



dr hab. Michał Kałdoński, prof. nadzw. UEP
Katedra Finansów Przedsiębiorstw
Wydział Zarządzania
Uniwersytet Ekonomiczny w Poznaniu

Poznań, dnia 18 lutego 2019 r.

Recenzja rozprawy doktorskiej pana Arkana Thomasa Kłozy'ego pod tytułem:
Rachunkowość kreatywna: badanie empiryczne z wykorzystaniem wskaźnika C-score (Creative Accounting: An Empirical Investigation Based on the C-score Model) przygotowanej pod kierunkiem prof. dr hab. Dariusza Zarzeckiego oraz promotora pomocniczego dra Pawła Barana

1. Aktualność i znaczenie tematyki rozprawy

Termin „rachunkowości kreatywnej” (*creative accounting*) jest z pewnością jednym z częściej używanych w piśmiennictwie naukowym i prasie w kontekście wielkich skandali finansowych, do których doszło głównie w USA, ale także w niektórych krajach europejskich, na przełomie XX i XXI wieku. Nie znaczy to jednak, że problem został zauważony dopiero wtedy. Można w tym miejscu przywołać choćby pracę Jones [1991], która wpisuje się w koncepcję rachunkowości kreatywnej, a dotyczy kształtowania zysków w spółkach starających się o działania ochronne państwa przed importerami. Ingerencja zarządzających w procesy raportowania wyników ma różne uwarunkowania, cele i może być realizowana za pomocą różnych instrumentów. Może też być różnie oceniana. Ronen i Yaari [2008] klasyfikują przyjmowane w literaturze definicje zgodnie z kryterium jakości sprawozdań finansowych w 3 grupy: 1) definicje „białe” odwołujące się do korzystnych konsekwencji tych procesów; 2) definicje „szare”, podkreślające zarówno korzyści, jak i negatywne konsekwencje podejmowanych działań; 3) definicje „czarne” zwracające uwagę głównie na negatywne konsekwencje w postaci ograniczonej transparentności sprawozdań finansowych i prywatnych korzyści osiąganych przez zarządzających. Wielowątkowość podejmowanego przez pana Arkana Thomasa Kłozy'ego w rozprawie zagadnienia jest już wystarczającą podstawą, aby z uznaniem patrzeć na przedstawioną do oceny dysertację.

Skala zainteresowania środowiska akademickiego problematyką rachunkowości kreatywnej wydaje się ogromna. Nie ma dziś bowiem czasopisma z zakresu finansów i rachunkowości, w którym w tytułach opublikowanych prac (lub przynajmniej w spisie słów kluczowych) nie pojawiałyby się takie sformułowania, jak „kształtowanie zysków” (*earnings management*), „błędne raportowanie finansowe” (*financial misreporting*), „nadużycia księgowość” (*accounting irregularities*), „manipulacje księgowość” (*accounting manipulations*) czy sama „rachunkowość kreatywna”. Warto także podkreślić, że problematyka rozprawy nie jest obca polskiej literaturze przedmiotu. Prace badawcze z tego obszaru prowadzone są w większości ośrodków akademickich w Polsce, a ich rezultaty ogłaszane są zarówno w postaci artykułów, jak również monografii.

Coraz większe zainteresowanie wspomnianym zagadnieniem wykazuje także praktyka gospodarcza. Z najnowszych badań kwestionariuszowych przeprowadzonych w USA wśród 400 dyrektorów finansowych spółek publicznych i prywatnych wynika, że 20% spośród nich aktywnie uczestniczy w procesie kształtowania wyników finansowych, zaś skala niewłaściwie zaraportowanych zysków może wynosić nawet 10% [Dichev i in., 2016]. Fakt aktywnych działań w obszarze rachunkowości kreatywnej w Polsce potwierdzają rezultaty postępowań kontrolnych wszczynanych przez Komisję Nadzoru



Finansowego wobec spółek publicznych. Rośnie także świadomość skali i konsekwencji takich praktyk wśród inwestorów.

Zatem wybór tematu rozprawy przygotowanej przez pana Arkana Thomasa Kłozy'ego uważam za trafny, interesujący i cenny poznawczo, świadczy on także o dobrym rozeznaniu Doktoranta w aktualnych nurtach badawczych z zakresu finansów i rachunkowości.

2. Ocena celu, hipotez, struktury rozprawy oraz przeglądu literatury

Recenzowana rozprawa doktorska koncentruje się na opracowaniu ilościowego narzędzia (modelu) wykrywania praktyk kreatywnej rachunkowości, który w zamierzeniu Doktoranta mógłby być stosowany w odniesieniu do podmiotów funkcjonujących w różnym otoczeniu gospodarczym i którego wykorzystanie mogłoby ograniczyć skalę zjawiska wprowadzania w błąd użytkowników informacji finansowych. Osiągnięcie tak określonego celu głównego rozprawy ma umożliwić realizację trzech celów pomocniczych. W mojej opinii Doktorantowi udało się zrealizować założony cel pracy.

