

Cenzurowanie danych w bankowości

Ryzyko operacyjne

Tomasz Szkutnik

Uniwersytet Ekonomiczny w Katowicach
Katedra Statystyki, Ekonometrii i Matematyki

2017

Cenzurowanie danych w bankowości

Plan wystąpienia

- ▶ Cenzurowanie danych- powszechne obszary zastosowań.
- ▶ Wpływ cenzurowania na wartości niektórych charakterystyk rozkładu.
- ▶ Cenzurowanie strat operacyjnych w banku w kontekście rozwiązań regulacyjnych.
- ▶ Wnioski instytucji finansowych z kilku lat rozwijania metod AMA.
- ▶ Metoda LDA w przypadku cenzurowania informacji. Przykład symulacyjny.
- ▶ Podsumowanie.

Cenzurowanie danych

Kilka przykładów cenzurowania

Obszary zastosowań, w których powszechnie wykorzystuje się cenzurowanie danych, to:

- ▶ analiza przeżycia, teoria niezawodności
- ▶ statystyka medyczna, badania biologiczne i chemiczne
- ▶ badania środowiskowe
- ▶ badania społeczne
- ▶ przemysł

Podstawowe obszary w których cenzurowanie danych może wpływać na wyniki analiz i interpretację wyników:

- ▶ statystyka opisowa
- ▶ testowanie hipotez
- ▶ korelacja i regresja

Cenzurowanie danych

Typy cenzurowania danych

Cenzurowanie informacji występuje w przypadku, gdy zachodzące zdarzenia nie mogą być w pełni zarejestrowane.

Rejestrowany jest tylko fakt zajścia zdarzeń, bez ich ilościowego charakteru.

Typy cenzurowania danych

- ▶ cenzurowanie lewostronne, prawostronne, przedziałowe.
- ▶ pojedyncze lub wielokrotne. Jeden lub różne poziomy cenzurowania
- ▶ cenzurowanie typu I lub typu II. Liczba obserwacji jest zmienną losową lub jest znana z góry.

np. Lewostronne jednokrotne cenzurowanie Typu I. Dane są w pełni rejestrowane powyżej określonego progu. Poniżej progu raportowania/ detekcji wystąpienie jest zmienną losową.

Cenzurowanie danych

Przykłady

Przykład 1. Cenzurowanie lewostronne

Czas do wystąpienia pewnego zdarzenia dla obserwowanej jednostki jest lewostronnie cenzurowany jeśli zdarzenie to wystąpiło przed objęciem tej jednostki badaniem.

- ▶ badania ankietowe dotyczące kontaktu z używkami.
- ▶ testy psychologiczne badające zdolności do wykonania określonych zadań (czynności).

Cenzurowanie danych

Przykłady

Przykład 2. Cenzurowanie Prawostronne

Badania, w których czas zajścia zdarzenia obserwowany jest tylko, gdy wystąpi przed określonym momentem cenzurowania. Moment ten może być różny dla różnych jednostek. Ze względu na czas trwania lub koszty badanie jest przerywane. Wszystkie cenzurowane obserwacje mają taki sam czas trwania (przeżycia).

- ▶ Badania niezawodności testujące wytrzymałość, długość pracy.
- ▶ Badania toksykologiczne.

Cenzurowanie danych

Badania oparte na danych cenzurowanych muszą uwzględniać odpowiednią postać funkcji wiarygodności.

Postać funkcji wiarygodności

$$L = \prod_{i \in D} f(x_i) \prod_{i \in R} S(C_r) \prod_{i \in L} (1 - S(C_l)) \prod_{i \in I} [S(L_i) - S(R_i)] \quad (1)$$

Dla obserwacji prawostronnie cenzurowanych informacja ta wyrażona jest przez funkcję przeżycia.

Analogicznie obserwacje lewostronnie cenzurowane związana jest z wartością dystrybuanty.

Ocena wpływu cenzurowania informacji

Przykład. Rozkład logarytmiczno-normalny

Trzy warianty procesu generującego dane:

- ▶ **Pełny.** $x_i \in \mathbb{R}^+$
- ▶ **Ucięty.** $x_i > c$.
- ▶ **Cenzurowany.** Wariant ucięty + wartości poniżej progu c są zliczane.

Próg c cenzurowania i ucinania obserwacji przyjęty jest jako wartość kwantyla z wyjściowego rozkładu.

Takie warunki dla procesu cenzurowania reprezentują hipotetyczną skalę utraty informacji.

