

**Marek Gruszczyński**

# Błędy doboru próby w badaniach bankructw przedsiębiorstw

**Ilościowe badania nad bankructwem i zagrożeniem finansowym w Polsce mają ponad 20-letnią historię. Od pierwszych światowych prac na ten temat dzieli nas kolejne 30 lat. Badania na poziomie międzynarodowym nadal nas wyprzedzają, przede wszystkim dlatego, że tworzy je bezpośrednio środowisko badaczy z zakresu finansów przedsiębiorstwa oraz rachunkowości.**

W Polsce takie rozumienie badań empirycznych w finansach i rachunkowości nie jest częste, wartościowe badania z tego zakresu są domeną metodyków (statystyków, ekonometryków, analityków danych itd.). Co więcej, jest rzadkością komunikowanie wyników tych badań w języku angielskim i uczestnictwo w światowym dyskursie naukowym. Chodzi mi na przykład o takie wiodące czasopisma jak: "Accounting Review" (45 punktów na liście A), "Journal of Business, Finance and Accounting" (20 pkt), "Journal of Empirical Finance" (25 pkt), "Journal of Corporate Finance" (40 pkt), "Contemporary Accounting Research" (40 pkt), "Journal of Accounting Research" (45 pkt), "Journal of Accounting and Economics" (50 pkt), "Review of Accounting Studies" (35 pkt). Rzadko też jesteśmy obecni w międzynarodowych czasopismach nieco niższej rangi.

Mamy olbrzymią literaturę krajową z zakresu badania upadłości, bankructwa i zagrożenia finansowego, niespecjalnie

znaną na świecie. Jednocześnie, popełniamy w tych badaniach rozmaite błędy metodyczne, co sami sobie często wytykamy, i co jest także przedmiotem rozważań w tym opracowaniu. Artykuł jest poświęcony jednemu z problemów metodycznych, występujących przy korzystaniu z klasycznych modeli upadłości, to jest modeli szacowanych w oparciu o nielosowe próby przedsiębiorstw. Chodzi o błędy wynikające z doboru próby i ich wpływ na szacowanie prawdopodobieństwa upadłości.

W kolejnych częściach przedstawione zostały uwagi na temat doboru próby w badaniach nad bankructwem w Polsce, komentarz o możliwych błędach związanych z doбором próby, metoda korekty logitowego modelu bankructwa, ogólny związek pomiędzy obciążonym i nieobciążonym prawdopodobieństwem bankructwa pojedynczego przedsiębiorstwa, a także warianty tej zależności dla modeli innych niż logitowy. Artykuł kończą uwagi o modelach klasyfikacji oraz podsumowanie.

## **Dobór próby w badaniach nad bankructwem i zagrożeniem finansowym w Polsce**

Doniesienia o wynikach badań bankructwa lub zagrożenia finansowego przedsiębiorstw w Polsce ukazują się często lub bardzo często. Poniżej przedstawiono kilka ostatnich pozycji mono-

graficznych – z komentarzem na temat doboru próby badawczej.

W książce pod redakcją A. Adamskiej i E. Mączyńskiej [2013] część metodyczna zawiera cztery teksty: W. Rogowski, K. Duleba [2013], M. Ciszewska [2013], M. Tymoszuć [2013], P. Antonowicz, M. Zwierzyńska [2013]. W pierwszym z nich [Rogowski, Duleba, 2013] posłużono się „klasycznymi” modelami badania upadłości w Polsce dla danych ze spółek Giełdy Papierów Wartościowych w Warszawie (GPW), które zgłosiły wniosek o upadłość w latach 2005-2010. Badanie pokazało niezłą jakość znanych modeli dyskryminacyjnych dla całkiem nowych danych. W swoim tekście M. Ciszewska [2013] analizuje 16 spółek z GPW, w tym osiem, które ogłosiły postępowanie upadłościowe lub upadłość w roku 2010 oraz osiem, które nie ogłosiły. Dla tych 16 obserwacji badano normalność rozkładu zmiennych objaśniających w potencjalnym modelu upadłości, a następnie zastosowano analizę dyskryminacyjną. Pomijając inne wady tego badania, trudny do akceptacji jest po prostu sam dobór próby oraz jej liczebność. W tekście M. Tymoszuć [2013], własny model autorki oparty jest na założeniu, że do próby trafiają spółki, które ogłosiły upadłość w latach 2007-2010 oraz mają aktywa w przedziale 30 mln – 435 mld zł. Dołączono do nich w proporcji 50:50 spółki „zdrowe”, łącznie próba liczy 68 spółek. Przyporządkowanie nie-bankrutów do grupy bankrutów odbyło się w sposób losowy. Autorka wskazuje wprawdzie, że udział bankrutów w populacji jest równy 18 proc., ale twierdzi (za B. Prusakiem, 2005), że próba o strukturze odpowiadającej populacji wszystkich przedsiębiorstw daje podobne efekty jak próba zbilansowana. To nie jest do końca prawda, ale warto pochwalić za zwrócenie uwagi na tę kwestię. Przy okazji, nie bardzo podoba mi się tytuł tego tekstu, kuriozalny jak na tytuł artykułu naukowego.

