

Ewa Wycinka

Modelowanie czasu do zaprzestania spłat rat kredytu lub wcześniejszej spłaty kredytu jako zdarzeń konkurujących

Problemy Zarządzania 13/3 (2), 146-157

2015

Artykuł został opracowany do udostępnienia w internecie przez Muzeum Historii Polski w ramach prac podejmowanych na rzecz zapewnienia otwartego, powszechnego i trwałego dostępu do polskiego dorobku naukowego i kulturalnego. Artykuł jest umieszczony w kolekcji cyfrowej bazhum.muzhp.pl, gromadzącej zawartość polskich czasopism humanistycznych i społecznych.

Tekst jest udostępniony do wykorzystania w ramach dozwolonego użytku.

Modelowanie czasu do zaprzestania spłat rat kredytu lub wcześniejszej spłaty kredytu jako zdarzeń konkurujących

Nadesłany: 15.10.15 | Zaakceptowany do druku: 28.12.15

Ewa Wycinka*

Ryzyko niewypłacalności i ryzyko przedwczesnej spłaty to dwa podstawowe zdarzenia powodujące przerwanie umowy kredytowej. Dla kredytodawcy istotna jest informacja, którzy kredytobiorcy i w jakim okresie od udzielenia kredytu są bardziej narażeni na każde z tych zdarzeń. W artykule zaproponowano wykorzystanie metody zdarzeń konkurujących umożliwiającej modelowanie w czasie obu powyższych zdarzeń. W części empirycznej artykułu przeprowadzono estymację na próbie 5000 pięcioletnich pożyczek, które były obserwowane przez okres 24 miesiące od udzielenia. Za pomocą wybranych estymatorów wyznaczone zostały prawdopodobieństwa analizowanych zdarzeń. Omawiane estymatory służą również do wyodrębniania kredytobiorców o różnym ryzyku niewypłacalności.

Słowa kluczowe: scoring kredytowy, zdarzenia konkurujące, analiza przeżycia.

Modelling Time to Default Or Early Repayment as Competing Risks

Submitted: 15.10.15 | Accepted: 28.12.15

Early repayment and default are two basic perils causing credit termination. Who of the borrowers and when are at the risk of both events is important information for the creditor. The article employs some of the estimators of the method of competing risks to model time to default or early repayment. The empirical part of the article consists of the results of the study on the sample of 5000 five-year loans that were observed for 24 months. Probabilities of default and early repayment have been evaluated by means of selected estimators. These estimators are also suitable for distinguishing borrowers with different risks of default.

Keywords: credit scoring, competing risks, survival analysis.

JEL: O16, C24

* **Ewa Wycinka** – dr, Katedra Statystyki, Wydział Zarządzania, Uniwersytet Gdański.

1. Wstęp

Ryzyko kredytowe jest jednym z najważniejszych rodzajów ryzyka zagrażających instytucjom finansowym. W ocenie ryzyka kredytowego ważne jest nie tylko czy kredyt zostanie spłacony w całości, lecz także czas, jaki upłynie od udzielenia kredytu do zaprzestania spłat rat kredytowych. Im większa liczba spłaconych rat, tym mniejsza jest strata kredytodawcy z tytułu udzielonego kredytu. Dodatkowym problemem przy analizie ryzyka kredytowego są kredyty przedwcześnie spłacone (Thomas, Banasik i Crook, 1999, s. 1185). Dla instytucji finansowej istotna jest identyfikacja cech charakteryzujących kredytobiorców narażonych na zaprzestanie spłat oraz tych skłonnych do wcześniejszej spłaty całości zobowiązania. Przy tak postawionym problemie zastosowanie w modelowaniu ryzyka kredytowego znajduje metoda analizy przeżycia, w szczególności analiza zdarzeń konkurujących (*competing risks*).

Celem artykułu jest przedstawienie przydatności trzech estymatorów: estymatora Kaplana-Meiera (KM), estymatora subdystrybuanty (*cumulative incidence function* – CIF lub *subdistribution*) oraz estymatora prawdopodobieństwa warunkowego (*cumulative conditional probability* – CPC) w modelowaniu ryzyka kredytowego. Estymatory te wykorzystano do oszacowania ryzyka niewypłacalności i przedwczesnej spłaty kredytu. Badanie empiryczne przeprowadzono na próbie 5000 kredytów udzielonych przez jedną z instytucji finansowych w Polsce.

2. Istota analizy przeżycia i jej zastosowania

Analiza przeżycia (trwania, historii zdarzeń) to zbiór statystycznych metod służących do badania procesów, których wspólną cechą jest ich trwanie między dwoma wyznaczonymi momentami: początkowym i końcowym (Balicki, 2006, s. 9). Analizie podlega czas trwania definiowany jako nieujemna zmienna losowa $T \geq 0$. Czas trwania od momentu początkowego do końcowego określany jest jako kompletny, natomiast jeśli dojdzie do przerwania obserwacji przed wystąpieniem zdarzenia końcowego, to mamy do czynienia z danymi cenzurowanymi (Balicki, 2006, s. 71–78).

