

Jadwiga Kostrzewska

Barbara Pawełek

Artur Lipieta

Katedra Statystyki

Uniwersytet Ekonomiczny w Krakowie

Problem wartości odstających w badaniu kondycji finansowej przedsiębiorstw budowlanych w Polsce*

Streszczenie

Wyniki analizy kondycji finansowej przedsiębiorstw są wykorzystywane m.in. w badaniach dotyczących zagrożenia upadłością. Do oceny kondycji finansowej przedsiębiorstw wykorzystuje się wskaźniki finansowe, podstawą badań są zatem dane pochodzące ze sprawozdań finansowych. Ocena jakości tych danych obejmuje m.in. wykrywanie wartości odstających. Celem artykułu jest przedstawienie wyników badań empirycznych nad wpływem wyboru metody wykrywania i eliminacji wartości odstających na skuteczność klasyfikacyjną modelu logitowego, budowanego na podstawie zbiorów uwzględniających lub pomijających wykryte wartości odstające. W badaniach empirycznych wykorzystano jedno- i wielowymiarowe metody wykrywania wartości odstających. Metody te dodatkowo połączono z analizą mocy dyskryminacyjnej wskaźników finansowych. Ocenę skuteczności modelu logitowego oparto na miernikach wiarygodności i specyficzności. Badaniem objęto przedsiębiorstwa budowlane w Polsce w latach 2005, 2007 i 2009.

* Artykuł powstał w ramach realizacji projektu badawczego sfinansowanego ze środków przyznanych Wydziałowi Zarządzania Uniwersytetu Ekonomicznego w Krakowie w ramach dotacji na utrzymanie potencjału badawczego.

Słowa kluczowe: wartości odstające, kondycja finansowa, wskaźnik finansowy, model logitowy, klasyfikacja.

1. Wprowadzenie

Analiza kondycji finansowej jest ważnym elementem zarządzania przedsiębiorstwem. Pogarszająca się sytuacja finansowa może prowadzić do bankructwa. Do oceny kondycji finansowej przedsiębiorstw wykorzystuje się wskaźniki pochodzące ze sprawozdań finansowych. Pierwszym etapem takiej analizy powinna być ocena jakości posiadanych danych finansowych.

Ocena jakości danych obejmuje m.in. wykrywanie wartości odstających. Obserwacje odstające są na ogół spowodowane błędami w danych. Powstają często na skutek błędnego pomiaru lub pomyłek przy wprowadzaniu informacji do baz danych, oszustw i nietypowych zachowań. Mogą też wynikać z naturalnych odchyłek jednostki od przyjętego wzorca wyznaczonego przez większość obserwacji [Hodge i Austin 2004]. Chociaż istnienie wartości odstających często jest traktowane jako błąd (zakłócenie, szum), one same mogą być nośnikiem ważnych informacji. W niniejszym opracowaniu przyjęto, że odstające wartości wskaźników finansowych dla pewnych przedsiębiorstw mogą świadczyć o ich złej kondycji finansowej, mogącej prowadzić do upadłości.

Obserwacja odstająca to taka, która wydaje się znacznie różnić od innych elementów zbiorowości, w której występuje [Grubbs 1969, Barnett i Lewis 1994, Hodge i Austin 2004, John 1995]. Odstającą nazywana jest więc często obserwacja, której odległość od innych wartości w rozważanym zbiorze jest nieporównywalnie większa od odległości pomiędzy pozostałymi elementami tej zbiorowości [Johnson i Wichern 1992]. Z punktu widzenia metod analizy skupień za odstającą może być uznana grupa obserwacji małej miary (tzn. o niewielkiej liczebności) [Kaufman i Rousseeuw 1990, Ramaswamy, Rastogi i Shim 2000].

Metody wykrywania wartości odstających można podzielić na: parametryczne (statystyczne), oparte na znajomości parametrów rozkładów empirycznych lub ich oszacowań (np. [Hawkins 1980]), i nieparametryczne, oparte na miarach odległości, np. mierze Mahalanobisa (np. [Williams i in. 2002]). Inny podział wyróżnia metody jednowymiarowe i wielowymiarowe. W pierwszych pracach analizujących wartości odstające stosowano metody jednowymiarowe [Barnett i Lewis 1994]. F.R. Hampel [1971 i 1974] zaproponował wykorzystanie mediany i mediany odchylenia absolutnego od mediany jako odpornych estymatorów położenia i rozproszenia. Jako miarę odporności estymatora na wartości odstające wprowadził on punkt załamania (najmniejszy odsetek odstających obserwacji, które powodują, że estymator zacznie dawać dowolnie duże wyniki). Powszechnie jako estymator rozproszenia stosuje się także odchylenie standar-

dowe obliczone na podstawie obserwacji pomijających wartości skrajne – dolne lub górne. Zazwyczaj pomija się 5% obserwacji [Ben-Gal 2005]. J.W. Tukey [1977] wprowadził wykres pudełkowy jako graficzne narzędzie pomocne do identyfikacji obserwacji odstających.

Aktualnie w pracach dotyczących wartości odstających częściej stosowane są metody wielowymiarowe, uwzględniające możliwość wystąpienia obserwacji odstających w ujęciu globalnym (wielowymiarowym) przy możliwym jednoczesnym braku odstawiania obserwacji po poszczególnych współrzędnych. W ramach wielowymiarowych, nieparametrycznych metod do wykrywania wartości odstających można stosować różne funkcje głębi (np. [Zuo i Serfling 2000, Zuo 2003]).

W pracach z zakresu badania kondycji finansowej przedsiębiorstw, w szczególności prognozowania bankructwa, można znaleźć rozważania dotyczące występowania w danych wartości odstających. Propozycje rozwiązania tego problemu wahają się od ignorowania [Spicka 2013], przez zamianę lub usunięcie wartości odstających [De Andrés i in. 2011, Pociecha i in. 2014, Shumway 2001, Wu, Gaunt i Gray 2010] do stosowania metod odpornych. W badaniach empirycznych pojawiają się zatem wątpliwości dotyczące wyboru poprawnego podejścia do wartości odstających: wykrywać je, czy nie wykrywać? jeśli wykrywać, to w jaki sposób i co zrobić z wiedzą wynikającą z ich występowania w zbiorze danych?

Niniejsza praca wpisuje się w nurt badań nad istotą i znaczeniem identyfikacji wartości odstających w badaniu kondycji finansowej przedsiębiorstw (zob. np. [Tsai i Cheng 2012]). Celem artykułu jest prezentacja wyników badań nad wpływem wyboru metody wykrywania i eliminacji wartości odstających na skuteczność modelu logitowego, budowanego na podstawie zbiorów uwzględniających lub pomijających wykryte wartości odstające, w zakresie klasyfikacji przedsiębiorstw na „zdrowe” i zagrożone upadłością. Wyniki badania pilotażowego dla 2009 r. zaprezentowano w pracy [Pawełek, Kostrzewska i Lipieta 2015].