Ocena wpływu cenzurowania informacji

Warianty danych i estymacji

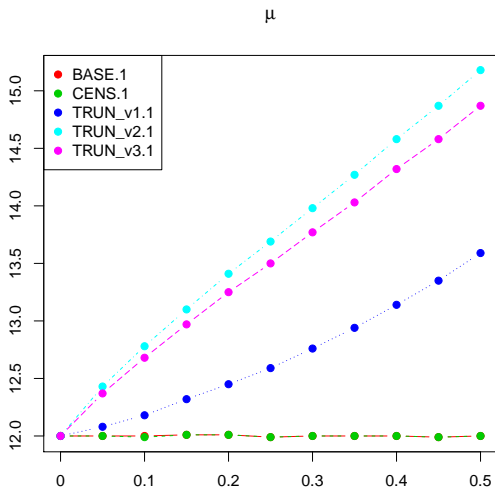
Przyjęto cztery warianty estymacji:

- ▶ **Bazowy.** Pełny zakres danych
- ▶ **Cenzurowany.** Dane cenzurowane- uwzględniony proces cenzurowania
- ▶ **Ucięty 1.** Dane ucięte- dołączone wartości ucięte zamienione na wartość progu ucinania
- ▶ **Ucięty 2.** Dane ucięte- wartości niepomniejszone o próg ucinania
- ▶ **Ucięty 3.** Dane ucięte- wartości pomniejszone o próg ucinania

Ocena wpływu cenzurowania informacji

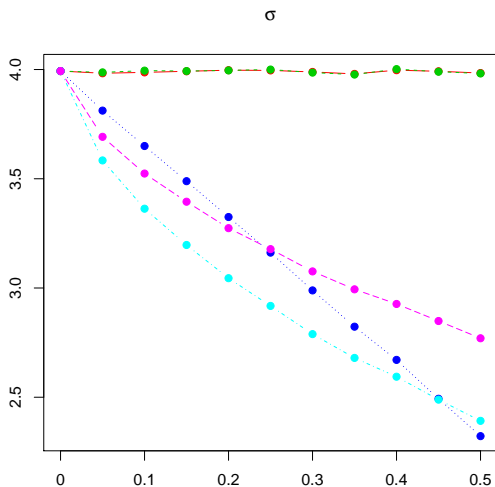
Oddziaływanie na parametry rozkładu

Oddziaływanie na parametr μ



Ocena wpływu cenzurowania informacji

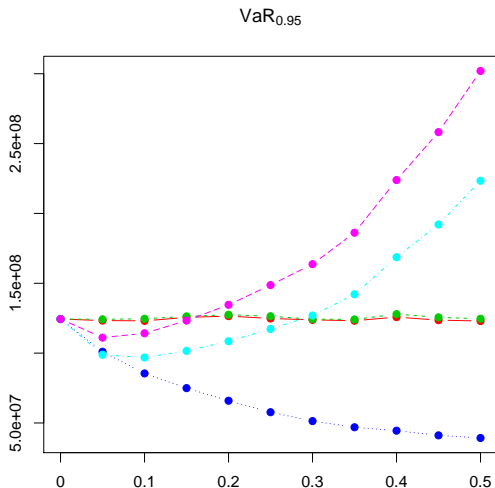
Oddziaływanie na parametr σ



Ocena wpływu cenzurowania informacji

Oddziaływanie na wartość $VaR_{0.95}$

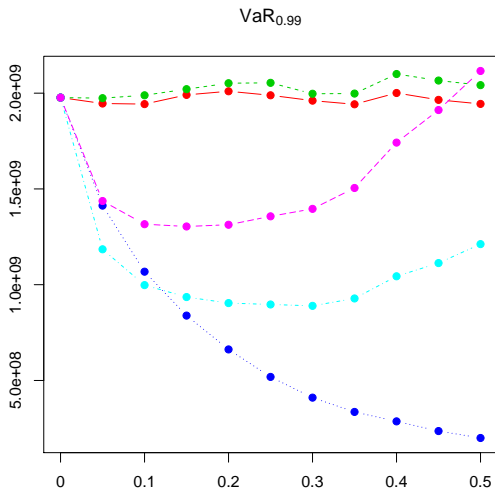
Oddziaływanie na wartość Value-at-Risk
 $VaR_{0.95}$



Ocena wpływu cenzurowania informacji

Oddziaływanie na wartość $VaR_{0.99}$

Oddziaływanie na wartość Value-at-Risk
 $VaR_{0.99}$



Wykorzystanie informacji cenzurowanej w bankowości

Cenzurowanie informacji w ryzyku operacyjnym

- ▶ Historia rozwiązań Bazylejskim dla ryzyka operacyjnego.
- ▶ Metoda LDA- doświadczenia praktyki bankowej.
- ▶ Cenzurowanie strat operacyjnych.
- ▶ Przykłady symulacyjne.