Metoda analizy przeżycia została najszerzej rozwinięta w biostatystyce. Wykorzystywana jest również w innych naukach, w tym w technice (analiza niezawodności), demografii, ekonomii i finansach (Balicki, 2006, s. 20–25).

Narain (1992) jako pierwszy zaproponował wykorzystanie analizy przeżycia do budowy modeli scoringowych oceny ryzyka kredytowego. Zastosował on wykładniczy akceleracyjny model do opisu trwania 24-miesięcznych pożyczek i wykazał, że model ten daje lepsze oszacowanie ryzyka niewypłacalności niż standardowe modele zmiennej dychotomicznej (w tym modele logitowe). Ideę tę rozwinęli Thomas, Banasik i Crook (1999). Zwrócili oni uwagę między innymi na konieczność analizowania nie tylko ryzyka niewypłacalności, lecz także innych przyczyn rozwiązania umowy kredytowej. Jako pierwsi zaproponowali traktowa-

nie tych zdarzeń jako konkurujących i rozłączne analizowanie każdego z nich przy założeniu, że pozostałe zdarzenia traktowane są jako dane cenzurowane. Autorzy analizowali ponadto wykorzystanie modeli o rozkładach wykładniczych, Weibulla oraz modeli Coxa do badania niewypłacalności, a także przedwczesnej spłaty w dwóch okresach: pierwszych 12 miesiącach trwania kredytów oraz w okresie 12–24 miesięcy, pod warunkiem, że kredyty są czynne w 12 miesiącu (Thomas, Banasik i Crook, 1999). Badanie empiryczne było przeprowadzone na próbie 50 000 pożyczek udzielonych przez jedną z brytyjskich instytucji finansowych w latach 1994–1997. Tę samą metodę wykorzystwała Matuszyk (2015), przeprowadzając badanie empiryczne na próbie 6135 kredytów hipotecznych wybranych z bazy z brytyjskiego systemu bankowego. Modele Coxa budowane oddzielnie dla ryzyka niewypłacalności i przedwczesnej spłaty były także przedmiotem badań m.in. Stepanova i Thomas (2002), Glennon i Nigro (2005), Andreeva (2006) czy Bellotti i Crook (2009).

Inne podejście do modelowania ryzyka kredytowego zastosowali Watkins, Vasnev i Gerlach (2014) oraz Tong, Mues i Thomas (2012). Argumentują oni, że odrębne analizowanie ryzyka niewypłacalności i ryzyka przedwczesnej spłaty jest błędne i jako alternatywę proponują modelowanie ryzyka obu zdarzeń jednocześnie za pomocą mieszaniny rozkładów trwania (*mixture cure models*).

W niniejszym artykule zaproponowano inną metodę jednoczesnego badania ryzyka niewypłacalności i ryzyka przedwczesnej spłaty. Prawdopodobieństwa tych zdarzeń oszacowano za pomocą estymatorów subdystrybuanty i prawdopodobieństwa warunkowego. Estymatory te pozwalają na oszacowanie ryzyka niewypłacalności z jednoczesnym uwzględnieniem wpływu ryzyka przedwczesnej spłaty. Jak zostanie dalej wykazane, dodatkową zaletą tych estymatorów jest prostota, łatwość obliczania i interpretacji wyników. Omawiane estymatory są szeroko wykorzystywane w biostatystyce (Pintilie 2006; Kleinbaum i Klein, 2005), w literaturze nie ma wyników badań prezentujących ich zastosowanie do oceny ryzyka kredytowego.

3. Czas trwania kredytu jako zmienna losowa

Czas trwania kredytu od momentu jego udzielenia do zakończenia spłat będzie traktowany jako zmienna losowa. Zakończenie spłacania rat może nastąpić na skutek wystąpienia po raz pierwszy jednego ze zdarzeń:

- przeterminowania 90 dni w spłacie kredytu (niewypłacalność);
- przedwczesnej całkowitej spłaty kredytu (przedwczesna spłata);
- zmiany treści umowy kredytowej (np. przewalutowanie) i/lub okresu trwania kredytu;
- innego zdarzenia powodującego przerwanie spłat rat według ustalonego harmonogramu.

W niniejszym artykule analizowane są tylko dwa pierwsze z wymienionych zdarzeń, jednak prezentowane metody mogą być uogólnione na $p \geq 2$ zdarzeń konkurujących.