W rozprawie postawiona została także hipoteza główna oraz trzy hipotezy pomocnicze. Hipoteza główna w zasadzie pokrywa się z celem głównym pracy, z kolei hipotezy pomocnicze – w mojej ocenie – w dużej mierze mają charakter założeń badawczych. Skłania mnie to do poddania pod dyskusję następującej kwestii: czy w pracach o charakterze normatywno-projektowym, a do takich z pewnością można zaliczyć recenzowaną rozprawę, konieczne jest stawianie hipotez? Osobiście jestem zdania, że w przypadku recenzowanej rozprawy ograniczenie się do celu (-ów) pracy nie stanowiłoby uszczerbku dla logiki wywodu czy szerzej – dla poprawności procesu badawczego. Mam jednocześnie świadomość, że badacze wywodzący się z różnych szkół mogą mieć odmienny pogląd w tym zakresie.

Przedłożona do oceny praca liczy łącznie 411 stron, z tego właściwy tekst zawiera się na ok. 300 stronach. Tytuł rozprawy dość dobrze oddaje istotę rozważań w niej zawartych. Praca podzielona jest na 5 rozdziałów, zawiera wstęp, podsumowanie, załączniki, bibliografię, spisy tabel i schematów oraz streszczenie w języku polskim. Struktura pracy jest klarowna i podporządkowana realizacji celu pracy.

Rozdział pierwszy poświęcono obszernemu wprowadzeniu do własnego badania empirycznego. Rozdziały drugi i trzeci mają charakter metodyczno-empiryczny. Z kolei rozdział czwarty i piąty to obszerny raport z własnego badania. Rozdział pierwszy (ok. 50 stron) wprowadza w tematykę koncepcji, uwarunkowań, motywów i instrumentów rachunkowości kreatywnej. Jest to dobre kompendium istniejącego stanu wiedzy w tym obszarze. Rozdziały drugi (ok. 50 stron) i trzeci (ok. 60 stron) to próba połączenia warstwy metodycznej z warstwą empiryczną. Zawartość rozdziału drugiego to przede wszystkim opis doboru próby, procedur audytowych i ich rezultatów umożliwiających identyfikację podmiotów oraz obszarów, w których stosowane są praktyki kreatywnej rachunkowości. Podmioty te określane są przez Doktoranta najczęściej mianem firm manipulatorów (*manipulating companies*) lub firm praktykujących rachunkowość kreatywną (*firms practicing creative accounting*). Z kolei rozdział trzeci zawiera kompleksowy przegląd literatury poświęconej narzędziom wykrywania praktyk kreatywnej rachunkowości ze szczególnym naciskiem na tzw. „modele accruals”¹ oraz model Beneisha dedykowany wykrywaniu podmiotów zagrożonych dokonywaniem manipulacji księgowych. Istotną część rozdziału trzeciego stanowi opis rezultatów własnych badań nad skutecznością wspomnianych wcześniej modeli. Zwięźlenie pracy stanowią rozdziały czwarty i piąty (łącznie ok. 100

¹ Trudno jest znaleźć w polskiej literaturze odpowiednik pojęcia accruals w takim znaczeniu, w jakim jest ono stosowane w badaniach. Dla celów niniejszej recenzji, za Piosik [2013], termin accruals jest używany zamiennie z pojęciem „zmian niepieniężnych aktywów obrotowych netto”.



stron), w których Doktorant opisuje przebieg oraz rezultaty procesu budowania funkcji dyskryminacyjnych oraz w sposób krytyczny dokonuje oceny przydatności opracowanych modeli.

Praca oparta jest na dość dobrym przeglądzie literatury. Można odnieść wrażenie, że Doktorant chciał przeczytać „prawie wszystko, co należało przeczytać”. Szczególnie wartościowe są fragmenty pracy poświęcone narzędziom detekcji manipulacji księgowych. Analiza wyników dokonanego przeglądu publikacji daje podstawy, by twierdzić, że Autor dobrze orientuje się w dokonaniach światowej społeczności zajmującej się tematyką rachunkowości kreatywnej. Chcąc zachować obiektywizm, jednocześnie muszę stwierdzić, że Doktorant miał pewien problem z selekcją źródeł. Spis bibliografii ma 26 stron i obejmuje prawie 400 pozycji literaturowych. To oczywiście świadczy o dobrym czytaniu Doktoranta, ale nadal łatwo znaleźć teksty pominięte w tym spisie. Na przykład nie ma tu tekstów traktujących o wskaźniku F-score [Dechow i in. 2011], stanowiącym bezpośrednią konkurencję dla takich modeli, jak model Beneisha. Nie znalazłem także odniesienia do prac łączących poziom ochrony praw inwestorów ze skalą kształtowania zysków, takich jak np. badania Leuz i in. [2003]. Brakuje także nawiązania do publikacji, w których porównuje się narzędzia wykrywania manipulacji księgowych proponowane przez akademików z narzędziami o charakterze komercyjnym, np. Price i in. [2011]. Generalnie da się zauważyć pewną dysproporcję pomiędzy odniesieniami do prac opublikowanych przed i po 2010 r. (niestety na korzyść tych pierwszych).