2. Dane

Dane finansowe przedsiębiorstw budowlanych w Polsce za lata 2005–2009 pochodzą z serwisu EMIS Intelligence – Polska. Zbudowana baza zawiera informacje dotyczące 371 przedsiębiorstw, w tym siedmiu w stanie upadłości. Przedsiębiorstwa „zdrowe” są nazywane w pracy także „niebankrutami” (NB), natomiast będące w stanie upadłości – „bankrutami” (B). Słabą stroną analizowanej bazy jest mała liczba przedsiębiorstw w stanie upadłości (przedstawiających sprawozdania finansowe w całym analizowanym okresie), co uniemożliwiło stworzenie zbioru testowego dla modelu logitowego. Badania empiryczne objęły lata 2005, 2007 i 2009.

W badaniach wykorzystano 14 wskaźników finansowych (tabela 1), które podzielno na cztery grupy ważnych charakterystyk kondycji finansowej przedsiębiorstw, a mianowicie: płynności (R_{01} – R_{03}), poziomu zadłużenia (R_{04} – R_{06}), rentowności (R_{07} – R_{10}) i sprawności działania (R_{11} – R_{14}).

Tabela 1. Wskaźniki finansowe

Symbol	Opis	Symbol	Opis
R_{01}	Wskaźnik płynności bieżącej	R_{08}	Wskaźnik rentowności netto
R_{02}	Wskaźnik płynności szybkiej	R_{09}	Wskaźnik rentowności kapitałów własnych
R_{03}	Wskaźnik natychmiastowej wymagalności	R_{10}	Wskaźnik rentowności aktywów
R_{04}	Wskaźnik zadłużenia aktywów	R_{11}	Wskaźnik rotacji należności
R_{05}	Wskaźnik zadłużenia kapitału własnego	R_{12}	Wskaźnik rotacji aktywów stałych
R_{06}	Wskaźnik zadłużenia długoterminowego	R_{13}	Wskaźnik rotacji aktywów
R_{07}	Wskaźnik rentowności brutto	R_{14}	Wskaźnik poziomu kosztów operacyjnych

Źródło: opracowanie własne.

Statystyczna analiza zmian rozkładów wartości poszczególnych wskaźników finansowych w kolejno następujących po sobie latach okresu 2005–2009 została przeprowadzona w grupie wszystkich przedsiębiorstw oraz oddzielnie w grupach niebankrutów i bankrutów.

Na podstawie wykresów pudełkowych (względem mediany oraz względem średniej) oraz wybranych statystyk opisowych porównano rozkłady wartości wskaźników wszystkich badanych przedsiębiorstw. Dla poszczególnych wskaźników finansowych w kolejnych latach utrzymuje się na ogół podobny, lecz niekoniecznie taki sam, charakter rozkładów. Największe różnice występują w wartościach odstających, koncentracji wartości wokół średniej lub mediany i stopniu zróżnicowania. W całym badanym okresie silnie prawostronnie asymetryczne rozkłady odnotowano dla wartości wskaźników: płynności bieżącej (R_{01}), płynności szybkiej (R_{02}), natychmiastowej wymagalności (R_{03}), zadłużenia długoterminowego (R_{06}), rotacji należności (R_{11}) i rotacji aktywów stałych (R_{12}). Asymetria rozkładów wartości wymienionych wskaźników finansowych utrzymuje się we wszystkich latach badanego okresu, choć z różnym natężeniem. Dla wskaźników zadłużenia aktywów (R_{04}) oraz poziomu kosztów operacyjnych (R_{14}) rozkłady wartości są bliskie symetrycznym w całym badanym okresie. Ocena siły i kierunku asymetrii rozkładu wartości wskaźników finansowych

przydatna jest przy stosowaniu analizy kwantylowej do wykrywania obiektów o wartościach odstających.

W analizowanej bazie danych obejmującej lata 2005–2009 znalazło się siedem przedsiębiorstw budowlanych, które ogłosiły bankructwo do 2005 r. (włącznie). Mały udział bankrutów wśród wszystkich badanych przedsiębiorstw (ok. 1,89%) sprawia, że ich wpływ na rozkład wartości poszczególnych wskaźników finansowych nie jest duży.

Można jednak zauważyć, że wartości niektórych wskaźników odnotowane w grupie bankrutów są inne niż w grupie „zdrowych” przedsiębiorstw budowlanych, np. zdecydowanie niskie, jak w wypadku wszystkich trzech wskaźników płynności (płynności bieżącej – R_{01} , płynności szybkiej – R_{02} , i natychmiastowej wymagalności – R_{03}) oraz dwóch wskaźników sprawności działania (rotacji należności – R_{11} , oraz rotacji aktywów stałych – R_{12}). Ponadto w grupie bankrutów zdarzają się nietypowe (odstające na tle wartości dla „zdrowych” przedsiębiorstw) wartości wskaźników finansowych, np. w wypadku wszystkich trzech wskaźników zadłużenia (aktywów – R_{04} , kapitału własnego – R_{05} , i długoterminowego – R_{06}) oraz trzech wskaźników rentowności (brutto – R_{07} , netto – R_{08} , i kapitałów własnych – R_{09}). Wartości takie odnotowano tylko dla niektórych bankrutów i tylko w niektórych latach. Trzeba jednak podkreślić, że każde z przedsiębiorstw z grupy bankrutów charakteryzuje się odmiennymi wartościami rozważanych wskaźników finansowych w analizowanym okresie. Przy próbie określenia ogólnych tendencji zbyt silny może być wpływ indywidualnych różnic między przedsiębiorstwami.

3. Metodyka badań

Do wykrywania wartości odstających zastosowano po jednej metodzie: jednowymiarowej (opartej na kwantylach) i wielowymiarowej (opartej na funkcji głębi projekcyjnej). Metody te zastosowano oddzielnie lub z uwzględnieniem wyników analizy mocy dyskryminacyjnej wskaźników finansowych.

Na podstawie analizy rozkładów wskaźników finansowych (por. pkt 2) zauważono, że w niektórych przypadkach wartości odnotowane dla bankrutów kształtują się odmiennie na tle wartości dla „zdrowych” przedsiębiorstw. W szczególności dla bankrutów pojawiają się wartości wskaźników zdecydowanie powyżej lub zdecydowanie poniżej typowego zakresu wartości „zdrowych” przedsiębiorstw, tzn. znajdujące się w ogonach rozkładów. Z tego względu w jednowymiarowej analizie badano obszary wyznaczone przez kwantyl $q_{0,1}$ lub – oddzielnie – kwantyl $q_{0,9}$. Dla danego wskaźnika finansowego kwantyle odpowiedniego rzędu wyznaczono w grupie „zdrowych” przedsiębiorstw. Obiekty –

przedsiębiorstwa o wartościach silnie niższych od $q_{0,1}$, silnie wyższych od $q_{0,9}$ lub spoza obustronnego zakresu wyznaczonego przez kwantyle $q_{0,05}$ i $q_{0,95}$ uznano za odstające. Postępowano przy tym na dwa sposoby.