Niech T będzie nieujemną zmienną losową opisującą czas trwania kredytu, taką, że $T \geq 0$. Spłaty kredytu następują najczęściej w ratach miesięcznych, co powoduje, że czas spłaty kredytu powinien być modelowany jako zmienna dyskretna. Rozkład zmiennej losowej T może być opisany przez funkcję trwania $S(t)$, zdefiniowaną jako:

$$S(t) = P(T > t) = 1 - F(t),$$

gdzie $F(t)$ jest dystrybuantą rozkładu.

Funkcja trwania wyraża prawdopodobieństwo, że kredyt pozostanie czynny co najmniej do czasu t , natomiast funkcja dystrybuanty wyraża prawdopodobieństwo, że spłaty rat kredytu zostaną przerwane najpóźniej do czasu t . Wyznaczając te prawdopodobieństwa, należy wziąć pod uwagę to, iż w kolejnych momentach czasu, na które przypadają spłaty kolejnych rat, liczba kredytów zagrożonych w portfelu zmniejsza się zarówno na skutek wystąpienia we wcześniejszych okresach niewypłacalności, jak i przedwczesnych spłat całkowitych. W klasycznym podejściu do analizy przeżycia modelowany jest tylko jeden rodzaj zdarzenia, np. niewypłacalność. Zdarzenie takie jest wówczas określane mianem zdarzenia kompletnego. Wszystkie inne zdarzenia powodujące zakończenie obserwacji przed wystąpieniem zdarzenia kompletnego są traktowane jako obserwacje cenzurowane. Zakończenie spłat rat kredytu zgodnie z harmonogramem będzie również traktowane jako obserwacja cenzurowana. Analogicznie, modelowanie rozkładu przedwczesnych spłat całkowitych będzie traktowane jako kompletne, a wszystkie pozostałe jako cenzurowane. Podejście to było wykorzystywane we wspomnianych wcześniej pracach Naraina (1992), Thomasa i in. (1999).

Najczęściej stosowanym estymatorem funkcji trwania jest estymator Kaplana-Meiera (KM) postaci:

$$\hat{S}(t) = KM = \prod_{j=1}^t \left(\frac{n_j - d_j}{n_j} \right),$$

gdzie:

- n_j – liczba czynnych kredytów bezpośrednio przed momentem t_j ,
- d_j – liczbą kredytów przeterminowanych w momencie t_j ,
- $j = 1, \dots, m$ – momenty czasu w których powinny nastąpić płatności rat.

Powyższy estymator wyznaczany jest dla każdego momentu, w którym wystąpiło zdarzenie kompletne. Estymator Kaplana-Meiera modeluje prawdopodobieństwo zdarzenia przy założeniu, że jest ono jedynym możliwym zdarzeniem. Nie można więc jednocześnie analizować rozkładu prawdopo-

dobieństw wystąpienia niewypłacalności i przedwczesnej spłaty. Suma tych prawdopodobieństw nie jest prawdopodobieństwem.

Estymator KM zakłada ponadto, że cenzurowanie jest losowe tzn. rozkład czasu cenzurowania jest niezależny od rozkładu trwania (tzw. cenzurowanie nieinformacyjne). Kleinbaum i Klein (2005) podkreślają, że mimo iż w literaturze przedmiotu prezentowane są różne metody badania założenia nieinformacyjności cenzurowania, to żadna z nich nie pozwala jednoznacznie potwierdzić lub zaprzeczyć, że założenie nieinformacyjności cenzurowania jest spełnione. Stosowanie estymatora K-M w przypadku niespełnienia założenia nieinformacyjności cenzurowania prowadzi do obciążonych wyników – niedoszacowania lub przeszacowania prawdopodobieństw.

Założenie nieinformacyjności cenzurowania wydaje się być niespełnione w modelowaniu czasu trwania kredytu. Przedwczesna spłata jest zdarzeniem, które nie tylko uniemożliwia obserwację kredytu do czasu jego niewypłacalności, lecz także jest zdarzeniem konkurującym. Przedwczesna spłata może być wynikiem bardzo dobrej kondycji finansowej kredytobiorcy, niewypłacalność związana jest zaś ze złą kondycją finansową dłużnika. Niewypłacalność i przedwczesna spłata jako zdarzenia kończące trwanie kredytu, mogą być jednak modelowane jednocześnie przy wykorzystaniu modeli zdarzeń konkurujących, które nie wymagają spełnienia założenia nieinformacyjności cenzurowania.

4. Niewypłacalność i przedwczesna spłata jako zdarzenia konkurujące

Przyjmijmy, że trwaniu kredytu zagrażają dwa zdarzenia: niewypłacalność i przedwczesna spłata, z tym że jednocześnie wystąpić może tylko jedno z nich i wyklucza ono wystąpienie drugiego. W badanym zagadnieniu analizowane są dwa zdarzenia konkurujące. Prezentowana poniżej metoda umożliwia również uogólnienie rozważań na większą liczbę zdarzeń konkurujących. Przykładowo, jeżeli kredytodawca posiada informacje o przyczynach wcześniejszych spłat, to każda z nich może być traktowana jako odrębne zdarzenie konkurujące, np. wcześniejsza spłata ze środków kredytobiorcy, konsolidacja kredytów, zmiana warunków kredytu.