3. Ogólna ocena głównego osiągnięcia rozprawy

Jak wynika z poprzedniej części recenzji, główne dokonanie tego doktoratu zostało przedstawione w rozdziale czwartym i piątym. Pozwolę sobie powtórzyć wniosek, który sformułowałem w początkowych fragmentach recenzji, że Doktorantowi udało się osiągnąć zasadniczy cel rozprawy polegający na opracowaniu narzędzia detekcji praktyk kreatywnej rachunkowości (określonego przez Doktoranta mianem C-score), które po pewnych zabiegach o charakterze dostosowawczym może być wykorzystywane w odniesieniu do podmiotów funkcjonujących w różnym otoczeniu gospodarczym i którego zastosowanie może przyczynić się do ograniczania skali zjawiska wprowadzania w błąd użytkowników informacji finansowych.

Zaproponowany przez Doktoranta wskaźnik C-score opracowany został przy wykorzystaniu metodologii analizy dyskryminacyjnej.

Przy ustalaniu wskaźników, które zostały uwzględnione w funkcji dyskryminacyjnej, kierowano się zarówno względami merytorycznymi, jak również odpowiednimi metodami statystycznymi. Miały one zapewnić dobór wskaźników o relatywnie niskim poziomie skorelowania oraz dużej zdolności dyskryminacyjnej między populacjami.

Przed przystąpieniem do konstruowania modeli sprawdzone zostały podstawowe założenia leżące u podstaw analizy dyskryminacyjnej.

Podstawowy model analizy dyskryminacyjnej (model I) wyznaczony został na podstawie próby złożonej z 50 przedsiębiorstw prowadzących działalność w Kuwejcie, na którą składa się 25 przedsiębiorstw dokonujących „manipulacji księgowych” i 25 przedsiębiorstw wolnych od takich praktyk. Występujące w modelu wskaźniki (predyktory) to wskaźnik płynności szybkiej (*the quick acid ratio*), relacja zysku netto do zysku na sprzedaży (*net profit/gross profit ratio*), rentowność netto aktywów trwałych (*net profit/fixed assets ratio*), udział zobowiązań w aktywach ogółem (*total liabilities/total assets ratio*),



udział *accruals* w aktywach ogółem (*total accruals/total assets ratio*), udział należności w przychodach ze sprzedaży (*accounts receivable/sales ratio*) oraz w aktywach ogółem (*accounts receivable/total assets ratio*), a także wskaźnik zysku na jedną akcję (*earnings per share*). Dla zbudowanego modelu wartość statystyki lambda Wilksa wynosi 0,33. Ogólna trafność klasyfikacji wynosi 96%, przy czym w odniesieniu do przedsiębiorstw niedokonujących manipulacji jest ona szczególnie wysoka i wynosi 100%, zaś w przypadku firm manipulatorów wynosi ona 92%. Wartość funkcji dyskryminacyjnej dla centroidów obu populacji wynosi odpowiednio 1,396 dla firm dokonujących manipulacji oraz -1,396 dla grupy wolnej od takich praktyk. Największy (najmniejszy) wpływ na wartość funkcji dyskryminacyjnej ma wskaźnik *accruals* (wskaźnik płynności szybkiej).

W kolejnym kroku model I poddano procedurze walidacyjnej, starając się stworzyć w tym celu odpowiednią próbę. Na podstawie zidentyfikowanych dla próby analitycznej predyktorów zbudowano model II dla próby spółek z USA i Wielkiej Brytanii. Ogólna trafność klasyfikacji dla tak zbudowanego modelu pogorszyła się i wynosi 74,9%. Odpowiednie wskaźniki trafności klasyfikacji dla przedsiębiorstw dokonujących i niedokonujących manipulacji wyniosły odpowiednio 68,1% i 78,4%. Ranking wpływu poszczególnych zmiennych na wartość funkcji dyskryminacyjnej zasadniczo nie uległ większym zmianom. Krzyżowe porównanie skuteczności prognostycznej modelu I i modelu II dla próby analitycznej oraz próby walidacyjnej doprowadziło Autora do konkluzji, że model I sprawdza się lepiej dla spółek funkcjonujących na rynkach wschodzących z relatywnie słabo rozwiniętym systemem audytu przedsiębiorstw oraz o łagodnym systemie podatkowym. Z kolei model II jest bardziej adekwatny dla rynków rozwiniętych. Dodatkowo rezultaty analiz przeprowadzonych na próbie analitycznej wskazują, że trafność klasyfikacyjna obu modeli analizy dyskryminacyjnej jest wyższa aniżeli modeli *accruals* czy modelu Beneisha.

Nie mam większych zastrzeżeń związanych z samą procedurą budowy modeli analizy dyskryminacyjnej. W mojej ocenie została ona przeprowadzona zgodnie z zasadami sztuki, a uzyskane rezultaty są w obrębie prób badawczych dość przekonujące. Muszę jednocześnie zastrzec, że w przypadku niektórych predyktorów wyjaśnienie kierunku ich wpływu na wartość funkcji dyskryminacyjnej mogłoby być bardziej wyraziste.