W książce P. Antonowicza [2015] zamieszczono obszerny obraz polskich badań nad upadłością, w tym także uwagi metodyczne autora. Te uwagi, na przykład, wskazują na jakość danych o firmach zaliczanych w próbach badawczych do przedsiębiorstw upadłych. W samej książce, w jej części metodycznej, autor nie poświęca uwagi tematowi doboru przedsiębiorstw do prób badawczych. Koncentruje się na wnikliwym opisie własnej próby bankrutów. Ta właśnie próba, której początkiem jest populacja 4750 przedsiębiorstw upadłych w latach 2007-2013, jest cennym dorobkiem autora. Potrafi ją dobrze zdefiniować i określić jej mankamenty. Jednak dobór przedsiębiorstw „zdrowych” do tych upadłych nie jest szczególnie uwidoczniiony, albowiem go w ogóle nie ma. Autor wykorzystał swój zbiór danych o przedsiębiorstwach upadłych w Polsce między innymi do pokazania, jak dane o tych przedsiębiorstwach pokazują symptomy upadłości na rok, dwa itd. przed samym terminem jej ogłoszenia. Sama monografia jest opracowaniem bardzo cennym, jako szczegółowa analiza danych o spółkach upadłych w Polsce w określonych czasowo i geograficznie ramach. Szkoda bardzo, że tak obfity zbiór danych o upadłościach nie zetknął się z możliwością wykorzystania w konfrontacji modelowania ze zbiorem danych o przedsiębiorstwach niezagrożonych finansowo.

Praca J. Kitowskiego [2015] to przykład swego rodzaju encyklopedii polskich badań nad upadłością, oparty na metodzie analizy dyskryminacyjnej. Autor jest niezawodny w wykrywaniu nieściśłości w polskich raportach o modelach krajowych i zagranicznych. Układ książki nie zachęca do analizy metodycznej, koncentruje się na szczegółowym raportowaniu kolejnych i kolejnych badań polskich przedsiębiorstw. Czytelnik nie poznaje opinii autora o sposobie dobierania próby oraz o innych tematach, na przykład normal-

ności rozkładu zmiennych objaśniających. Komplementuję jednak autora za ściganie niedorzeczności w krajowej literaturze na temat modeli bankructwa i upadłości.

Poza tymi i innymi monografiami, w polskiej literaturze na te tematy jest oczywiście wiele doniesień w postaci artykułów. Niektóre z nich stanowią swego rodzaju benchmarki dla badaczy z rachunkowości empirycznej, niektóre nie, np. bardzo dobry model P. Ciesielskiego [2005], opublikowany w „Rachunkowości”. Artykuły powstające obecnie, mam takie wrażenie, w dużej mierze powielają metodyczne pomysły sprzed lat, starając się je dostosować do coraz nowszych zbiorów danych. Tak czy inaczej, nie widać naszych autorów i naszych badań w światowym nurcie empirycznych finansów przedsiębiorstwa i rachunkowości stosowanej, o czym była mowa wyżej.

Jako czytelnikowi krajowych publikacji na temat zagrożenia finansowego i bankructwa, brak mi zawartych tam komentarzy o jakości modelowania i prognozowania w kontekście stosowanej próby firm upadłych/zagrożonych oraz firm niezagrożonych finansowo. Stąd pomysł, by przyrzeć się tym zagadnieniom na nowo, szczególnie że pojawiają się nowe prace kwestionujące dotychczasowe sposoby wnioskowania na podstawie szacowanych modeli.