Niech C będzie zmienną losową skokową informującą o realizacji i -tego zdarzenia konkurującego, $i = 1, \dots, p$. Subdystrybuanta (*cumulative incidence function* lub *subdistribution*) zdarzenia i wyrażona jest wzorem:

$$F_i(t) = P(T \leq t, C = i)$$

i określa prawdopodobieństwo, że zdarzenie i wystąpi przed czasem t . Suma subdystrybuant dla wszystkich zdarzeń konkurujących jest równa dystrybuancie rozkładu:

$$F(t) = \sum_{i=1}^p F_i(t).$$

Należy mieć na uwadze, że dystrybuanta rozkładu przyjmuje wartości z przedziału $[0; 1]$, natomiast subdystrybuanta przyjmuje wartości z przedziału $[0; P(C = i)]$, gdzie $P(C = i)$ jest udziałem zdarzeń i w badanej próbie (Pintilie, 2006, s. 45). Subdystrybuanta jest oszacowaniem brzegowego prawdopodobieństwa wystąpienia zdarzenia przy uwzględnieniu zdarzeń konkurujących, nie wymaga więc spełnienia założenia niezależności zdarzeń konkurujących.

Estymator subdystrybuanty ma postać:

$$\hat{F}_i(t) = CIF = \sum_{j \leq t} \frac{d_{ij}}{n_j} \hat{S}(t_{j-1}),$$

gdzie:

- d_{ij} – jest liczbą zdarzeń typu i w czasie t_j ,
- n_j – liczbą jednostek narażonych w momencie t_j ,
- $\hat{S}(t_{j-1})$ – estymatorem Kaplana-Meiera wyznaczonym dla wszystkich zdarzeń konkurujących rozważanych łącznie.

Zachodzi własność: $\hat{F}_i(t) \leq 1 - \hat{S}(t)$, przy czym równość ma miejsce tylko, gdy występuje jedno zdarzenie (brak zdarzeń konkurujących). Suma subdystrybuant dla niewypłacalności oraz wcześniejszej całkowitej spłaty w każdym momencie t jest równa prawdopodobieństwu przerwania spłat (z jakiegokolwiek przyczyny) wyznaczonemu jako dopełnienie do jedności estymatora KM.

Zdarzenia konkurujące mogą być również modelowane za pomocą prawdopodobieństwa warunkowego (*cumulative conditional probability* – CPC). Wyraża ono prawdopodobieństwo zajścia zdarzenia typu i do czasu t , pod warunkiem, że jednostka nie doznała żadnego ze zdarzeń konkurujących do tego czasu (Kleinbaum i Klein, 2005, s. 420):

$$CPC_i = P(T_i \leq t | T_p \geq t),$$

gdzie:

- T_p – jest czasem do wystąpienia jakiegokolwiek ze zdarzeń konkurujących.

Prawdopodobieństwo warunkowe może być oszacowane za pomocą wzoru:

$$CPC_i = \frac{CIF_i}{1 - CIF_p},$$

gdzie:

- CIF_p – jest subdystrybuantą wyznaczoną dla wszystkich pozostałych zdarzeń konkurujących razem wziętych. Wartość CPC_i zmienia się nie tylko wtedy gdy wystąpi zdarzenie i , lecz także na skutek wystąpienia zdarzeń konkurujących (Pintilie, 2006, s. 128). Prawdopodobieństwo warunkowe jest obliczane jako uzupełnienie CIF, dlatego że wielkość

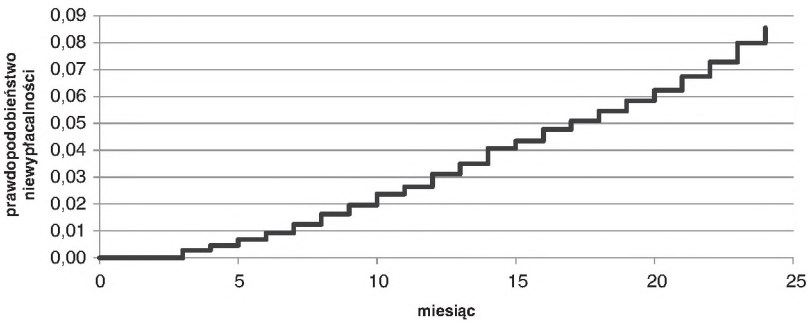
CIF dla badanego zdarzenia może być niska tylko ze względu na to, iż CIF dla zdarzenia konkurującego jest wysokie.