4. Uwagi krytyczne, postulaty i sugestie dotyczące procedury badawczej

Rola recenzenta każe dzielić się wątpliwościami, zgłaszać zastrzeżenia i formułować postulaty umożliwiające podniesienie wartości uzyskanych rezultatów na jeszcze wyższy poziom. Poniżej pozwoliłem sobie na zamieszczenie kilku uwag, które mają być przede wszystkim zachętą do kontynuacji badań i nie zmieniają mojej pozytywnej całościowej oceny pracy.

Dobór próby badawczej

Jednym z istotniejszych problemów w badaniach nad zbiorami firm zagrożonych dokonywaniem (a także dokonujących) manipulacji księgowych jest dobór próby. W wielu państwach (podejrzewam, że także w Kuwejcie, choć nie jest to wyraźnie określone w pracy) brakuje baz danych o jednostkach dokonujących manipulacji. W tej sytuacji każdy z badaczy dobiera własną próbę – lepiej lub gorzej. Takie podejście zastosowane zostało także przez Autora recenzowanej pracy.

Wyjściowa próba badawcza obejmowała 89 przedsiębiorstw przemysłowych (*industrial companies* – s. 86), z czego finalnie badaniem objętych zostało 50 spółek, dla których dostępne były sprawozdania finansowe oraz raporty audytorów za lata 2010–2015. Na podstawie sprawozdań finansowych, raportów z badań, przeglądów dokonanych przez niezależnych audytorów oraz analiz własnych Autora,



ustalono metodą ekspercką (do tej kwestii odniosę się jeszcze w późniejszej części recenzji) listę 25 spółek dokonujących manipulacji księgowych, a także listę 25 spółek wolnych od tego typu zabiegów. Te 50 spółek stanowiło tzw. próbę analityczną, dla której szacowano zarówno modele regresyjne, jak i modele analizy dyskryminacyjnej. Takie postępowanie – choć nie do końca poprawne statystycznie – jest stosowane w badaniach prowadzonych na świecie. W idealnym przypadku próba powinna być w miarę jednorodna pod względem branży i rozmiarów działalności, a także powinna odzwierciedlać faktyczny udział firm dokonujących i niedokonujących manipulacji księgowych. Ten ostatni postulat jest w praktyce bardzo często trudny do spełnienia, a jedną z możliwych metod rozwiązania problemu jest zastosowanie próby dobieranej. Choć w recenzowanym tekście kilkakrotnie pojawia się stwierdzenie wskazujące, że mamy do czynienia z próbą dobieraną ze względu na branżę (np. na str. 227 *industry matched sample*), należy podkreślić, że zaprezentowanie przekroju sektorowego wnioskoby wartość dodaną do pracy.

Zupełnie inaczej przedstawia się sytuacja dotycząca tzw. próby walidacyjnej. Z wyjściowej próby 2200 spółek notowanych na giełdach w USA i Wielkiej Brytanii, Autor wyselekcjonował 1305 jednostek (w tym 448 dokonujących manipulacji), dla których dostępne były dane umożliwiające przeprowadzenie zaprojektowanych badań. Przy czym, o ile w przypadku próby analitycznej do identyfikacji spółek dokonujących „manipulacji księgowych” wykorzystane zostało podejście eksperckie, w przypadku próby walidacyjnej Doktorant zastosował powszechnie stosowane przez badaczy rozwiązanie polegające na wykorzystaniu danych o stwierdzonych manipulacjach czy o przesłankach wystąpienia manipulacji, dostępnych w dwóch bazach danych: AAERS – *Accounting and Auditing Enforcement Releases* (USA) oraz *Financial Reporting Review Panel* (UK). Także w przypadku próby walidacyjnej brak jest zestawień ilustrujących przynależność sektorową, choć na str. 87 Autor wspomina, że spółki z tej próby reprezentują różne sektory przemysłu (*different industrial sectors*).