W pierwszym podejściu dla każdego z 14 analizowanych wskaźników finansowych wyznaczono wartości odstające dla „zdrowych” przedsiębiorstw za pomocą kwantyla odpowiedniego rzędu w zależności od kierunku asymetrii rozkładu danego wskaźnika (silna asymetria lewostronna – lewy ogon rozkładu wyznaczony przez kwantyl $q_{0,1}$; prawostronna – prawy ogon wyznaczony przez $q_{0,9}$; brak silnej asymetrii – obustronny obszar wyznaczony przez $q_{0,05}$ i $q_{0,95}$).

W drugim podejściu przy wyborze prawego, lewego lub obustronnego ogona rozkładu kierowano się analizą mocy dyskryminacyjnej wskaźników finansowych. Podstawą wykrywania wartości odstających w tym podejściu są tylko te wskaźniki, które mają większą moc dyskryminacyjną od pozostałych (por. tabela 2).

Tabela 2. Wskaźniki finansowe o większej mocy dyskryminacyjnej niż pozostałe wskaźniki w latach 2005, 2007 i 2009 (m – liczba wskaźników)

Rok	Wskaźniki finansowe	m
2005	$R_{01} R_{02} R_{03} R_{04}^a R_{05} R_{07} R_{08} R_{09}^b R_{10} R_{11} R_{12} R_{13} R_{14}^a$	13
2007	$R_{04}^a R_{05} R_{06}^b R_{08} R_{09}^a R_{10} R_{14}^a$	7
2009	$R_{03} R_{06}^a R_{07} R_{09}^b R_{10} R_{11} R_{14}^a$	7

^a Moc dyskryminacyjna w prawym ogonie. ^b Moc dyskryminacyjna w obu ogonach.

Uwaga: pozostałe – moc dyskryminacyjna w lewym ogonie.

Źródło: opracowanie własne.

W celu ustalenia, które ze wskaźników mogłyby wystarczająco wyraźnie sygnalizować problemy z kondycją finansową, wyznaczono ich moc dyskryminacyjną (zob. też [Yu i in. 2014]). Jako kryterium przyjęto liczbę bankrutów o wartościach należących do wyznaczonych 10-procentowych zakresów wartości skrajnych dla „zdrowych” przedsiębiorstw (w prawym lub lewym ogonie rozkładu). Gdy wartości przyjmowane przez bankrutów wystąpiły w obu ogonach rozkładu dla „zdrowych” przedsiębiorstw, sprawdzono odsetek bankrutów w obustronnym 10-procentowym obszarze wyznaczonym przez kwantyle $q_{0,05}$ oraz $q_{0,95}$. Im wyższa jest wartość tego kryterium, tym większa jest moc dyskryminacyjna danego wskaźnika. Za wskaźniki o mocy dyskryminacyjnej uznano te, dla których w jednym z obszarów wyznaczonych przez kwantyle dla „zdrowych” przedsiębiorstw odnotowano wartości wskaźnika dla co najmniej 2 z 7 przedsiębiorstw bankrutów, co było bardzo łagodnym kryterium. W poszczególnych latach były to różne zestawy (por. tabela 2). W szczególności na początku badanego okresu, tj.

w 2005 r., aż 13 wskaźników finansowych miało moc dyskryminacyjną (zgodnie z przyjętym kryterium), zaś w następnych dwóch badanych latach – tylko 7. Oznacza to, że wartości wskaźników finansowych przedsiębiorstw budowlanych w niedługim czasie po ogłoszeniu upadłości znajdowały się w ogonach rozkładu wartości wskaźników dla „zdrowych” przedsiębiorstw. Im więcej czasu upłynęło od ogłoszenia upadłości, tym mniej wskaźników charakteryzowało się mocą dyskryminacyjną, a zatem tym mniej przedsiębiorstw bankrutów odnotowało skrajne wartości wskaźników finansowych.

Do wykrywania wartości odstających w przestrzeni wielowymiarowej wykorzystano głębię projekcyjną [Zuo 2003]. Koncepcja głębi danych to zagadnienie nieparametrycznej, odpornej wielowymiarowej analizy statystycznej rozwijane w ramach eksploracyjnej analizy danych. Umożliwia ona określenie liniowego porządku wielowymiarowych obserwacji z wykorzystaniem wielowymiarowej mediany, definiowanej jako wielowymiarowe centrum zbioru obserwacji. Istnieje wiele propozycji funkcji, zwanych funkcjami głębi, przyporządkowujących każdej obserwacji pochodzącej z pewnego rozkładu dodatnią liczbę będącą miarą jej odstawania od centrum ze względu na ten rozkład [Kosiorowski 2012].

W wielowymiarowej metodzie również zastosowano dwa podejścia: bez uwzględnienia lub z uwzględnieniem wiedzy o mocy dyskryminacyjnej wskaźników finansowych. W pierwszym podejściu podstawą budowy wielowymiarowego centrum zbioru obserwacji jest 14 wskaźników finansowych, w drugim – m wskaźników finansowych o mocy dyskryminacyjnej. „Zdrowe” przedsiębiorstwa o wartościach odstających to 10% obiektów leżących najdalej od 14- lub m -wymiarowego centrum.

Korzystając z funkcji głębi projekcyjnej w wykrywaniu wartości odstających, należy pamiętać, że metoda ta wskazuje obiekty oddalone od centrum zbioru danych bez uwzględniania kierunku odstawania (tzn. wśród wymienionych przedsiębiorstw odstających mogą być zarówno przedsiębiorstwa charakteryzujące się bardzo dobrą kondycją finansową, jak i przedsiębiorstwa borykające się z poważnymi problemami finansowymi).

Dla poszczególnych lat zastosowano zatem metody analizy jednowymiarowej opartej na kwantylach oraz wielowymiarowej opartej na funkcji głębi. Metody oznaczono w taki sposób, by pierwsza litera informowała o rodzaju analizy (K oparta na kwantylach, G oparta na funkcji głębi), następnie podano dwucyfrowy symbol roku (05, 07 lub 09) oraz liczbę wskaźników finansowych będących podstawą wykrywania obiektów o wartościach odstających ($m = 14, 13$ lub 7). Przy tym m różne od 14 wskazuje na uwzględnienie w metodzie mocy dyskryminacyjnej wskaźników finansowych.

Dla każdego roku osobno wspomnianymi metodami utworzono zbiory, które są podstawą dalszej analizy. Przyjęto następujące warianty:

- wariant I – zbiory obejmujące wszystkie przedsiębiorstwa w danym roku (pełna baza);
 - wariant II – zbiory obejmujące wszystkich bankrutów i „zdrowe” przedsiębiorstwa niebędące obiektami odstającymi przy zastosowaniu danej metody;
 - wariant III – zbiory składające się z bankrutów i „zdrowych” przedsiębiorstw będących obiektami odstającymi przy zastosowaniu danej metody.
- Liczebności zbiorów w poszczególnych wariantach zestawiono w tabeli 3.