Różnice w przebiegu krzywych wyznaczonych przez powyższe estymatory dla grup kredytobiorców mogą być wykorzystane do identyfikacji kredytobiorców o najwyższym ryzyku wystąpienia badanych zdarzeń. Wśród testów statystycznych weryfikujących hipotezę o równości dwóch lub więcej krzywych trwania najpopularniejszy jest test *log-rank* (test Mantela-Henszela). Test opiera się na różnicy między obserwowaną a oczekiwaną liczbą zdarzeń kompletnych we wszystkich momentach. Statystyka testowa ma rozkład chi-kwadrat o $k - 1$ stopniach swobody, gdzie k jest liczbą porównywanych krzywych (Klein i Moeschberger, 2003, s. 208). Test ten wykorzystywany jest do badania różnicy między dystrybuantami lub krzywymi trwania wyznaczonymi za pomocą estymatora KM, gdy analizowany jest tylko jeden rodzaj zdarzenia. W przypadku zdarzeń konkurujących, równość subdystrybuant weryfikuje test Graya (1988). Statystyka testowa ma również rozkład chi-kwadrat o $k - 1$ stopniach swobody. Test Graya jest odpowiednikiem testu *log-rank* dla sytuacji zdarzeń konkurujących. Różnice w ocenie istotności w testach *log-rank* i Graya wskazują na wpływ zdarzeń konkurujących na zdarzenie badane (Pintilie, 2006, s. 7). Inne testy zbudowane w oparciu o analizowane estymatory opisane są w pracach: Klein i Moeschberger (2003), Pintilie (2006) oraz Balakrishnan i Rao (2004).

5. Charakterystyka próby i wyniki badań empirycznych

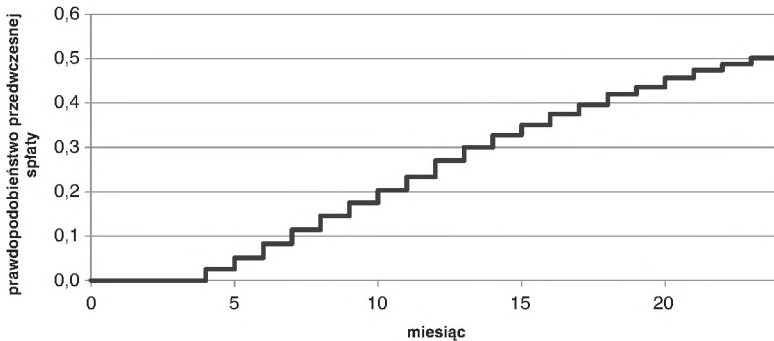
Analizę empiryczną przeprowadzono na próbie pięciu tysięcy 60-miesięcznych pożyczek udzielonych osobom fizycznym przez jedną z instytucji finansowych w Polsce. Każda pożyczka była obserwowana przez okres pierwszych 24 miesięcy trwania umowy lub do momentu przedwczesnej całkowitej spłaty lub niewypłacalności (*default*). Niewypłacalność została zdefiniowana jako co najmniej 90-dniowe przeterminowanie spłaty raty. W badanym okresie 90-dniowe przeterminowanie w spłacie kredytu wystąpiło co najmniej 1 raz w 297 umowach (5,9%). Przedwcześnie spłaconych zostało 2429 umów (48,58%). Pozostałe 2274 umowy (45,5%) były czynne w 24 miesiącu i nie wystąpiło do tego czasu przeterminowanie 90-dniowe. Charakterystyka próby została szerzej opisana w pracy Wycinka (2015).

W badanej próbie pierwszym momentem w którym mogła wystąpić niewypłacalność był 3 miesiąc, a ostatnim 24 miesiąc. W każdym z pośrednich miesięcy wystąpiła co najmniej jedna niewypłacalność, stąd wyznaczona dystrybuanta ma kształt krzywej schodkowej z równoodległymi skokami występującymi w każdym miesiącu. Na rysunku 1 przedstawiono oszacowanie dystrybuanty wyznaczonej za pomocą dopełnienia do jedności estymatora KM. Krzywą tę można interpretować następująco: prawdopodobieństwo niewypłacalności dla kredytu w ciągu pierwszych 12 miesięcy wynosi 0,031 (3,1%), a w pierwszych 24 miesiącach 0,0798 (7,98%).



Rys. 1. Prawdopodobieństwo niewypłacalności oszacowane za pomocą dopełnienia do jedności estymatora Kaplana-Meiera. Źródło: opracowanie własne.

Przy modelowaniu prawdopodobieństwa przedwczesnej spłaty kredytu wystąpienie niewypłacalności jest traktowane jako zdarzenie cenzurowane. Oszacowania prawdopodobieństwa przedwczesnej spłaty za pomocą dopełnienia do jedności estymatora K-M przedstawia rysunek 2.

