03
Fit.cens	0,5	4,00	1,50	0,01	0,04
Fit.cens	0,6	4,00	1,50	0,02	0,04
Fit.cens	0,7	4,00	1,50	0,02	0,05
Fit.cens	0,8	4,00	1,50	0,03	0,06
Fit.trunc	0	4,00	1,50	0,01	0,02
Fit.trunc	0,1	4,00	1,50	0,02	0,04
Fit.trunc	0,2	3,99	1,50	0,03	0,05
Fit.trunc	0,3	3,99	1,50	0,04	0,06
Fit.trunc	0,4	3,98	1,50	0,06	0,07
Fit.trunc	0,5	3,98	1,50	0,08	0,09
Fit.trunc	0,6	3,97	1,50	0,12	0,11
Fit.trunc	0,7	3,91	1,51	0,20	0,16
Fit.trunc	0,8	3,47	1,57	0,82	0,30

* Fit.base – wariant bazowy. Pełny zakres danych i estymacja pełnej dziedziny; Fit.naive – estymacja naiwna. Ucięty zakres danych i estymacja pełnej dziedziny; Fit.trunc – estymacja ucięta. Ucięty zakres danych i estymacja uciętej dziedziny danych; Fit.cens – estymacja cenzurowana. Cenzurowany zakres danych i estymacja cenzurowana.

Źródło: Obliczenia własne.

Literatura

- BIS (2010), *Recognizing the Risk-Mitigating Impact of Insurance in Operational Risk Modelling*, October 28, <http://www.bis.org/publ/bcbs181.htm> (dostęp: 05.06.2016).
- BIS (2011), *Principles for the Sound Management of Operational Risk – Final Document*, June 30, <http://www.bis.org/publ/bcbs195.htm> (dostęp: 05.06.2016).
- BIS (2013), *Principles for Effective Risk Data Aggregation and Risk Reporting*, January 9, <http://www.bis.org/publ/bcbs239.htm> (dostęp: 05.06.2016).
- Chernobai A.S., Rachev S.T., Fabozzi F.J. (2007), *Operational Risk: a Guide to Basel II Capital Requirements, Models and Analysis*, John Wiley & Sons, Hoboken N.J.
- Ergashev B., Pavlikov K., Uryasev S., Sekeris E. (2015), *Estimation of Truncated Data Samples in Operational Risk Modeling*, Wiley, doi: 10.1111/jori.12062.
- Klein J.P., Moeschberger M.L. (2003), *Survival Analysis: Techniques for Censored and Truncated Data*, Springer, New York.
- Klugman S.A., Panjer H.H., Willmot G.E. (2008), *Loss Models: From Data to Decisions*, Wiley, Hoboken N.J.
- Millard S.P. (2013), *EnvStats*, Springer, New York.
- Shevchenko P.V. (2011), *Modelling Operational Risk Using Bayesian Inference*, Springer Berlin Heidelberg, Berlin.
- Szkutnik T., Basiaga K. (2013), *The Application of Generalized Pareto Distribution and Copula Functions in the Issue of Operational Risk*, "Econometrics", No. 39(1), s. 133-142.

ANALYSIS OF SOURCES OF UNCERTAINTY IN MODELLING OPERATIONAL RISK

Summary: The article focuses on the problem of identifying the sources of uncertainty in modelling operational risk based on the LDA method. Modelling operational risk in banking is based mostly on information collected above a certain threshold. Truncation of the observation as one of the specific types of censoring information in many practical cases is not taken into account when estimating the distribution parameters of the severity of operating losses in the LDA method. The article compared the different forms of censorship data together with the assessment of the degree of the impact of loss of information on the variability of the estimated parameters of the distribution and size of the risk. The proposed estimation method is not widely used in banking practice. This solution allows to improve the quality of the estimation without full reporting losses of the bank. In addition, censoring of information is a valuable source of information not only in the severity of losses, but also provides additional information in modelling frequency process of losses. Using the simulation is shown the impact of different ways of collecting information on the quality of the estimates of the distribution of severity of losses.

Keywords: operational risk, LDA, censoring information, truncation of the information.