Tabela 3. Liczebności zbiorów w wariantach II oraz III w zależności od zastosowanej metody i liczby uwzględnionych wskaźników finansowych (m) w latach 2005, 2007 i 2009

Metoda	Wariant II			Wariant III		
	2005	2007	2009	2005	2007	2009
<i>K.rr.14</i>	176	169	181	202	209	197
<i>K.rr.m</i>	211 ($m = 13$)	246 ($m = 7$)	243 ($m = 7$)	167 ($m = 13$)	132 ($m = 7$)	135 ($m = 7$)
<i>G.rr.14</i> <i>G.rr.m</i>	335 ($m = 14$ lub 13)	335 ($m = 14$ lub 7)	335 ($m = 14$ lub 7)	43 ($m = 14$ lub 13)	43 ($m = 14$ lub 7)	43 ($m = 14$ lub 7)

Źródło: opracowanie własne.

W wypadku zastosowania analizy jednowymiarowej opartej na kwantylach z uwzględnieniem lub bez uwzględnienia informacji o mocy dyskryminacyjnej wskaźników finansowych dla różnych lat otrzymano różne liczebności zbiorów wariantów II i III (por. tabela 3). Jest to spowodowane tym, że obiekty o wartościach odstających ze względu na wartości różnych wskaźników nie muszą być takie same. W wypadku zastosowania analizy wielowymiarowej opartej na funkcji głębi w obu wersjach dla różnych lat otrzymano równoliczne zbiory wariantu II (335 przedsiębiorstw) lub wariantu III (43 przedsiębiorstwa). Przy czym zarówno w jedno-, jak i wielowymiarowej metodzie, w zależności od uwzględnienia lub nieuwzględnienia wiedzy o mocy dyskryminacyjnej wskaźników, niekoniecznie te same obiekty zostały wskazane jako odstające. Można zatem spodziewać się różnic w modelach logitowych szacowanych z wykorzystaniem wiedzy o wartościach odstających, uzyskanej za pomocą różnych metod.

Do oceny kondycji finansowej przedsiębiorstw budowlanych w Polsce zastosowano model logitowy o postaci:

$$P(y_i = \text{bankrut} | \mathbf{x}_i) = \frac{\exp(\mathbf{x}_i \boldsymbol{\beta})}{1 + \exp(\mathbf{x}_i \boldsymbol{\beta})}, \quad (1)$$

gdzie:

- \mathbf{x}_i – wektor wartości zmiennych niezależnych dla obiektu i -tego,
- $\boldsymbol{\beta}$ – wektor parametrów.

Parametry modelu szacowano metodą krokową wsteczną. Za pomocą oszacowanego modelu dokonano klasyfikacji przedsiębiorstw do grup przedsiębiorstw w dobrej (niebankrutów) lub w złej kondycji finansowej (bankrutów). Do oceny skuteczności klasyfikacyjnej modelu logitowego wykorzystano miarę wrażliwości, obliczaną jako odsetek przedsiębiorstw w stanie upadłości, które zostały poprawnie zaklasyfikowane przez model do zbioru przedsiębiorstw zagrożonych upadłością, oraz miarę specyficzności modelu, wyznaczaną jako odsetek przedsiębiorstw „zdrowych” poprawnie zaklasyfikowanych przez model do grupy przedsiębiorstw niezagrażonych upadłością [Pociecha i in. 2014].

Wiedza o obiektach odstających wśród „zdrowych” przedsiębiorstw wykrytych różnymi metodami została wykorzystana do konstrukcji zbiorów (por. tabela 3) będących podstawą budowy modeli logitowych. Celem analizy zaprezentowanej w następnych punktach jest sprawdzenie, czy wiedza o wartościach odstających może być przydatna do poprawienia skuteczności klasyfikacyjnej oszacowanych modeli.

4. Estymacja i ocena skuteczności klasyfikacyjnej modelu logitowego dla 2005 roku

W pierwszym etapie analizy dla 2005 r. na podstawie całej bazy (wariant I, tj. 371 przedsiębiorstw) oszacowano dwa modele logitowe. Wejściowy zbiór zmiennych objaśniających stanowiło 14 wskaźników finansowych (model $M_{O.05.14}$) albo 13 wytypowanych wskaźników (model $M_{O.05.13}$) o mocy dyskryminacyjnej zgodnie z kryterium podanym w pkt 3 (por. tabela 2). Przy konstrukcji tych modeli ignorowano problem występowania obiektów – przedsiębiorstw o odstających wartościach wskaźników finansowych ($M_{O.05.14}$) lub uwzględniono jedynie moc dyskryminacyjną wskaźników ($M_{O.05.13}$) przez zamieszczenie ich we wstępnym zbiorze zmiennych objaśniających.

W drugim etapie analizy zbudowano modele logitowe na podstawie zbiorów uzyskanych z wejściowej bazy przez usunięcie z niej przedsiębiorstw „zdrowych”, uznanych za obiekty odstające (wariant II – por. tabela 3). Analizę przeprowadzono osobno dla wyników uzyskanych po zastosowaniu jedno- (modele $M_{K.05}$) lub wielowymiarowej (modele $M_{G.05}$) metody wykrywania wartości odstających. Wejściowy zbiór zmiennych objaśniających zawierał 14 wskaźników – modele $M_{K.05.14}$ i $M_{G.05.14}$, lub 13 wskaźników o mocy dyskryminacyjnej (por. tabela 2) – modele $M_{K.05.13}$ i $M_{G.05.13}$. Zmienne niezależne o statystycznie istotnym wpływie

na zmienną zależną w oszacowanych modelach logitowych dla danych z 2005 r. zestawiono w tabeli 4.

Tabela 4. Zmienne objaśniające w oszacowanych modelach logitowych dla 2005 r.

Model	Zmienne objaśniające	Zbiór uczący
$M_{O.05.14}$	R_{04}	wariant I
$M_{O.05.13}$	R_{04}	wariant I
$M_{K.05.14}$	$R_{04} R_{13}$	wariant II
$M_{K.05.13}$	R_{04}	wariant II
$M_{G.05.14}$	$R_{04} R_{12}$	wariant II
$M_{G.05.13}$	$R_{04} R_{12} R_{14}$	wariant II

Źródło: opracowanie własne.