### **Błąd próby dobieranej i błąd doboru próby**

Podstawowa „niekonsekwencja”, jaka do dziś wydaje się badaczom nierozwiązana, wynika z niskiej częstości zdarzenia/stanu „bankructwo” w całej populacji przedsiębiorstw. Na ogół dociera się do danych o pewnej określonej grupie bankrutów, a później czyni się rozmaite zabiegi dobierania „zdrowych” przedsiębiorstw do tej ustalonej już grupy. Pierwszy zwrócił na to uwagę M. Zmijewski [1984]. Jego pracę cytowano wielokrotnie w kolejnych latach po jej publikacji, jednak w

większości badań temat „problematicznej próby” był raczej pomijany. Ważniejsze były dobre prognozy/klasyfikacje, jakie z modeli otrzymywano.

Skoncentrujmy się na sytuacji zerojedynkowej (binarnej), to znaczy na modelowaniu zmiennej  $Y$  reprezentującej upadłość – na podstawie próby firm upadłych i nieupadłych [por. Gruszczyński, 2012]. Modelowanie polega na objaśnieniu zmiennej  $Y$  za pomocą modelu zawierającego dobrane zmienne objaśniające  $X$  (*covariates*). Chodzi o model dwumianowy zmiennej  $Y$ , w którym przyjmuje ona dwie wartości:  $Y=1$  (bankrut) lub  $Y=0$  (nie-bankrut). Firmy dobrane do próby nie stanowią na ogół próby losowej z całej populacji. To powoduje obciążenia. Dwa rodzaje obciążeń próby obecne w większości badań to:

- błąd próby dobieranej (*choice-based sample bias*), to znaczy takiej, do której wybiera się jednostki na podstawie wcześniejszej wiedzy o zmiennej zależnej (na przykład zbiera się najpierw dane o grupie bankrutów); prawdopodobieństwo dostania się do takiej próby zależy właśnie od cech zmiennej zależnej: próbę konstruuje się na przykład w ten sposób, że włącza się do niej wszystkie jednostki upadłe, a pozostałe dobiera za pomocą jakiegoś schematu dopasowania (*matching*);
- błąd doboru próby (*sample selection bias*) polega na tym, że z określonego powodu do próby trafiają tylko niektóre jednostki; w przypadku badań nad upadłością typowe jest dobieranie tylko tych jednostek, dla których mamy pełne dane.

Badania M. Zmijewskiego [1984] były prowadzone dla modelu probitowego i wskazały, że błąd próby dobieranej maleje gdy struktura próby (bankruci – nie-bankruci) zbliża się do struktury w populacji, a jest całkiem eliminowany gdy stosuje się odpowiednią metodę estymacji modelu.

Pokazał on także, że efektem obu rodzajów błędów jest asymptotyczne obciążenie ocen parametrów oraz ocen prawdopodobieństwa bankructwa. Jednocześnie, okazało się, że oba typy błędów nie miały znaczącego wpływu na dokładność klasyfikacji i predykcji na podstawie szacowanych modeli. Wskazał jednak, że te błędy mają wyraźny wpływ na oceny odnoszące się do pojedynczych przypadków, np. na oceny prawdopodobieństwa upadłości konkretnej firmy.

H.D. Platt i M.B. Platt [2002] potwierdzili na pojedynczym przykładzie istnienie obciążenia doboru próby w badaniach nad zagrożeniem finansowym. Ćwiczenie polegało na wielokrotnym wykonaniu estymacji modelu wczesnego ostrzegania dla prób losowanych z posiadanego przez autorów zbioru danych o bankrutach i o przedsiębiorstwach zdrowych.

Aby uniknąć obciążeń, jakie wynikają z prób dobieranych, niektórzy badacze korzystają z prób, w których proporcja bankrutów jest bliska tej w populacji, to znaczy bardzo niska. Na przykład J. Chen, et al. [2006] wykorzystali próbę z 89 bankrutami oraz 940 nie-bankrutami. Czasami próbuje się też dobierać próbę, ale nie w proporcji 50:50. Na przykład H.D. Platt, M.B. Platt, J.G. Pedersen [1994] wykorzystali 35 bankrutów, do których dobrali 89 nie-bankrutów. W nowszych badaniach wykorzystuje się liczniejsze próby, na przykład S. Jones i in. [2017] wzięli pod uwagę 990 bankrutów, a sama próba objęła 3960 obserwacji przestrzenno-czasowych (firmo-lat) dla bankrutów oraz 26 169 obserwacji dla nie-bankrutów (udział obserwacji dla bankrutów wyniósł ok. 13 proc.).