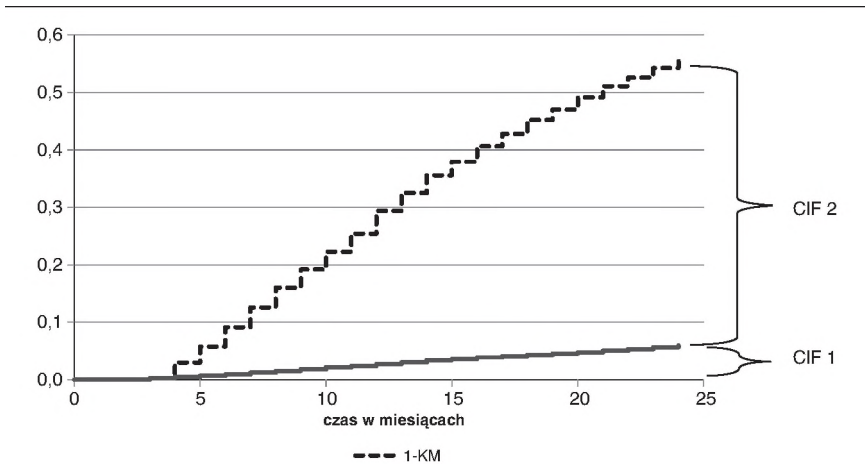


Rys. 2. Prawdopodobieństwo przedwczesnej spłaty kredytu oszacowane za pomocą dopełnienia do jedności estymatora Kaplana-Meiera. Źródło: opracowanie własne.

Przedwczesne spłaty występowały od 4 miesiąca. Prawdopodobieństwo, że do 24 miesiąca dojdzie do przedwczesnej spłaty wynosi 0,52 (52%). Interpretując wyniki, należy jednak pamiętać, że obie powyższe krzywe zostały oszacowane przy założeniu, że badane zdarzenie jest jedynym możliwym, a zdarzenie konkurujące nie istnieje.

W każdym portfelu kredytowym mogą wystąpić zarówno niewypłacalność, jak i przedwczesne całkowite spłaty kredytów dlatego większą wartość praktyczną ma oszacowanie subdystrybuant dla obu zdarzeń konkurujących. Oszacowania prawdopodobieństwa niewypłacalności oraz przedwczesnej

splaty za pomocą estymatorów CIF zostały przedstawione na rysunku 3. Suma tych subdystrybuant jest równa dopełnieniu do jedności estymatora KM wyznaczonego dla obu zdarzeń łącznie.



CIF 1 – subdystrybuanta dla zdarzenia „niewypłacalność”,
 CIF 2 – subdystrybuanta dla zdarzenia „przedwczesna spłata”.

Rys. 3. Prawdopodobieństwo przedwczesnej spłaty oraz niewypłacalności oszacowane za pomocą estymatora subdystrybuanty. Źródło: opracowanie własne.

Przykładowo, prawdopodobieństwo, że umowa kredytowa zostanie przewzrana do 24 miesiąca wynosi 0,5624 (56,24%) i składa się na nie prawdopodobieństwo niewypłacalności – 0,0594 (5,94%) oraz prawdopodobieństwo przedwczesnej spłaty – 0,503 (50,3%).

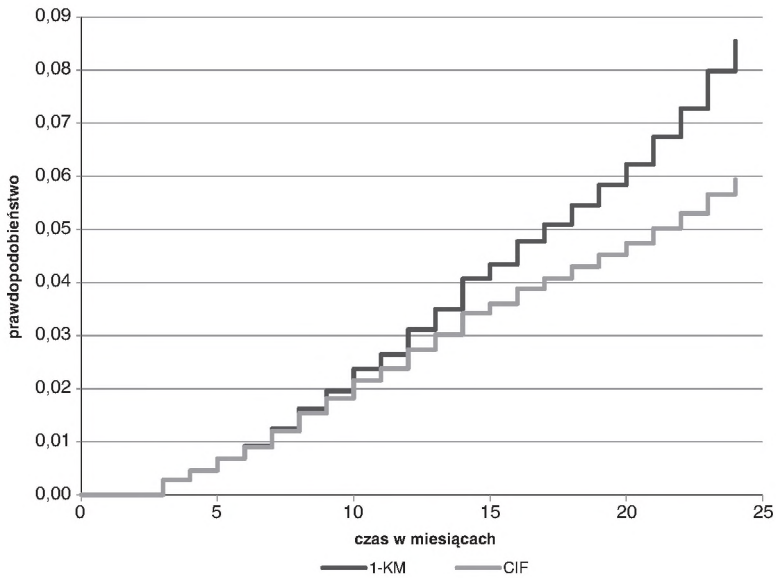
Zestawienie oszacowań prawdopodobieństw niewypłacalności uzyskanych za pomocą estymatorów KM oraz CIF przedstawiono na rysunku 4.