W każdym z sześciu oszacowanych modeli wśród zmiennych objaśniających znajduje się wskaźnik zadłużenia aktywów (R_{04}). W niektórych modelach ponadto znajdują się niektóre wskaźniki z grupy wskaźników sprawności działania (R_{12} , R_{13} , R_{14} – por. tabela 1). Wskaźniki z pozostałych grup (płynności i rentowności) nie mają statystycznie istotnego wpływu na klasyfikację za pomocą modelu logitowego do grupy przedsiębiorstw w złej lub dobrej kondycji finansowej. Warto przy tym nadmienić, że badani bankruci ogłosili upadłość do 2005 r. (włącznie), a w modelach na ich złą kondycję finansową wskazuje tylko kilka wskaźników, m.in. zadłużenia aktywów.

Skuteczność klasyfikacyjna modeli zbudowanych na podstawie pełnego zbioru ($M_{O.05.14}$, $M_{O.05.13}$) została oceniona z wykorzystaniem mierników wrażliwości i specyficzności na zbiorze uczącym wariantu I. Ocenę skuteczności klasyfikacyjnej modeli zbudowanych na zbiorach wariantu II, również z wykorzystaniem mierników wrażliwości i specyficzności, przeprowadzono na: całej bazie (wariant I), zbiorze uczącym (wariant II) i zbiorze kontrolnym (wariant III) (por. tabela 3). Wyniki zamieszczono w tabeli 5.

Najwyższą wartość miary wrażliwości (0,714, co oznacza 5 poprawnie zaklasyfikowanych przedsiębiorstw w złej kondycji finansowej – por. tabela 5) oraz wysokie wartości miary specyficzności odnotowano w przypadku modelu logitowego $M_{G.05.13}$. Model ten oszacowano, łącząc wielowymiarową metodę wykrywania wartości odstających (funkcja głębi) z wiedzą o mocy dyskryminacyjnej wskaźników finansowych. Dość wysokie wartości miary wrażliwości (powyżej 0,5) odnotowano w drugim modelu, $M_{G.05.14}$, oszacowanym z wykorzystaniem wiedzy uzyskanej za pomocą wielowymiarowej metody wykrywania wartości odstających, oraz modelach $M_{K.05.14}$ i $M_{K.05.13}$, oszacowanych z wykorzystaniem wiedzy uzyskanej za pomocą jednowymiarowej metody wykrywania wartości

odstających (analiza kwantylowa). We wspomnianych czterech modelach zbiorami uczącymi są odpowiednie zbiory wariantu II, zaś zbiorami kontrolnymi – wariantu I lub III.

Tabela 5. Ocena skuteczności klasyfikacyjnej oszacowanych modeli logitowych za pomocą miary wrażliwości oraz miary specyficzności dla 2005 r.

Model	N	Wariant I		Wariant II		Wariant III	
		wrażliwość	specyficzność	wrażliwość	specyficzność	wrażliwość	specyficzność
$M_{O.05.14}$	371	0,143 (1)	0,997	×	×	×	×
$M_{O.05.13}$	371	0,143 (1)	0,997	×	×	×	×
$M_{K.05.14}$	176	0,571 (4)	0,951	0,571 (4)	1,000	0,571 (4)	0,908
$M_{K.05.13}$	211	0,571 (4)	0,970	0,571 (4)	1,000	0,571 (4)	0,931
$M_{G.05.14}$	335	0,571 (4)	0,975	0,571 (4)	0,997	0,571 (4)	0,778
$M_{G.05.13}$	335	0,714 (5)	0,981	0,714 (5)	0,997	0,714 (5)	0,833

Uwagi: N – liczebność zbioru uczącego. W nawiasie podano liczbę przedsiębiorstw w złej kondycji finansowej poprawnie zaklasyfikowanych do grupy bankrutów. Pogrubioną czcionką wyróżniono wartości miary wrażliwości wyższe niż 0,5.

Źródło: opracowanie własne.

W każdym z modeli najwyższa skuteczność klasyfikacji mierzona miarą specyficzności występowała na zbiorze uczącym, zaś najniższa – na zbiorach wariantu III składających się z trudno rozróżnialnych dla modelu obiektów: „zdrowych” przedsiębiorstw o odstających wartościach wskaźników finansowych oraz przedsiębiorstw bankrutów.

Wyniki wskazują, że uwzględnienie informacji o mocy dyskryminacyjnej wskaźników finansowych w połączeniu z wielowymiarową metodą wykrywania wartości odstających dało najwyższą skuteczność klasyfikacyjną przedsiębiorstw budowlanych w Polsce w 2005 r.

5. Estymacja i ocena skuteczności klasyfikacyjnej modelu logitowego dla 2007 roku

Podobnie jak poprzednio w pierwszym etapie analizy dla 2007 r. na podstawie całej bazy (tj. 371 przedsiębiorstw – wariant I) oszacowano dwa modele logitowe. Wejściowy zbiór zmiennych objaśniających stanowiło 14 wskaźników finansowych (model $M_{O.07.14}$) albo 7 wskaźników (model $M_{O.07.7}$), mających moc dyskryminacyjną zgodnie z kryterium podanym w pkt 3 (por. tabela 2). Modele te oszacowano, ignorując problem wartości odstających ($M_{O.07.14}$) lub biorąc pod

uwagę tylko moc dyskryminacyjną wskaźników finansowych ($M_{O.07.7}$) przez uwzględnienie ich we wstępnym zbiorze zmiennych objaśniających.

W drugim etapie analizy zbudowano modele logitowe na podstawie zbiorów uzyskanych z wejściowej bazy przez usunięcie z niej przedsiębiorstw „zdrowych”, uznanych za obiekty odstające (wariant II). Analizę przeprowadzono osobno dla wyników uzyskanych po zastosowaniu jedno- (modele $M_{K.07}$) lub wielowymiarowej (modele $M_{G.07}$) metody wykrywania wartości odstających. Wejściowy zbiór zmiennych objaśniających zawierał 14 wskaźników (modele $M_{K.07.14}$ i $M_{G.07.14}$) lub 7 wskaźników o mocy dyskryminacyjnej (zob. tabela 2) – modele $M_{K.07.7}$ i $M_{G.07.7}$.

Tabela 6. Zmienne objaśniające w oszacowanych modelach logitowych dla 2007 r.

Model	Zmienne objaśniające	Zbiór uczący
$M_{O.07.14}$	$R_{05} R_{06}$	wariant I
$M_{O.07.7}$	$R_{05} R_{06}$	wariant I
$M_{K.07.14}$	$R_{05} R_{06} R_{11} R_{14}$	wariant II
$M_{K.07.7}$	$R_{05} R_{06} R_{10} R_{14}$	wariant II
$M_{G.07.14}$	$R_{05} R_{06} R_{11}$	wariant II
$M_{G.07.7}$	$R_{05} R_{06} R_{08} R_{14}$	wariant II

Źródło: opracowanie własne.