Ryzyko operacyjne w bankowości

Historia wdrażania rozwiązań Bazylejskich

Rok 2003 Bazylea II. Trzy metody pomiaru:

1. Podstawowa (BIA)
2. Standardowa (TSA)
3. Zaawansowana (AMA)- punkt docelowy dla instytucji finansowych.

Rok 2014 październik. Projekt nowej metody standardowej
Komitet Bazylejski proponuje nowe rozwiązanie w zamian za BIA
oraz TSA.

Rok 2016 marzec. Zrewidowane podejście z 2014 roku.
Metoda SMA jako jedyna forma wyliczania kapitału regulacyjnego.

Ryzyko operacyjne w bankowości

Historia wdrażania rozwiązań Bazylejskich

Rok 2003 Bazylea II. Trzy metody pomiaru:

1. Podstawowa (BIA)
2. Standardowa (TSA)
3. Zaawansowana (AMA)- punkt docelowy dla instytucji finansowych.

Rok 2014 październik. Projekt nowej metody standardowej
Komitet Bazylejski proponuje nowe rozwiązanie w zamian za BIA oraz TSA.

Rok 2016 marzec. Zrewidowane podejście z 2014 roku.
Metoda SMA jako jedyna forma wyliczania kapitału regulacyjnego.

Ryzyko operacyjne w bankowości

Historia wdrażania rozwiązań Bazylejskich

Rok 2003 Bazylea II. Trzy metody pomiaru:

1. Podstawowa (BIA)
2. Standardowa (TSA)
3. Zaawansowana (AMA)- punkt docelowy dla instytucji finansowych.

Rok 2014 październik. Projekt nowej metody standardowej
Komitet Bazylejski proponuje nowe rozwiązanie w zamian za BIA oraz TSA.

Rok 2016 marzec. Zrewidowane podejście z 2014 roku.
Metoda SMA jako jedyna forma wyliczania kapitału regulacyjnego.

Ryzyko operacyjne w bankowości

Metoda AMA

Od podejścia AMA oczekiwano się:

- ▶ poprawy zarządzania ryzykiem operacyjnym
- ▶ niższego poziomu kapitału regulacyjnego

Komitet Bazylejski dla metod AMA formułuje wskazówki na wysokim szczeblu ogólności.

Ryzyko operacyjne w bankowości

Metoda rozkładu strat LDA

Metoda LDA

Standardowo wielkości $VaR(0.999)$ wyznaczana jest ze zagregowanego rozkładu, na którego składają się dwa procesy cząstkowe:

- ▶ proces częstości występowania strat
- ▶ proces dotkliwości strat

Ograniczenia wskazywane przez instytucje finansowe:

- ▶ LDA ma retrospektywny charakter
- ▶ Ekstremalna strata lub straty mogą trwale zmienić wielkość kapitału regulacyjnego
- ▶ Konieczność skalowania i korekty strat po zaniknięciu przyczyn i wdrożeniu działań naprawczych.

Ryzyko operacyjne w bankowości

Kilka lat doświadczeń w rozwoju metod zaawansowanych

Doświadczenia instytucji finansowych i Komitetu Bazylejskiego z procesu wdrażania metod AMA:

- ▶ ograniczenia w łączeniu i wykorzystaniu wewnętrznych i zewnętrznych źródeł informacji,
- ▶ konieczność poniesienia dużych nakładów początkowych, tworzenie baz danych, tworzenie polityki w zakresie ryzyka operacyjnego,
- ▶ wyższy poziom skomplikowania niż pierwotnie zakładano,
- ▶ niedostosowanie rozwiązań AMA do prawdziwego profilu ryzyka instytucji,
- ▶ brak możliwości porównania wyników pomiędzy bankami.

Ryzyko operacyjne w bankowości

Sceptycyzm w odniesieniu do wyników metod AMA

Nadzór bankowy hamuje zbytni optymizm negując wiele z ulg kapitałowych wypracowanych przez AMA

Instytucje finansowe oczekują powiązania wyników modelu z potrzebami zarządzania. Instytucje częściowo powstrzymują się od bezpośredniego wykorzystania modeli.