Różnica między krzywymi wynika z wpływu zdarzenia konkurującego na prawdopodobieństwo niewypłacalności.

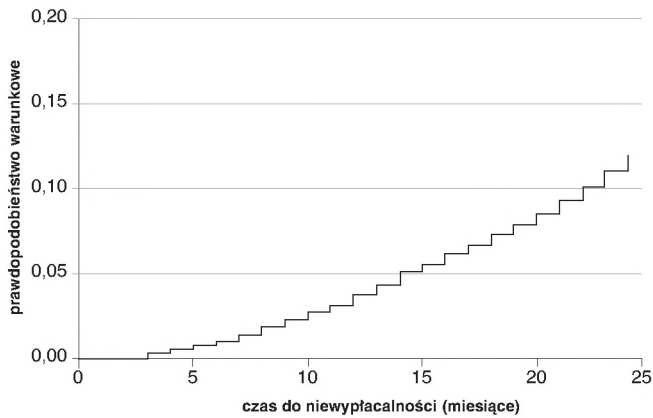
Dodatkowo prawdopodobieństwo warunkowe niewypłacalności pod warunkiem, że nie doszło wcześniej do spłacenia kredytu oszacowane za pomocą estymatora CPC przedstawiono na rysunku 5.

Prawdopodobieństwo, że do 24 miesiąca wystąpi niewypłacalność, pod warunkiem, że kredyt nie zostanie przedwcześnie spłacony wynosi 0,1195 (11,95%).

Wszystkie wymienione powyżej estymatory mogą służyć do wyodrębniania grup kredytobiorców różniących się poziomem ryzyka niewypłacalności. Cechy demograficzno-ekonomiczne kredytobiorców różnicujące ich ryzyko niewypłacalności oraz przedwczesnej spłaty były analizowane przez autorkę (Wycinka, 2015). Do pokazania różnic w przebiegu krzywych i możliwości



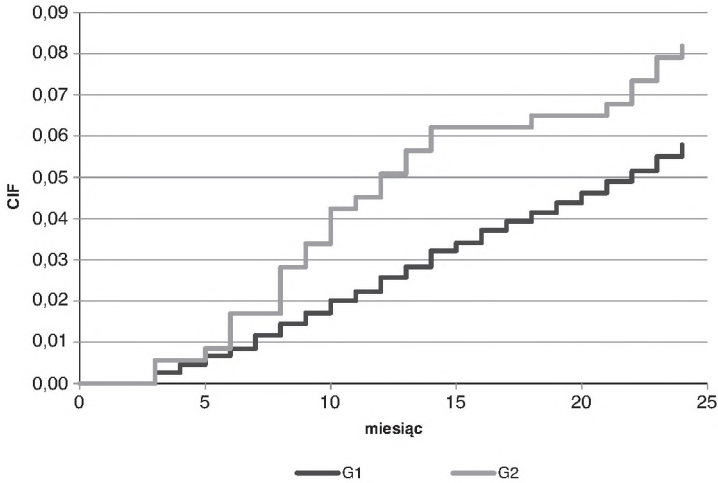
Rys. 4. Oszacowania prawdopodobieństwa niewypłacalności za pomocą estymatora subdystrybuanty oraz estymatora Kaplana-Meiera. Źródło: opracowanie własne.



Rys. 5. Oszacowanie warunkowego prawdopodobieństwa niewypłacalności pod warunkiem, że nie doszło wcześniej do spłaty kredytu. Źródło: opracowanie własne.

ich statystycznej oceny za pomocą wymienionych wcześniej testów wybrana została zmienna dychotomiczna opisująca status majątkowy kredytobiorcy oznaczona w przywołanym wcześniejszym badaniu jako X14. Prawdopodobieństwo niewypłacalności oszacowane za pomocą estymatora CIF dla

dwóch grup kredytobiorców wyodrębnionych według wariantów tej cechy (grupy G1 i G2) przedstawiono na rysunku 6.



Rys. 6. Oszacowanie prawdopodobieństwa niewypłacalności w dwóch grupach kredytobiorców. Źródło: opracowanie własne.

Kredytobiorcy należący do grupy pierwszej (G1) charakteryzują się niższym ryzykiem niewypłacalności niż kredytobiorcy z grupy drugiej (G2), ponadto w grupie drugiej intensywność ryzyka nie jest stała. Ryzyko niewypłacalności dla tej grupy w pierwszych miesiącach trwania kredytu jest wysokie (duże skoki krzywej), natomiast między 15 a 21 miesiącem praktycznie nie występuje. W grupie pierwszej ryzyko jest stałe w czasie (równe przyrosty krzywej). Krzywe dla obu grup nie różnią się jednak w sposób statystycznie istotny (test LR $p = 0,0708$; test Graya $p = 0,0632$).