Zmienne niezależne o statystycznie istotnym wpływie na zmienną zależną w oszacowanych modelach logitowych dla danych z 2007 r. zestawiono w tabeli 6. We wszystkich sześciu otrzymanych modelach wśród zmiennych objaśniających pozostają wskaźniki zadłużenia kapitału własnego (R_{05}) oraz zadłużenia długoterminowego (R_{06}). W modelach oszacowanych na zbiorach składających się z typowych (według danej metody) „zdrowych” przedsiębiorstw oraz przedsiębiorstw bankrutów, znajdują się także niektóre wskaźniki z grupy rentowności (R_{08} – wskaźnik rentowności netto, lub R_{10} – rentowności aktywów) oraz z grupy sprawności działania (R_{11} – wskaźnik rotacji należności, i/lub R_{14} – poziomu kosztów operacyjnych). Wskaźniki płynności nie mają statystycznie istotnego wpływu na klasyfikację przedsiębiorstw ze względu na kondycję finansową w 2007 r.

Podobnie jak wcześniej skuteczność klasyfikacyjna modeli $M_{O.07.14}$ oraz $M_{O.07.7}$ została oceniona z wykorzystaniem mierników wrażliwości i specyficzności na zbiorze uczącym wariantu I, natomiast pozostałych modeli logitowych – na całym zbiorze (wariant I), zbiorze uczącym (wariant II) i zbiorze kontrolnym (wariant III). Wyniki zamieszczono w tabeli 7.

Najwyższą wartość miernika wrażliwości (0,714, tj. 5 poprawnie zaklasyfikowanych przedsiębiorstw w złej kondycji finansowej) odnotowano w przypadku modelu logitowego $M_{K.07.7}$ (por. tabela 7). Model ten oszacowano na podstawie zbioru wariantu II skonstruowanego za pomocą jednowymiarowej metody wykrywania wartości odstających (analiza kwantylowa) z wykorzystaniem wiedzy o mocy dyskryminacyjnej wskaźników finansowych. Pominięcie tej dodatkowej wiedzy doprowadziło do powstania modelu $M_{K.07.14}$, dla którego wartości miary wrażliwości są wyższe niż 0,5, co jest dobrym wynikiem na tle pozostałych modeli.

Tabela 7. Ocena skuteczności klasyfikacyjnej oszacowanych modeli logitowych za pomocą miary wrażliwości oraz miary specyficzności dla 2007 r.

Model	N	Wariant I		Wariant II		Wariant III	
		wrażliwość	specyficzność	wrażliwość	specyficzność	wrażliwość	specyficzność
$M_{O.07.14}$	371	0,143 (1)	0,997	×	×	×	×
$M_{O.07.7}$	371	0,143 (1)	0,997	×	×	×	×
$M_{K.07.14}$	169	0,571 (4)	0,923	0,571 (4)	1,000	0,571 (4)	0,861
$M_{K.07.7}$	246	0,714 (5)	0,913	0,714 (5)	0,996	0,714 (5)	0,760
$M_{G.07.14}$	335	0,286 (2)	0,995	0,286 (2)	1,000	0,286 (2)	0,944
$M_{G.07.7}$	335	0,286 (2)	0,962	0,286 (2)	0,997	0,286 (2)	0,639

Uwaga: jak w tabeli 5.

Źródło: opracowanie własne.

Tak jak poprzednio wartości miary specyficzności na zbiorach kontrolnych nie są wyższe niż na zbiorze uczącym. W wypadku zbiorów wariantu III miara specyficzności osiąga najniższe wartości dla danego modelu.

Wyniki otrzymane na podstawie badania przeprowadzonego dla danych z 2007 r. wskazują, że uwzględnienie informacji o mocy dyskryminacyjnej wskaźników finansowych – tym razem w połączeniu z jednowymiarową analizą kwantylową – znacznie poprawiło skuteczność klasyfikacyjną modeli logitowych.

6. Estymacja i ocena skuteczności klasyfikacyjnej modelu logitowego dla 2009 roku

W pierwszym etapie analizy dla danych z 2009 r. na podstawie bazy wariantu I oszacowano dwa modele logitowe, ignorując problem wartości odstających

(14 wskaźników finansowych w wejściowym zbiorze zmiennych objaśniających – model $M_{O.09.7}$) lub uwzględniając tylko moc dyskryminacyjną wskaźników finansowych (7 wskaźników o mocy dyskryminacyjnej w wejściowym zbiorze zmiennych objaśniających – model $M_{O.09.7}$).

W drugim etapie analizy zbudowano modele logitowe na podstawie zbiorów uzyskanych z wejściowej bazy przez usunięcie z niej „zdrowych” przedsiębiorstw uznanych za obiekty odstające – wariant II. Analizę przeprowadzono osobno dla wyników uzyskanych po zastosowaniu jedno- (modele $M_{K.09}$) lub wielowymiarowej (modele $M_{G.09}$) metody wykrywania wartości odstających. Wejściowy zbiór zmiennych objaśniających zawierał 14 (modele $M_{K.09.14}$ i $M_{G.09.14}$) lub 7 wskaźników o mocy dyskryminacyjnej (zob. tabela 2) – modele $M_{K.09.7}$ i $M_{G.09.7}$.

Tabela 8. Zmienne objaśniające w oszacowanych modelach logitowych dla 2009 r.

Model	Zmienne objaśniające	Zbiór uczący
$M_{O.09.14}$	$R_{01} R_{04} R_{06} R_{09} R_{11} R_{13} R_{14}$	wariant I
$M_{O.09.7}$	$R_{07} R_{09} R_{11}$	wariant I
$M_{K.09.14}$	$R_{06} R_{07}$	wariant II
$M_{K.09.7}$	$R_{06} R_{14}$	wariant II
$M_{G.09.14}$	$R_{03} R_{09} R_{14}$	wariant II
$M_{G.09.7}$	$R_{06} R_{09} R_{11} R_{14}$	wariant II

Źródło: opracowanie własne.

Zestaw zmiennych niezależnych o statystycznie istotnym wpływie na zmienną zależną w otrzymanych modelach logitowych oszacowanych dla danych z 2009 r. zamieszczono w tabeli 8. Wraz z upływem czasu od momentu ogłoszenia upadłości przez badane przedsiębiorstwa – bankrutów, więcej wskaźników ma statystycznie istotny wpływ na klasyfikację przedsiębiorstw ze względu na ich kondycję finansową. Niemal w każdym modelu statystycznie istotny wpływ mają wskaźniki z grup: sprawności działania, zyskowności, a także zadłużenia. Wskaźniki płynności (R_{01} – płynności bieżącej, lub R_{03} – natychmiastowej wymagalności) występują tylko w dwóch modelach.

Tak jak poprzednio skuteczność klasyfikacyjną oszacowanych modeli oceniono z wykorzystaniem mierników wrażliwości i specyficzności (por. tabela 9). W modelach $M_{O.09.14}$ oraz $M_{O.09.7}$ mierniki te obliczono tylko na zbiorze uczącym wariantu I, natomiast w pozostałych modelach oszacowanych na podstawie zbiorów wariantu II – na całej bazie (wariant I), zbiorze uczącym (wariant II) i zbiorze kontrolnym (wariant III) (por. także tabela 3).