W tym miejscu nasuwa się pytanie, na ile tak zdefiniowana próba walidacyjna wydaje się właściwa do oceny trafności klasyfikacji proponowanego modelu. Na str. 88–89 Autor stwierdza, że już pobieżna analiza porównawcza warunków funkcjonowania spółek wchodzących w skład obu grup wskazuje na ich duże podobieństwo. Przywołuje przy tym silne powiązania finansowo-handlowe łączące Kuwejt z USA i Wielką Brytanią. Co ciekawe, jako jeden z obszarów podobieństw Doktorant wskazuje regulacje podatkowe, z czym nie do końca mogę się zgodzić, gdyż dla przykładu w Kuwejcie podatkiem CIT objęte są przede wszystkim podmioty z udziałem kapitału zagranicznego, z kolei w USA i Wielkiej Brytanii CIT ma charakter podatku powszechnego. Trudno także mówić o zbliżonej konstrukcji podatku CIT w tych państwach. Wiadomo jednocześnie, że regulacje podatkowe, a w szczególności zakres zbieżności definicji „wyniku księgowego” i „dochodu podatkowego” (*book-tax conformity*), są istotnym czynnikiem skali kształtowania zysków [Blaylock i in., 2015]. W tym miejscu rola recenzenta każe mi także zgłosić bardziej generalną wątpliwość co do porównywalności obu prób ze względu na charakterystykę otoczenia instytucjonalnego, w szczególności w zakresie odnoszącym się do *corporate governance*. Z jednej strony istnieją opracowania, które wskazują, że zarówno skala [Leuz i in., 2003], jak i stosowane mechanizmy kształtowania wyników finansowych [Enomoto i in., 2015]) determinowane są przez poziom ochrony praw inwestorów. Jednocześnie według ostatniego raportu Doing Business, przygotowywanego cyklicznie przez Bank Światowy, indeks ochrony praw inwestorów mniejszościowych (*Protecting Minority Investors Score*) dla Kuwejtu kształtuje się na poziomie 58,33, zaś dla USA i Wielkiej Brytanii odpowiednio na poziomie 64,67 i 75,00. Innymi słowy, w rankingu poziomu ochrony praw inwestorów wśród 190 sklasyfikowanych przez Bank Światowy państw Kuwejt zajmuje 72 miejsce, podczas gdy USA i Wielka Brytania zostały sklasyfikowane odpowiednio na 50 i 15 miejscu. W tym stanie rzeczy zdecydowanie bardziej pożądanym rozwiązaniem byłoby stworzenie



próby walidacyjnej obejmującej spółki funkcjonujące w zbliżonym co do charakterystyki Kuwejt otoczeniu instytucjonalnym.

Uzasadniając taki, a nie inny, dobór próby walidacyjnej, Doktorant odwołuje się także do podobieństw pomiędzy spółkami reprezentującymi obie populacje ze względu na ich przynależność sektorową, specyfikę zarządzania, strukturę własności, strukturę kapitału czy wielkość przedsiębiorstwa. Niestety, stwierdzenia powyższe nie zostały poparte jakąkolwiek statystyką opisową – czy bardziej formalnym testem, np. na różnice wartości średnich, co w moim przekonaniu w dużej mierze mogłoby rozwiązać wątpliwości dotyczące doboru próby walidacyjnej. Brak takich testów jest o tyle zaskakujący, że narzędzia testowania są Autorowi znane, gdyż z powodzeniem stosuje je do porównania obu prób (niestety, tylko w odniesieniu do zmiennych, które finalnie znalazły się w modelach dyskryminacyjnych) w rozdziale 5 (zakres tych testów doskonale ilustruje schemat 5.9 na str. 259).

Dokonując oceny doboru próby badawczej, należy stwierdzić, że w zakresie próby analitycznej Doktorant zastosował często stosowaną przez badaczy zasadę 50 : 50. Do dostępnej próby spółek dokonujących manipulacji księgowych dobrano równie liczną próbę przedsiębiorstw wolnych od takich praktyk. Jednocześnie zgodnie z zasadami sztuki poprawność klasyfikacyjną modelu sprawdzano dla próby odmiennej od tej, która służyła do jego skonstruowania. Jednoznaczna ocena wyboru próby badawczej nie jest łatwa, trzeba bowiem wczuć się w rolę badacza, który zazwyczaj musi dokonać wyboru pomiędzy próbą jak najlepiej odzwierciedlającą rzeczywistą proporcję obu grup podmiotów występujących w gospodarce (lub branży itp.) a jej wyrazistością pozwalającą na wyodrębnienie dobrych predyktorów manipulacji księgowych. Dylematy wynikające z niedostatków próby trzeba rozstrzygać dla każdego zadania badawczego i tak też postąpił Doktorant, decydując się z jednej strony na małą liczebność próby analitycznej, charakteryzującej się jednocześnie danymi o wysokiej jakości oraz z drugiej strony obfitą, wychodzącą naprzeciw postulatowi tzw. zewnętrznej oceny modelu, ale nie do końca porównywalną co do warunków funkcjonowania, próbę walidacyjną. Z punktu widzenia recenzenta rozprawy, unikatowa pod względem jakości danych próba analityczna wystarczająco rekompensuje niedostatki próby walidacyjnej.

Empiryczna definicja firmy dokonującej manipulacji księgowych

Stosowane w literaturze przedmiotu empiryczne definicje firmy dokonującej manipulacji księgowych nie są jednolite. W przypadku najbardziej rozwiniętego pod względem badań w tym obszarze rynku amerykańskiego wielu badaczy za firmy dokonujące manipulacji księgowych uznaje wyłącznie te firmy, w przypadku których Komisja Papierów Wartościowych i Giełd (SEC) stwierdziła występowanie nadużyć [Dechow i in. 2011]. Informacje o takich przypadkach, po zakończeniu postępowań, zamieszczane są w publicznie dostępnej bazie AAER (*Accounting and Auditing Enforcement Releases*). Inni badacze stosują szerszą definicję i do grupy firm dokonujących manipulacji księgowych zaliczają także przypadki firm, które dokonały tzw. restatements (wprowadziły korekty do sprawozdań finansowych). Podejście takie niesie jednak ze sobą zagrożenie traktowania firm, które korygują *de facto* błędy księgowe, jako dokonujących manipulacji. W przypadku spółek amerykańskich informacje o korektach do sprawozdań można znaleźć m.in. w bazie *The Government Accountability Office (GAO) Financial Statement Restatement Database*. Baza ta opracowana została na podstawie publicznie dostępnych informacji przez firmę Lexis Nexis i obejmuje wyłącznie lata 1997–2005. Jeszcze inni badacze korzystają z bazy *Stanford Law Database on Shareholder Lawsuits* zawierającej dane o procesach wszczętych przez akcjonariuszy wobec spółek. Wszczęcie procesu nie zawsze jednak musi być związane z nadużyciami



księgowymi, co sprawia, że kwalifikacja podmiotów jako dokonujących manipulacji może być także obciążona błędem.