Świadomość błędów doboru próby jest w badaniach coraz powszechniejsza. W praktyce potwierdzają się ustalenia M. Zmijewskiego [1984] o mniejszym zakłóceniu tymi błędami przy modelach nakierowanych na jakość klasyfikacji. Błędy mają jednak znaczący wpływ na ocenę

prawdopodobieństwa bankructwa pojedynczego przedsiębiorstwa.

### Praktyka dla modelu logitowego: korekta Andersona-Maddali

Znane jest rozwiązanie przybliżające model upadłości oszacowany dla próby dobieranej do modelu dla populacji. Warunkiem jest stosowanie dwumianowego modelu logitowego oraz znajomość proporcji doboru przedsiębiorstw do próby. W tym przypadku stosuje się korektę Andersona-Maddali [Anderson 1973; Maddala 1983; Gruszczynski, 2012].

Jak to działa? Jak poprzednio,  $y=1$  oznacza firmę-bankruta, a  $y=0$  nie-bankruta. Przedmiotem modelowania jest prawdopodobieństwo  $P(y=1)$ . Jeśli znamy proporcje firm wybranych do próby z obu grup, to oznaczmy je  $p_1$  dla firm z  $y=1$  oraz  $p_2$  dla firm z  $y=0$ . Wówczas, zgodnie z korektą, w modelu logitowym po estymacji należy zmniejszyć wyraz wolny o wartość korekty

$$\delta = \ln p_1 - \ln p_2 \quad (1)$$

Korekta równa się zero, gdy te odsetki są identyczne, jak na przykład przy próbie losowej z obu grup łącznie. Przypomnijmy, że w modelu logitowym zmienną objaśnianą jest właśnie logit. Jest to logarytm szansy, to znaczy logarytm ilorazu prawdopodobieństw:  $P(y=1)$  przez  $1-P(y=1)$ . Zmiennymi objaśniającymi (zmiennie  $X$ ) są rozmaite charakterystyki przedsiębiorstw wybranych do próby. Zatem, po prawej stronie modelu jest liniowa funkcja zmiennych  $X$  oraz wyraz wolny. Ten właśnie element należy skorygować zgodnie ze wzorem (1).

Dla przykładu [Gruszczynski (red.), 2012] weźmy pod uwagę 1000 firm, z których 60 to bankruci, a 940 to nie-bankruci. Zwykle do próby wybiera się wszystkich bankrutów (to znaczy 60). Następnie spośród 940 nie-bankrutów losuje się 60 firm. Otrzymujemy w efekcie próbę, w której jest po 50 proc. firm z każdej grupy. Wtedy  $p_1=1$  (wzięliśmy

wszystkich bankrutów) oraz  $p_2=60/940$ . Po oszacowaniu modelu na podstawie próby można obliczyć oceny (wartości teoretyczne) prawdopodobieństwa upadłości dla firm należących do próby. To są oceny obciążone błędem doboru próby. Oceny nieobciążone otrzymujemy biorąc pod uwagę całą populację, z której wykonano losowanie. Model dla wszystkich 1000 firm uzyskuje się z modelu oszacowanego dla próby poprzez zmniejszenie wyrazu wolnego o wielkość (1) to znaczy o  $\ln(1)-\ln(60/940)$  czyli o 2,7515.

Jeśli nie znamy populacji i nie dokonujemy losowania, a mamy jedynie szacunki wielkości  $p_1$  i  $p_1$ , to korekta powinna być stosowana z odpowiednim komentarzem. W każdym razie, korektę stosujemy, gdy wnioskujemy w kontekście całej populacji, a nie samej próby.

Korekta Andersona-Maddali daje wynik tożsamy ze wzorem Skogsvików dla nieobciążonego prawdopodobieństwa upadłości, co pokazujemy niżej. Szersze konsekwencje praktyczne stosowania korekty można znaleźć w M. Gruszczyński [2012]. Warto podkreślić, że w tej ekspozycji nie chodzi o problem klasyfikacyjny, lecz o szacunek prawdopodobieństwa upadłości, co jest przedmiotem rosnącego popytu ze strony analityków rachunkowości i finansów przedsiębiorstwa.