Na podstawie wyników zamieszczonych w tabeli 9 można stwierdzić, że najwyższą wartość miary wrażliwości (0,429, tj. 3 poprawnie zaklasyfikowane

przedsiębiorstwa w złej kondycji finansowej) odnotowano dla modeli logitowych oszacowanych na zbiorach wariantu II skonstruowanych z wykorzystaniem wiedzy o wartościach odstających uzyskanej za pomocą jednowymiarowej metody wykrywania wartości odstających ($M_{K.09.14}$ i $M_{K.09.7}$). Spośród wspomnianych dwóch modeli wyższą wartością miary specyficzności charakteryzuje się model, przy którego budowie wykorzystano wiedzę o mocy dyskryminacyjnej wskaźników finansowych. Modele $M_{G.09.14}$ i $M_{G.09.7}$ skonstruowane z wykorzystaniem wielowymiarowej metody wykrywania wartości odstających (funkcja głębi) z uwzględnieniem lub bez uwzględnienia wiedzy o mocy dyskryminacyjnej wskaźników finansowych, mają jeszcze niższą ocenę skuteczności klasyfikacyjnej (miara wrażliwości na poziomie 0,286) mimo dość wysokich na tle pozostałych modeli wartości miary specyficzności.

Tabela 9. Ocena skuteczności klasyfikacyjnej oszacowanych modeli logitowych za pomocą miary wrażliwości oraz miary specyficzności dla 2009 r.

Model	N	Wariant I		Wariant II		Wariant III	
		wrażliwość	specyficzność	wrażliwość	specyficzność	wrażliwość	specyficzność
$M_{O.09.14}$	371	0,143 (1)	1,000	×	×	×	×
$M_{O.09.7}$	371	0,000 (0)	1,000	×	×	×	×
$M_{K.09.14}$	181	0,429 (3)	0,896	0,429 (3)	0,948	0,429 (3)	0,847
$M_{K.09.7}$	243	0,429 (3)	0,901	0,429 (3)	1,000	0,429 (3)	0,719
$M_{G.09.14}$	335	0,286 (2)	0,984	0,286 (2)	1,000	0,286 (2)	0,833
$M_{G.09.7}$	335	0,286 (2)	0,975	0,286 (2)	1,000	0,286 (2)	0,750

Uwagi tak jak do tabeli 5.

Źródło: opracowanie własne.

Trzeba podkreślić, że wszystkie wartości miary wrażliwości w oszacowanych modelach dla danych z 2009 r. nie przekraczają progu 0,5. Można odnieść wrażenie, że tym razem zbudowane modele nie są dobrym narzędziem klasyfikacyjnym dla rozważanej grupy przedsiębiorstw budowlanych w Polsce. Należy jednak zwrócić uwagę, że 2009 r. to najdalszy punkt czasu od momentu ogłoszenia upadłości przez badane przedsiębiorstwa. Otrzymane wyniki mogą zatem sugerować, że przedsiębiorstwa, które były w złej kondycji finansowej na początku badanego okresu (do 2005 r.), po kilku latach (w 2009 r.) kondycję tę poprawiły. Z tego względu za pomocą zbudowanych modeli nie da się ich na ogół odróżnić od przedsiębiorstw w dobrej kondycji finansowej.

7. Podsumowanie

Podjęte badania są próbą odpowiedzi na pytania, w jaki sposób uwzględniać wartości odstające w ocenie kondycji finansowej przedsiębiorstw budowlanych w Polsce, oraz czy wybór metody wykrywania obiektów odstających ma wpływ na skuteczność klasyfikacyjną modelu logitowego w przypadku analizowania kondycji finansowej przedsiębiorstw budowlanych w Polsce.

W pracy przedstawiono wyniki badania przeprowadzonego na podstawie danych finansowych przedsiębiorstw budowlanych w Polsce w latach 2005, 2007 i 2009. Uzyskane wyniki sugerują przydatność uwzględniania wiedzy o wartościach odstających w badaniu kondycji finansowej przedsiębiorstw, co więcej – wskazują, że ignorowanie problemu wartości odstających prowadzi do niezadowolających wyników (tj. niskiej skuteczności klasyfikacyjnej modeli).

Wyniki uzyskane dla 2005 r. wskazują na najwyższą skuteczność klasyfikacyjną w wypadku modelu logitowego oszacowanego po zastosowaniu wielowymiarowej metody wykrywania przedsiębiorstw o wartościach odstających (funkcja głębi) w połączeniu z informacją o mocy dyskryminacyjnej wskaźników finansowych.

Wyniki uzyskane dla 2007 r. są nieco inne. Najwyższą skuteczność klasyfikacyjną odnotowano w wypadku modelu logitowego oszacowanego po zastosowaniu jednowymiarowej metody wykrywania wartości odstających (analiza kwantylowa) w połączeniu z informacją o mocy dyskryminacyjnej wskaźników.

Obliczenia wykonane dla 2009 r. wskazują na większą przydatność jednowymiarowej metody wykrywania wartości odstających (analiza kwantylowa) niż zastosowanej metody wielowymiarowej (funkcja głębi projekcyjnej). Wskazują również na poprawę skuteczności klasyfikacyjnej modeli, gdy wykorzystana zostaje wiedza o mocy dyskryminacyjnej wskaźników finansowych. Niestety dla danych z 2009 r. żaden z oszacowanych modeli nie osiągnął skuteczności klasyfikacyjnej bankrutów wyższej niż 50% (miara wrażliwości). Należy jednak zauważyć, że jest to najdalszy moment analizy od czasu ogłoszenia upadłości przez badane przedsiębiorstwa. Otrzymane wyniki sugerują, że działania podjęte przez większość analizowanych przedsiębiorstw są skuteczne. W rezultacie przedsiębiorstw tych na ogół nie da się odróżnić od przedsiębiorstw w dobrej kondycji finansowej. Szersza analiza tego zagadnienia jest przedmiotem dalszych badań autorów.

Drugi etap analiz zaprezentowanych w pkt 4–6 miał na celu sprawdzenie wpływu sposobu wykorzystania wiedzy o obiektach odstających na skuteczność klasyfikacyjną modeli logitowych. W szczególności rozważenie wejściowego zbioru zmiennych objaśniających, opartego na 13 lub 7 wskaźnikach finansowych o mocy dyskryminacyjnej, jest próbą wykorzystania wiedzy o obiektach

odstających w ujęciu jednowymiarowym bez tradycyjnej eliminacji obiektów z bazy. Badania przeprowadzone dla lat 2005 i 2007 wskazały, że podejście to (uwzględnienie informacji o obiektach odstających przez analizę mocy dyskryminacyjnej wskaźników finansowych) przyczyniło się do wyższej skuteczności klasyfikacyjnej modeli logitowych oszacowanych na zbiorach wariantu II utworzonych z wykorzystaniem jedno- (analiza kwantylowa) lub wielowymiarowej analizy (funkcja głębi).