Z powyższego wynika, że nawet kiedy dostępne są bazy danych, podział na firmy dokonujące manipulacji księgowych i firmy niestosujące takich praktyk, jest w jakimś stopniu arbitralny. Oczywiście aktualny jest postulat, że definicja jednostki dokonującej manipulacji księgowych powinna być możliwie jasna i jednoznaczna, jednakże natura materii sprawia, że w praktyce jest on bardzo trudny do spełnienia.

Wspomniany problem arbitralności odnosi się także w pewnym stopniu do recenzowanej pracy. Choć trzeba w tym miejscu przyznać, że Doktorant dołożył wielu starań, aby w obrębie obu prób zobiektywizować proces identyfikacji firm dokonujących manipulacji księgowych. W szczególności na duże uznanie zasługuje proces badawczy będący *de facto* bardzo wnikliwym przeglądem sprawozdań finansowych oraz raportów audytorów, uzupełnionym o informacje z przeprowadzonych wywiadów, w rezultacie którego możliwe było szczegółowe zidentyfikowanie obszarów manipulacji księgowych. Owoce ogromu wykonanej pracy jest rozdział 2.6 oraz rezultaty zestawione w tabelach 2.7–2.23. Nie znalazłem jednak w pracy informacji o kryteriach, które ostatecznie zdecydowały o kwalifikacji spółek z próby analitycznej do jednej z dwóch grup – manipulatorów i niemanipulatorów. Praca z pewnością mogłaby zyskać na wartości, gdyby udało się wspomniane kryteria wyrazić w postaci jakiegoś syntetycznego miernika.

Zestaw zmiennych objaśniających

Kolejną istotną kwestią, którą chciałbym poruszyć, jest dobór wyjściowego zestawu zmiennych objaśniających, na podstawie których wyznaczana jest funkcja dyskryminacyjna. Jest to bowiem wspólny problem różnorodnych podejść do modelowania mikroekonometrycznego. W przypadku większości badań z zakresu szeroko rozumianej rachunkowości kreatywnej wstępny zbiór zmiennych objaśniających ustalany jest zazwyczaj według dwóch kryteriów: ich popularności w literaturze przedmiotu oraz relatywnie dobrych wyników w prognozowaniu we wcześniejszych badaniach. Podobne podejście zastosował także Autor recenzowanej pracy, który zidentyfikował 52 wskaźniki przynależne do jednej z pięciu grup miar charakteryzujących płynność, wypłacalność finansową, rentowność, efektywność gospodarowania oraz strukturę finansowania. Ostatecznie w wyniku przeprowadzonej analizy w modelu dyskryminacyjnym znalazło się 8 wcześniej przeze mnie wymienionych wskaźników.

Dominacja wskaźników finansowych wśród determinant praktyk rachunkowości kreatywnej nie jest niczym nadzwyczajnym. Po prostu w obliczu braku dobrych teorii oraz relatywnie dobrej dostępności danych finansowych badacze bazują na statystycznych własnościach zmiennych oraz własnościach próby. U podstaw większości badań nadal – także tych prowadzonych w ramach recenzowanej rozprawy – leży przekonanie, że wyjaśnienia przyszłych stanów firmy należy szukać w raportach finansowych z poprzednich okresów. W konsekwencji koncentrujemy się na eksplorowaniu wskaźników finansowych, pomijając inne, często bardziej znaczące determinanty manipulacji księgowych. Zaliczyć do nich można przede wszystkim różne charakterystyki nadzoru korporacyjnego, takie jak struktura własności (np. udział insiderów, struktura udziałów różnych typów inwestorów instytucjonalnych itp.), liczebność i skład rady dyrektorów (np. udział ekspertów z zakresu finansów), stosowane systemy zachęt dla menedżerów (np. udział opcji na akcje w łącznym wynagrodzeniu) czy charakterystyka audytora (np. przynależność do tzw. Wielkiej Czwórki, staż w badaniu podmiotu itp.). Wymienione powyżej tzw. czynniki „miękkie” – w mojej opinii – mogłyby znacznie zwiększyć zdolności



klasyfikacyjne modelu, zwłaszcza że literatura, w tym wiele pozycji, które możemy znaleźć w bibliografii recenzowanej pracy, wskazuje na ich istotne znaczenie.