### Wzór Skogsvików

Powtórzmy za ostatnim akapitem, że oprócz modelu „dobrze klasyfikującego” bankrutów i nie-bankrutów potrzebne są w praktyce bardziej szczegółowe informacje. Taką informacją jest nieobciążona ocena prawdopodobieństwa upadłości dla konkretnej firmy w próbie lub spoza próby. Dla modeli opartych na próbach dobieranych takie prawdopodobieństwo jest na ogół obciążone z uwagi na ich nielosowość.

K. Skogsvik i S. Skogsvik [2013] akcentują, że w prognozowaniu bankrutwa korzysta się zwykle z prób nielosowych i

w związku z tym otrzymywane z modeli prognostycznych prawdopodobieństwa (teoretyczne) upadłości zależą od udziału firm upadłych w próbie i dlatego są obciążone. Autorzy znaleźli arytmetyczny związek między obciążonym prawdopodobieństwem upadłości danej firmy (z modelu – *sample based*), a prawdopodobieństwem nieobciążonym, które wynika z udziału bankrutów w populacji. Ten ostatni udział został nazwany prawdopodobieństwem *a priori* bankrutwa (w danym roku).

Prawdopodobieństwo bankrutwa danej firmy wynikające z modelu (z próby) oznaczone przez autorów jako  $p_{fail}^{prop}$  jest zatem wielkością obciążoną.

Jest to funkcja:

- prawdopodobieństwa nieobciążonego  $p_{fail}^{\pi}$  (to nas właśnie interesuje),
- udziału (*prop*) bankrutów w próbie,
- udziału  $\pi$  bankrutów w populacji.

Wzór (wyprowadzony z twierdzenia Bayesa [por. Skogsvik, Skogsvik, 2013] jest następujący:

$$p_{fail}^{prop} = \left[ 1 + \left( \frac{\pi}{1-\pi} \right) \left( \frac{1-prop}{prop} \right) \left( \frac{1-p_{fail}^{\pi}}{p_{fail}^{\pi}} \right) \right]^{-1} \quad (2)$$

Na podstawie znalezionej zależności można przede wszystkim obliczyć  $p_{fail}^{\pi}$  dla każdej firmy w próbie. Wykazuje się też, że  $p_{fail}^{prop} > p_{fail}^{\pi}$  jeśli  $prop > \pi$  i vice versa. Ponadto, ranking firm ze względu na prawdopodobieństwo wynikające z modelu  $p_{fail}^{prop}$  jest tożsamy z rankingiem ze względu na prawdopodobieństwo  $p_{fail}^{\pi}$ .

Ważna uwaga: w dowodzie wzoru (2) autorzy zakładają, że firmy-bankruci są dobrane do próby w sposób losowy z populacji wszystkich bankrutów, to samo dotyczy populacji nie-bankrutów.

Jaki jest pożytek ze wzoru Skogsvików? Autorzy sami wskazują, że wprawdzie prawdopodobieństwo teoretyczne z modelu  $p_{fail}^{prop}$  na ogół wystarcza do klasyfi-

kacji firm (czyż tutaj się nie zajmujemy), to jednak konkretne prawdopodobieństwo nieobciążone  $p_{fail}^\pi$  jest potrzebne, np. w zarządzaniu ryzykiem finansowym, w wycenie akcji bądź obligacji firmy.

Warto zauważyć, że gdy  $prop = \pi$ , to wówczas  $p_{fail}^{prop} = p_{fail}^\pi$ .

To znaczy, że próba o proporcji bankrutów równej proporcji w populacji zapewnia nieobciążoność prawdopodobieństwa bankructwa danej firmy wynikającego z modelu. Wynikałby stąd wniosek, że gdy interesujemy się szacunkiem prawdopodobieństwa upadłości, to próba powinna mieć proporcje odpowiadające proporcjom dla populacji.