6. Wnioski

Modele wykorzystujące metody analizy przeżycia umożliwiają monitorowanie ryzyka nie tylko poszczególnych umów kredytowych, lecz także całych portfeli kredytowych od ich udzielenia przez cały okres trwania. Możliwa jest identyfikacja okresów o różnym prawdopodobieństwie niewypłacalności oraz ryzyka przedwczesnej spłaty. W zależności od wybranego estymatora analizowane może być prawdopodobieństwo pojedynczego zdarzenia przy założeniu, że inne rodzaje ryzyka nie występują lub prawdopodobieństwo zdarzenia przy uwzględnieniu zdarzeń konkurujących. Estymator K-M może służyć do szacowania prawdopodobieństwa niespłacenia kredytu jako jedyne ryzyka przy założeniu, że nie występują zdarzenia konkuru-

jące. Estymator CIF jest narzędziem do szacowania prawdopodobieństwa niespłacenia kredytu przy uwzględnieniu ryzyka przedwczesnej całkowitej spłaty. Dodatkową informację dostarcza estymator prawdopodobieństwa warunkowego CPC. Wszystkie z tych estymatorów można wykorzystać do badania prawdopodobieństw zarówno niewypłacalności, jak i przedwczesnej całkowitej spłaty kredytu w różnych grupach kredytobiorców. Umożliwia to identyfikację cech kredytobiorców charakteryzujących się najwyższym ryzykiem każdego z analizowanych zdarzeń. Jeżeli zdarzenia konkurujące są niezależne, to do oceny istotności różnic krzywych prawdopodobieństwa można stosować test *log-rank* (LR) lub test Graya. Gdy zdarzenia są zależne, powinien być wykorzystywany test Graya. Reasumując, analiza przeżycia dostarcza użytecznych narzędzi do szacowania prawdopodobieństwa przerwania kredytu, które mogą wspomóc proces decyzyjny zarówno na etapie udzielania kredytów, jak i monitorowania portfeli kredytowych.

Bibliografia

- Andreeva, G. (2006). European Generic Scoring Models Using Survival Analysis. *The Journal of the Operational Research Society*, 57(10), 1180–1187.
- Balakrishnan, N. i Rao, C.R. (2004). *Advances in Survival Analysis*. Elsevier.
- Balicki, A. (2006). *Analiza przeżycia i tablice wymieralności*. Warszawa: PWE.
- Bellotti, T. i J. Crook, J. (2009). Credit scoring with macroeconomic variables using survival analysis. *Journal of the Operational Research Society*, 60, 1699–1707.
- Glennon, D. i Nigro P. (2005). Measuring Default Risk of Small Business Loans: A Survival Analysis Approach. *Journal of Money, Credit and Banking*, 37(5), 923–947.
- Gray, R. (1988). A Class of K-Sample Tests for Comparing the Cumulative Incidence of a Competing Risk. *The Annals of Statistics*, 16, 1141–1154.
- Kleinbaum, D. i Klein, M. (2005). *Survival Analysis. A Self-Learning Text*. Springer.
- Klein, J. i Moeschberger, M. (2003). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.
- Matuszyk, A. (2015). *Zastosowanie analizy przetrwania w ocenie ryzyka kredytowego klientów indywidualnych*. Warszawa: CeDeWu.
- Narain, B. (1992). Survival Analysis and the Credit Granting Decision. W: L.C. Thomas, J.N. Crook, D.B. Edelman (red.), *Credit Scoring and Credit Control*, (s. 109–121). Oxford: Oxford University Press.
- Pintilie, M. (2006). *Competing risks. A Practical Perspective*. J.Wiley & Sons.
- Stepanova, M. i Thomas, L.C. (2002). Survival Analysis Methods for Personal Loan Data. *The Journal of the Operational Research Society*, 50(2), 277–289.
- Thomas, L., Banasik, J. i Crook, J. (1999). Not if but when will borrowers default. *Journal of the Operational Research Society*, 50(12), 1185–1190.
- Tong, E., Mues, Ch. i Thomas, L. (2012). Mixture cure models in credit scoring: if and when borrowers default. *European Journal of Operational Research*, 218(1), 132–139, <http://dx.doi.org/10.1016/j.ejor.2011.10.007>.
- Watkins, J., Vasnev, A. i Gerlach, R. (2014). Multiple Event Incidence and Duration Analysis for Credit Data Incorporating Non-Stochastic Loan Maturity. *Journal of Applied Econometrics*, 29(4), 627–648, <http://dx.doi.org/10.1002/jae.2329>.
- Wycinka, E. (2015). Time to default analysis in personal credit scoring. *Research Papers of Wrocław University of Economics*, 381, 527–536, <http://dx.doi.org/10.15611/pn.2015.381.381>.