Inną, choć nie mniej ważną kwestią, związaną z doбором zmiennych objaśniających jest relatywnie uboga reprezentacja zmiennych reprezentujących motywów manipulacji księgowych. Na podstawie rocznego sprawozdania finansowego można wyznaczyć wystarczającą liczbę wskaźników reprezentujących tzw. *market incentives* (motywy rynkowe). Poza (dobrze reprezentowanymi w badaniu) wskaźnikami zadłużenia aproksymującymi motyw związany z covenantami finansowymi mogłyby to być wskaźniki informujące o skali emisji akcji czy szerzej z wielkością zapotrzebowania na tzw. finansowanie zewnętrzne. Z kolei łącząc dane ze sprawozdań finansowych z danymi rynkowymi, można łatwo wyznaczyć wskaźniki M/B aproksymujące stan rynku. Wydaje się, że uwzględnienie powyższych zmiennych mogłoby istotnie zwiększyć skuteczność proponowanych przez Doktoranta narzędzi, zwłaszcza że bardziej współczesne modele predykcji manipulacji księgowych (np. wspomniany wskaźnik F-score) tego typu charakterystyki coraz częściej uwzględniają. O wielu z nich Autor wspomina w podrozdziale 1.4 poświęconym motywom manipulacji księgowych.

Metodologia szacowania modeli accruals

W większości badań poświęconych kształtowaniu zysków czy pomiarze ich jakości (*earnings quality*) stosuje się metodologię zmian niepieniężnych aktywów obrotowych netto (*accruals*). Fakt ten znajduje także odzwierciedlenie w recenzowanej rozprawie. Spośród ilościowych modeli wykrywania manipulacji księgowych opisywanych w rozdziale 3 rozprawy, najwięcej miejsca poświęcono właśnie metodologii accruals. Ponadto prawie wszystkie modele zmian niepieniężnych aktywów obrotowych netto (wyjątek stanowią wczesne koncepcje: model Healy oraz model de Angelo) wraz z modelem Beneisha zostały poddane empirycznej weryfikacji, gdyż w zamierzeniu Doktoranta mają stanowić punkt odniesienia do oceny proponowanego narzędzia C-score.

Filozofia modeli accruals zakłada wykorzystanie różnic pomiędzy memoriałowym a kasowym podejściem do pomiaru dokonań operacyjnych przedsiębiorstw. Przyjmuje się przy tym, że sama zmiana niepieniężnych aktywów obrotowych netto nie jest celem kształtowania wyniku finansowego, lecz naturalną konsekwencją tego procesu. Z tego powodu w badaniach próbuje się oddzielić normalne (nieuznaniowe, *non-discretionary*) zmiany niepieniężnych aktywów obrotowych netto od zmian uznaniowych (*discretionary accruals*). Te pierwsze są rezultatem normalnego przebiegu procesów gospodarczych, a ich wielkość powiązana jest zazwyczaj ze zmianami poziomu aktywności, stosowaną strategią biznesową, specyfiką branży czy szeroko rozumianą sytuacją makroekonomiczną. Z kolei zmiany uznaniowe są konsekwencją kształtowania struktury transakcji oraz zabiegów księgowych podporządkowanych kształtowaniu zysków. Syntetyczny opis tej metodologii można znaleźć w pracy na str. 128–130.

Z przeglądu modeli dokonanego przez Doktoranta bardzo dobrze wyłania się proces ewolucji modeli accruals. O ile początkowo modele szacowane były dla indywidualnych przedsiębiorstw, współcześnie dominuje podejście polegające na ich wyznaczeniu dla grupy przedsiębiorstw (*cross-sectional analysis*) przynależnych do tej samej branży. Każdy kolejny model zawiera także propozycję modyfikacji zestawu zmiennych objaśniających. W czasie zmieniało się również podejście do szacowania całkowitych zmian niepieniężnych aktywów obrotowych netto. Wczesne prace wykorzystywały tzw. podejście bilansowe (zmiana niepieniężnych aktywów obrotowych netto z wyłączeniem opodatkowania i amortyzacji). Obecnie dominuje podejście polegające na tym, że accruals wyznaczone są jako różnica pomiędzy



wynikiem finansowym (zwykle w wyłączeniu efektów zdarzeń o charakterze nadzwyczajnym oraz efektów operacji niekontynuowanych) a przepływami z działalności operacyjnej.

Lektura rozprawy daje podstawy (a w zasadzie pewność), że Doktorant zaznajomił się ze stosowanymi powszechnie w literaturze światowej modelami zmian niepieniężnych aktywów obrotowych netto, a pomimo tego zdecydował się na – mocno dyskusyjny w moim przekonaniu – eksperyment polegający na wyznaczaniu accruals nie na bazie przepływów z działalności operacyjnej, lecz na bazie całkowitych przepływów pieniężnych. Wybór swój Doktorant uzasadnia zamiarem uwzględnienia manipulacji księgowych, które mogą się odnosić do wszystkich obszarów działalności przedsiębiorstwa (str. 131 rozprawy).