### Obciążona i nieobciążona ocena prawdopodobieństwa upadłości w różnych modelach

Wynik uzyskany przez K. Skogsvik i S. Skogsvik [2013] okazuje się być uogólnieniem wyniku Andersona-Maddali dla modelu logitowego, o czym była mowa wyżej. Oznaczmy przez  $N$  liczebność całej populacji, w której  $N_1$  oznacza liczebność bankrutów oraz  $N_2$  liczebność nie-bankrutów. Wówczas  $\pi = N_1/N_2$ . W próbie mamy  $n$  obserwacji, w tym  $n_1$  bankrutów oraz  $n_2$  nie-bankrutów. To znaczy  $prop = n_1/n$ . We wzorze (1) Andersona-Maddali mamy  $p_2 = n_2/N_2$  oraz  $p_1 = n_1/N_1$ . Korzystając z tych oznaczeń wzór (2) można zatem zapisać jako:

$$p_{fail}^{prop} = \left[ 1 + \frac{p_2}{p_1} \left( \frac{1 - p_{fail}^\pi}{p_{fail}^\pi} \right)^{-1} \right] \quad (3)$$

czyli

$$\frac{p_2}{p_1} \left( \frac{1 - p_{fail}^\pi}{p_{fail}^\pi} \right) = \left( \frac{1 - p_{fail}^{prop}}{p_{fail}^{prop}} \right) \quad (4)$$

Wzór (4), czyli inna postać wzoru (2) mówi, że iloczyn  $p_2/p_1$  oraz nieobciążonej szansy równa się szansie obciążonej. Dalej mamy (4):

$$\ln\left(\frac{p_2}{p_1}\right) - \ln\left(\frac{p_{fail}^\pi}{1 - p_{fail}^\pi}\right) = -\ln\left(\frac{p_{fail}^{prop}}{1 - p_{fail}^{prop}}\right)$$

albo

$$\ln p_1 - \ln p_2 + \text{logit}(p_{fail}^\pi) = \text{logit}(p_{fail}^{prop}) \quad (5)$$

Z tego wynika, że logit dla obciążonego prawdopodobieństwa bankructwa  $p_{fail}^{prop}$ , czyli to, co otrzymujemy z oszacowanego modelu logitowego należy zmniejszyć o wielkość

$$\delta = \ln p_1 - \ln p_2 \quad [\text{wzór (1)}]$$

aby otrzymać logit dla nieobciążonego prawdopodobieństwa bankructwa  $p_{fail}^\pi$ . Zatem – korekta Andersona-Maddali jest zgodna ze wzorem Skogsvików.

Zaletą wzoru (2) jest możliwość jego stosowania także do ocen prawdopodobieństwa bankructwa otrzymywanych z innych modeli dwumianowych, np. probitowego lub liniowego modelu prawdopodobieństwa (LMP). W modelu probitowym ocena (teoretyczna) prawdopodobieństwa upadłości nie da się przedstawić w postaci algebraicznie zamkniętej formuły. Oczywiście, można ją obliczyć i jako  $p_{fail}^{prop}$  podstawić do wzoru (4).

Jeśli chodzi o LMP, to prawdopodobieństwo (obciążone) odczytuje się bezpośrednio z modelu jako ocenę tzw. wartości teoretycznej zmiennej objaśnianej (zerojedynkowej) [por. Gruszczynski, 2012].

Tę wartość podstawiamy do równania (4) w miejsce  $p_{fail}^{prop}$  i dalej z tego równania obliczamy poszukiwane nieobciążone  $p_{fail}^\pi$ , pod warunkiem, że znane są proporcje  $p_1$  oraz  $p_2$ . Jeśli są one jednakowe, to znaczy, że ocena prawdopodobieństwa upadłości obliczona bezpośrednio z modelu jest nieobciążona.

Zależności (2) lub (4) pozwalają na obliczenie nieobciążonej oceny prawdopodobieństwa upadłości dla konkretnego przedsiębiorstwa na podstawie „standardowych” modeli: logitowego, probitowego oraz LMP.

### Efekty dla klasyfikacji

Jeśli szacunki prawdopodobieństw upadłości są jedynie podstawą dla klasyfikacji firm (na bankrutów i nie-bankrutów) to błędy doboru próby mogą nie

mieć istotnego znaczenia. To jest prawda, o ile punkt odcięcia (prawdopodobieństwo graniczne przy klasyfikacji) jest odpowiednio dobrany.

Zasada prognozy z punktem odcięcia 0,5 powinna być stosowana jedynie przy proporcji bankrutów i nie-bankrutów w próbie równej 1:1. Przy innych proporcjach można użyć na przykład zasady Cramera, która za punkt odcięcia przyjmuje odsetek jedynek w próbie [por. Gruszczyński, 2001]. Alternatywnie – obciążone prawdopodobieństwa upadłości powinny być przekształcone na nieobciążone przy pomocy zależności omawianych w poprzednich sekcjach [Skogsvik, Skogsvik, 2013].