Autorzy planują przeprowadzić podobne badania na bazie danych, w której kosztem zakresu czasowego zwiększony zostanie zbiór przedsiębiorstw w stanie upadłości. Rozważane jest także: rozszerzenie badania o wariant uwzględniający wszystkie lata łącznie, zwiększenie zbioru metod wykrywania wartości odstających oraz wzięcie pod uwagę innych metod szacowania parametrów w modelu logitowym (w tym metod odpornych na wartości odstające).

Literatura

- Barnett V., Lewis T. [1994], *Outliers in Statistical Data*, John Wiley, New York.
- Ben-Gal I. [2005], *Outlier Detection [w:] Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, red. O. Maimon i L. Rokach, Kluwer Academic Publishers, Dordrecht.
- De Andrés J., Sánchez-Lasheras F., Lorca P., De Cos Juez F.J. [2011], *A Hybrid Device of Self Organizing Maps (SOM) and Multivariate Adaptive Regression Splines (MARS) for the Forecasting of Firms' Bankruptcy*, „Accounting and Management Information Systems”, vol. 10, nr 3.
- Grubbs F.E. [1969], *Procedures for Detecting Outlying Observations in Samples*, „Technometrics”, vol. 11, nr 1, <http://dx.doi.org/10.1080/00401706.1969.10490657>.
- Hampel F.R. [1971], *A General Qualitative Definition of Robustness*, „The Annals of Mathematical Statistics”, vol. 42, nr 6, <http://dx.doi.org/10.1214/aoms/1177693054>.
- Hampel F.R. [1974], *The Influence Curve and Its Role in Robust Estimation*, „Journal of the American Statistical Association”, vol. 69, nr 346, <http://dx.doi.org/10.1080/01621459.1974.10482962>.
- Hawkins D. [1980], *Identification of Outliers*, Chapman and Hall, London–New York.
- Hodge V.J., Austin J. [2004], *A Survey of Outlier Detection Methodologies*, „Artificial Intelligence Review”, vol. 22, nr 2, <http://dx.doi.org/10.1023/b:aire.0000045502.10941.a9>.
- John G.H. [1995], *Robust Decision Trees: Removing Outliers from Databases [w:] Proceedings of the First International Conference on Knowledge Discovery and Data Mining*, red. U.M. Fayyad i R. Uthurusamy, AAAI Press, Menlo Park, CA.
- Johnson R., Wichern D.W. [1992], *Applied Multivariate Statistical Analysis*, Prentice Hall, Upper Saddle River, NJ.
- Kaufman L., Rousseeuw P.J. [1990], *Finding Groups in Data: An Introduction to Cluster Analysis*, Wiley, New York.

- Kosiorowski D. [2012], *Statystyczne funkcje głębi w odpornej analizie ekonomicznej*, Zeszyty Naukowe Uniwersytetu Ekonomicznego w Krakowie, Seria specjalna: Monografie, nr 208, Kraków.
- Pawelek B., Kostrzewska J., Lipieta A. [2015], *The Problem of Outliers in the Research on the Financial Standing of Construction Enterprises in Poland* [w:] *Proceedings of the 9th Professor Aleksander Zeliaś International Conference on Modelling and Forecasting of Socio-economic Phenomena*, red. M. Papież i S. Śmiech, Foundation of the Cracow University of Economics, Cracow.
- Pociecha J., Pawelek B., Baryła M., Augustyn S. [2014], *Statystyczne metody prognozowania bankructwa w zmieniającej się koniunkturze gospodarczej*, Fundacja Uniwersytetu Ekonomicznego w Krakowie, Kraków.
- Ramaswamy S., Rastogi R., Shim K. [2000], *Efficient Algorithms for Mining Outliers from Large Data Sets* [w:] *Proceedings of the 2000 ACM SIGMOD International Conference on Management of Data*, May 16–18, 2000, Dallas, Texas, red. W. Chen, J. Naughton i P.A. Bernstein, Association for Computing Machinery, New York.
- Shumway T. [2001], *Forecasting Bankruptcy More Accurately: A Simple Hazard Model*, „The Journal of Business”, vol. 74(1).
- Spicka J. [2013], *The Financial Condition of the Construction Companies before Bankruptcy*, „European Journal of Business and Management”, vol. 5, nr 23.
- Tsai Ch.F., Cheng K.Ch. [2012], *Simple Instance Selection for Bankruptcy Prediction*, „Knowledge-Based Systems”, vol. 27, <http://dx.doi.org/10.1016/j.knosys.2011.09.017>.
- Tukey J.W. [1977], *Exploratory Data Analysis*, Addison-Wesley, Reading, PA.
- Williams G.J., Baxter R.A., He H.X., Hawkins S., Gu L. [2002], *A Comparative Study of RNN for Outlier Detection in Data Mining*, IEEE International Conference on Data-mining (ICDM'02), Maebashi City, Japan, CSIRO Technical Report CMIS-02/102.
- Wu Y., Gaunt C., Gray S. [2010], *A Comparison of Alternative Bankruptcy Prediction Models*, „Journal of Contemporary Accounting and Economics”, vol. 6, nr 1, <http://dx.doi.org/10.1016/j.jcae.2010.04.002>.
- Yu Q., Miche Y., Séverin E., Lendasse A. [2014], *Bankruptcy Prediction Using Extreme Learning Machine and Financial Expertise*, „Neurocomputing”, vol. 128, <http://dx.doi.org/10.1016/j.neucom.2013.01.063>.
- Zuo Y. [2003], *Projection-Based Depth Functions and Associated Medians*, „The Annals of Statistics”, vol. 31, nr 5, <http://dx.doi.org/10.1214/aos/1065705115>.
- Zuo Y., Serfling R. [2000], *General Notions of Statistical Depth Functions*, „The Annals of Statistics”, vol. 28, nr 2, <http://dx.doi.org/10.1214/aos/1016218226>.

The Problem of Outliers in Research on the Financial Standing of Construction Enterprises in Poland

(Abstract)

The results of an analysis of financial standing can be used to study the threat of going bankrupt. Financial indicators are used to evaluate enterprises' financial standing. Thus, the data from financial statements is the basis for the examination of the financial position. The evaluation of data quality includes the identification of outliers, among other factors. This article presents the results of an empirical study done on how the method of detecting and eliminating outliers chosen influences the effectiveness of a logit model

constructed on the basis of samples that either included the outliers or left them out. The research for the paper employed one- and multi-dimensional methods of detecting outliers and their combinations with an analysis of the discriminatory power of the financial indicators. Classification effectiveness of the logit model was assessed by sensitivity and specificity measures. The research covered the years 2005, 2007 and 2009.

Keywords: outliers, financial standing, financial indicator, logit model, classification.