Dane o stratach banku

Wpływ ucinania informacji

Proces ucinania informacji w przypadku RO:

- ▶ Nie istnieje jeden ogólnie przyjęty próg ucinania informacji
- ▶ Może być różny w odniesieniu do instytucji, procesu jakiego dotyczy, wykorzystanej bazy danych
- ▶ Może wpływać na szacunki ryzyka
- ▶ Może uniemożliwić porównywalność modeli RO.

Ocena wpływu cenzurowania informacji

Założenia badania symulacyjnego

- ▶ $VaR(0.999)$ dla zagregowanego rozkładu strat
- ▶ Proces bazowy: hipotetyczny okres 4 lat:
 - ▶ Proces częstości. Rozkład Poissona z $\lambda = 50$
 - ▶ Proces dotkliwości. Rozkład Weibulla z parametrem kształtu $\tau \in 0.5, 0.75, \dots, 3$ i parametrem skali $\theta = 4$, gdzie:
Rozkład gruboogonowy dla $\tau < 1$

Ocena wpływu cenzurowania informacji

Warianty danych i estymacji

Trzy warianty procesu generującego dane:

- ▶ **Pełny.** Straty operacyjne $x_i \in \mathbb{R}^+$
- ▶ **Ucięty.** Straty operacyjne $x_i > c$.
- ▶ **Cenzurowany.** Wariant ucięty + straty poniżej progu c są zliczane.

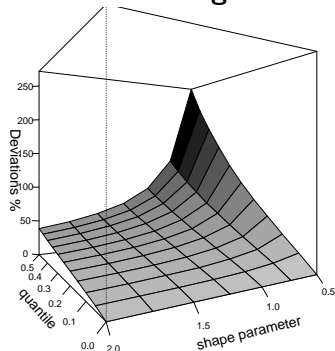
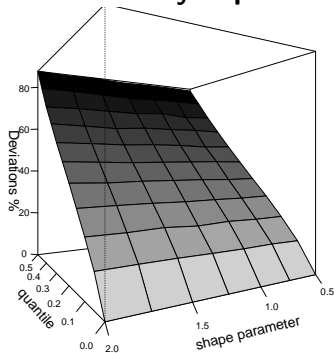
Przyjęto cztery warianty estymacji:

- ▶ **Bazowy.** Pełny zakres danych
- ▶ **Naiwny.** Dane ucięte- nieuwzględniony proces ucinania
- ▶ **Ucięty.** Dane ucięte- uwzględniony proces ucinania
- ▶ **Cenzurowany.** Dane cenzurowane- uwzględniony proces cenzurowania

Ocena wpływu cenzurowania informacji

Wyniki badania

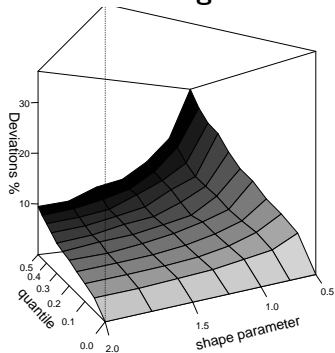
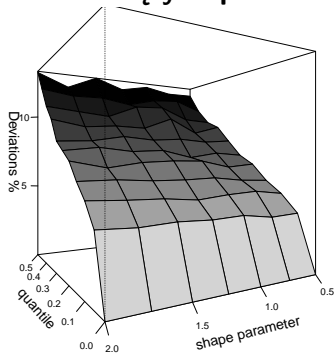
**Wpływ utraty informacji na stabilność parametrów.
Wariant Naiwny w porównaniu do wariantu bazowego**



Ocena wpływu cenzurowania informacji

Wyniki badania

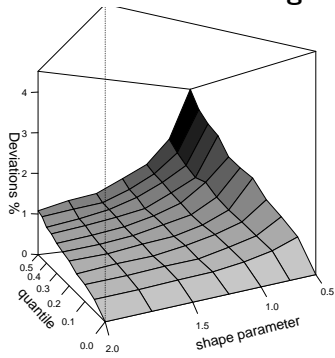
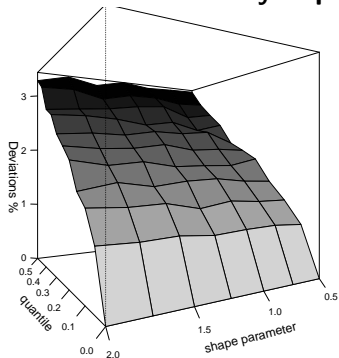
**Wpływ utraty informacji na stabilność parametrów.
Wariant Ucięty w porównaniu do wariantu bazowego**