Część modeli bankructwa nie może być bezpośrednio wykorzystywana do szacunku prawdopodobieństw upadłości, mają raczej cel klasyfikacyjny. Takim modelem jest model liniowej analizy dyskryminacyjnej (LAD). M. Zmijewski [1984] podaje, że nielosowy dobór próby może być skorygowany poprzez uwzględnienie struktury „bankruci – nie-bankruci” w populacji przy klasyfikacji na podstawie LAD (próba dobierana wpływa w tym przypadku jedynie na wyraz wolny [Morrison, 1990]). Problemem w stosowaniu LAD jest jednakże założenie o normalności rozkładu zmiennych  $X$ . Podkreślimy, że analiza dyskryminacyjna zakłada, iż zmienne objaśniające mają wielowymiarowy rozkład normalny w populacjach bankrutów i nie-bankrutów. Wskaźniki finansowe, które są proponowane jako zmienne  $X$ , nie spełniają na ogół tego założenia.

Warto w tym miejscu wskazać artykuł S. Jones, et al. [2017], w którym autorzy koncentrują się na nowych metodach klasyfikacji danych (klasyfikatorach) w odniesieniu do modelowania bankructwa. Te metody są oparte na koncepcjach wywodzących się z *data miningu*, takich jak sieci neuronowe, maszyny wektorów nośnych (*support vector machines*) czy techniki uczenia statystycznego nowej generacji,

takie jak *generalised boosting*, *AdaBoost* i drzewa losowe. Klasyczne klasyfikatory, takie jak logit czy LAD, okazują się całkiem dobre do prognozowania bankructwa, jednak autorzy proponują korzystanie z nowych technik. Sprawdzają się one lepiej dla danych przekrojowo-czasowych, a także – mimo formy „czarnej skrzynki” – mają całkiem akceptowalne interpretacje. Warto podkreślić, że kwestie błędów doboru próby nie są przedmiotem komentarzy w przytaczanym tu artykule.

Wracając do głównego wątku, jeśli przedmiotem modelowania jest otrzymanie nieobciążonego prawdopodobieństwa bankructwa (nadającego się do oceny ryzyka dla konkretnego przedsiębiorstwa), to powinno się korzystać z przekształceń proponowanych w przytoczonych wyżej wzorach. Podkreślimy raz jeszcze, że czynimy tak wówczas, gdy wnioskujemy w kontekście całej populacji, a nie samej próby.

## Podsumowanie

Dobór próby w modelowaniu bankructwa jest zazwyczaj słabszą stroną badań empirycznych z tego zakresu. Koncepcja tego artykułu polega na wskazaniu błędów doboru próby, ich efektów dla modeli prognostycznych i klasyfikacyjnych, a także sposobów radzenia sobie z nimi w praktyce. W szczególności, chodzi o szacunki prawdopodobieństwa bankructwa dla pojedynczego przedsiębiorstwa.

Otrzymywane z modeli oceny prawdopodobieństwa upadłości są na ogół obciążone, z powodu niezgodności proporcji bankrut – nie-bankrut w próbie i w populacji. Centralną część artykułu poświęcono analizie relacji pomiędzy prawdopodobieństwem obciążonym i nieobciążonym, która została podana przez Skogsvików [2013]. Ten wzór powinien być używany w praktyce, jeśli model wykorzystuje się do szacunku prawdopodobieństwa bankructwa konkretnej firmy. Okazuje się, że dla modelu logitowego ten wzór jest zgod-

ny z korektą Andersona-Maddali [Anderson, 1973; Maddala, 1983], co pokazano w artykule.

Dodajmy, że rzadko podnosi się temat jakości „klasycznych” polskich modeli upadłości, które do dzisiaj ochoczo są przykładane do aktualnych danych, a które – poza swoją wiekowością – były prze-

cież szacowane z wieloma wadami metodycznymi. Nie ma powodu, aby właśnie teraz do tego wracać, ale trudno oprzeć się wrażeniu, że spora część polskich autorów od finansów i rachunkowości nie sięga po znane, lepsze podejścia metodyczne, zwłaszcza że bardzo poprawiła się jakość danych.

### Bibliografia:

1. Adamska A., Mączyńska E. (red.) [2013], *Upadłości, bankructwa i naprawa przedsiębiorstw. Wybrane zagadnienia*, Warszawa, Oficyna Wydawnicza SGH.
2. Anderson J.A. [1972], *Separate sample logistic discrimination*, „Biometrika”, Vol. 59, p. 19-35.
3. Antonowicz P. [2013], *Założenia analizy dyskryminacyjnej w konstruowaniu nowych wielowymiarowych modeli klasyfikacyjnych w celu prognozowania upadłości przedsiębiorstw*, w: *Upadłości, bankructwa i naprawa przedsiębiorstw. Wybrane zagadnienia*, A. Adamska, E. Mączyńska (red.), Warszawa, Oficyna Wydawnicza SGH.
4. Chen J., Marshall B.R., Zhang J., Ganesh S. [2006], *Financial Distress Prediction in China*, “Review of Pacific Basin Financial Markets and Policies”, Vol. 09, No. 02, p. 317-336.
5. Ciesielski P. [2005], *Prognozowanie upadłości podmiotów gospodarczych w Polsce*, „Rachunkowość”, nr 8.
6. Ciszewska M. [2013], *Prognozowanie upadłości przedsiębiorstw z wykorzystaniem analizy dyskryminacyjnej przy spełnieniu założenia dotyczącego rozkładu badanych zmiennych*, w: *Upadłości, bankructwa i naprawa przedsiębiorstw. Wybrane zagadnienia*, A. Adamska, E. Mączyńska (red.), Warszawa, Oficyna Wydawnicza SGH.
7. Gruszczyński M. [2001], *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Warszawa, Oficyna Wydawnicza SGH.
8. Gruszczyński M. (red.) [2012], *Mikroekonometria*, wyd. 2, Warszawa, Wolters Kluwer.
9. Gruszczyński M. [2012], *Empiryczne finanse przedsiębiorstw. Mikroekonometria finansowa*, Warszawa, Difin.
10. Jones S., Johnstone D., Wilson R. [2017], *Predicting Corporate Bankruptcy: An Evaluation of Alternative Statistical Frameworks*, “Journal of Business Finance and Accounting”, Vol. 44(1) & (2), p. 3-34.
11. Kitowski J. [2015], *Metody dyskryminacyjne jako instrument oceny zagrożenia upadłością przedsiębiorstwa*, Rzeszów, Wydawnictwo Uniwersytetu Rzeszowskiego.
12. Maddala G.S. [1983], *Limited dependent and qualitative variables in econometrics*, Cambridge, Cambridge University Press.
13. Morrison D.F. [1990], *Wielowymiarowa analiza statystyczna*, Warszawa, PWN.
14. Platt H.D., Platt M.B., Pedersen J.G. [1994], *Bankruptcy discrimination with real variables*, “Journal of Business Finance and Accounting”, Vol. 1(4), p. 491-510.
15. Platt H.D., Platt M.B. [2002], *Predicting corporate financial distress. Reflections on choice-based sample bias*, “Journal of Economics and Finance”, Vol. 26/2, p. 184-199.
16. Prusak B. [2005], *Nowoczesne metody prognozowania zagrożenia finansowego przedsiębiorstw*, Difin, Warszawa.
17. Rogowski W., Duleba K. [2013], *Możliwość wykorzystania modeli oceny zagrożenia bankructwem w praktyce audytorów*, w: *Upadłości, bankructwa i naprawa przedsiębiorstw. Wybrane zagadnienia*, A. Adamska, E. Mączyńska (red.), Warszawa, Oficyna Wydawnicza SGH.
18. Skogsvik K., Skogsvik S. [2013], *On the choice based sample bias in probabilistic bankruptcy prediction*, “Investment Management and Financial Innovations”, Vol. 10, Issue 1, p. 29-37.
19. Tymoszek M. [2013], *Skuteczność modeli prognozowania upadłości przedsiębiorstw a upływ czasu – porównanie popularnych polskich modeli wielowymiarowej analizy dyskryminacyjnej z modelem zbudowanym przez autorkę*, w: *Upadłości, bankructwa i naprawa przedsiębiorstw. Wybrane zagadnienia*, A. Adamska, E. Mączyńska (red.), Warszawa, Oficyna Wydawnicza SGH.
20. Zmijewski M. [1984], *Methodological issues related to the estimation of financial distress prediction models*, “Journal of Accounting Research”, Vol. 20, p. 59-82.