Ocena wpływu cenzurowania informacji

Wyniki badania

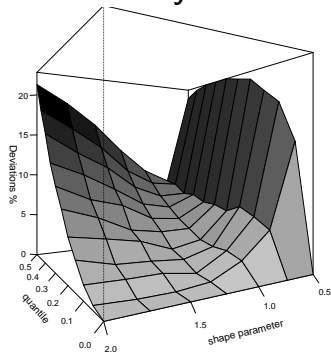
**Wpływ utraty informacji na stabilność parametrów.
Wariant Cenzurowany w porównaniu do wariantu bazowego**



Ocena wpływu cenzurowania informacji

Wyniki badania

Różnice w wielkościach miary VaR .
Wariant Naiwny do Wariantu Bazowego



Wykorzystanie informacji cenzurowanej w bankowości

Funkcja gęstości i dystrybuanta rozkładu Burra:

$$f(x) = \alpha \cdot \beta x^{-\alpha-1} (1 + x^{-\alpha})^{-\beta-1}; F(x) = (1 + x^{-\alpha})^{-\beta} \quad (2)$$

Funkcji wiarygodności w przypadku lewostronnie cenzurowanym:

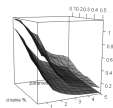
$$L(\alpha, \beta) = (1 + C_l^{-\alpha})^{-r\beta} \cdot (\alpha\beta)^{n-r} \cdot \prod_{i=r+1}^n x_{(i)}^{-\alpha-1} \cdot (1 + x_{(i)}^{-\alpha})^{-\beta-1} \quad (3)$$

lub odpowiednio logarytm z funkcji wiarygodności:

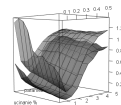
$$l(\alpha, \beta) = -r\beta \cdot \ln(1 + C_l^{-\alpha}) + (n-r) \ln \alpha\beta - (\alpha+1) \sum_{i=r+1}^n \ln(x_{(i)}) - (\beta+1) \cdot \sum_{i=r+1}^n \ln(1 + x_{(i)}^{-\alpha}) \quad (4)$$

Wykorzystanie informacji cenzurowanej w bankowości

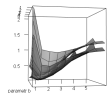
Ucinanie obserwacji



Iloraz VaR_B/VaR_U . (q/α)

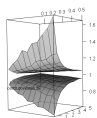


Iloraz VaR_B/VaR_U . (q/β)

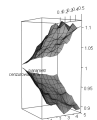


Iloraz VaR_B/VaR_U .
(α/β)

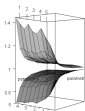
Cenzurowanie obserwacji



Iloraz VaR_B/VaR_C . (q/α)



Iloraz VaR_B/VaR_C . (q/β)



Iloraz VaR_B/VaR_C .
(α/β)

Wnioski

Rozważane warianty danych i estymacji

Wpływ przyjętego scenariusza na własności rozkładu dotkliwości strat:

- ▶ Wariant Naiwny nie pozwala na ocenę właściwego stopnia narażenia na ryzyko operacyjne.
- ▶ Wariant Ucięty zapewnia znaczną poprawę względem rozwiązania naiwnego
- ▶ Wariant Cenzurowany zapewnia najmniejszą skalę utraty informacji co przekłada się bezpośrednio na bardziej zbliżone szacunki parametrów względem rozwiązania bazowego niż poprzednie scenariusze.

Wnioski

Cenzurowanie danych

Uwzględnienie cenzurowania danych:

- ▶ Daje możliwość łączenia danych pochodzących z różnych źródeł i o różny progach raportowania.
- ▶ Upraszcza sposób rejestracji danych poniżej c na potrzeby kwantyfikacji i zarządzania ryzykiem operacyjnym.
- ▶ Daje możliwość kontroli częstości występowania strat pełnej dziedziny danych.
- ▶ Cenzurowanie może zapewnić lepszą porównywalność danych
- ▶ Brak jest ryzyka zmiany progu raportowania w celu uzyskania satysfakcjonującego wyniku
- ▶ Cenzurowanie może stanowić wyzwanie przy testowaniu dopasowania rozkładów.