Rozumiem przesłanki, jakimi kierował się Doktorant. Zgodnie z logiką pracy operacje polegające na sprzedaży aktywów i ich leasingu zwrotnym czy wykupie akcji własnych zwiększającym EPS to instrumenty, które – tak samo jak tworzenie i rozwiązywanie odpisów i rezerw czy klasyfikacja operacji wynikowych – mogą być wykorzystywane w celach manipulacyjnych. Innymi słowy, skoro rachunkowość kreatywna może dotyczyć różnych obszarów działalności przedsiębiorstwa, a celem rozprawy jest opracowanie uniwersalnego narzędzia umożliwiającego identyfikację podmiotów dokonujących manipulacji księgowych, „logiczne” wydaje się zastosowanie szerokiej definicji przepływów pieniężnych. W mojej ocenie, podążając takim tokiem rozumowania, Doktorant wpadł w swoistą pułapkę, gdyż modele accruals, które stosuje w pracy, koncentrują się przede wszystkim na działalności operacyjnej. Wskazuje na to choćby zestaw zmiennych objaśniających odnoszących się głównie do takich pozycji, jak przychody ze sprzedaży czy należności, a niekiedy także koszty działalności operacyjnej oraz zobowiązania operacyjne. Według mojego stanu wiedzy, badacze stosują odrębne narzędzia do identyfikacji manipulacji księgowych w różnych obszarach działalności przedsiębiorstwa. Stąd pomysł Doktoranta uznaję za oryginalny, ale jednocześnie mocno kontrowersyjny.

Analiza porównawcza modeli accruals z modelem Beneisha oraz modelem C-score

Na str. 176 rozprawy znajdujemy informację o tym, w jaki sposób określana była trafność klasyfikacyjna modeli accruals. W przypadku modeli budowanych na podstawie danych przekrojowo-czasowych obiekt był klasyfikowany jako dokonujący manipulacji, jeżeli wartość bezwzględna uznaniowych accruals była większa od zera lub od wartości średniej. Z kolei w przypadku modeli bazujących na danych przekrojowych, obiekt był klasyfikowany jako manipulator, jeżeli obserwowana wartość accruals różniła się od wartości teoretycznej wyznaczonej na podstawie modelu regresji. Tak dokonana klasyfikacja następnie była konfrontowana z faktyczną przynależnością obiektu do jednej z dwóch grup – manipulatorów i niemanipulatorów.

Moim zdaniem, zastosowane rozwiązanie jest dużym uproszczeniem, zwłaszcza w przypadku modeli budowanych dla danych przekrojowych. Literatura przedmiotu dostarcza bardziej zaawansowanych rozwiązań umożliwiających porównanie różnych narzędzi wykrywania manipulacji księgowych. Można tu przywołać, stanowiącą kanon literatury przedmiotu, pracę Dechow i in. [1995], w której do oceny różnych modeli accruals proponuje się wykorzystanie podejścia symulacyjnego.

Z kolei, jeżeli chodzi o porównywanie skuteczności różnych narzędzi detekcji manipulacji księgowych, to jednym z możliwych rozwiązań jest wykorzystanie łącznej estymacji pozornie niezależnych regresji (SUR), w których zmienna objaśniana miałaby charakter dychotomiczny, a jedną ze zmiennych objaśniających byłaby wartość uznaniowych accruals, wskaźnik M-score czy będący autorską



propozycją Doktoranta C-score. Przykład zastosowania tego podejścia można znaleźć w pracy Price i in. [2011].

Choć zastosowane przez Doktoranta do porównań poszczególnych narzędzi detekcji manipulacji księgowych rozwiązanie nie jest wolne od zastrzeżeń natury metodologicznej, z perspektywy realizacji celu głównego rozprawy uznaję je jednak za wartościowe. Każda propozycja nowej metody powinna być konfrontowana z dotychczas istniejącymi i tak też postrzegam podjętą przez Doktoranta próbę.

5. Pytania do Doktoranta

Po lekturze rozprawy nasuwają mi się trzy pytania, które chciałbym skierować do Doktoranta:

- 1) Badania na rachunkowością kreatywną, z powodu relatywnie dobrej dostępności do danych, dotyczą głównie spółek publicznych. Zatem, czy, a jeżeli tak, to jakie znaczenie dla doboru zmiennych objaśniających (predyktorów) do modelu może mieć prywatna forma własności spółki? Innymi słowy, czy istnieje możliwość budowania uniwersalnych modeli detekcji manipulacji księgowych?
- 2) We wnioskach końcowych Doktorant wspomina, że jednym z czynników, który różnicuje przydatność modelu I i modelu II, są stosowane standardy sprawozdawczości finansowej. Proszę wskazać kilka przykładowych różnic pomiędzy amerykańskimi standardami sprawozdawczości (US GAAP) a Międzynarodowymi Standardami Sprawozdawczości Finansowej (IFRS), które mogą wpływać na zróżnicowaną sprawność proponowanych modeli.
- 3) Rezultaty badań prowadzonych na świecie wskazują, że wprowadzanie Ustawy Sarbanesa-Oxleya w 2002 r. (SOX) doprowadziło do ograniczenia zjawiska kształtowania zysków typu memoriałowego (*accrual-based earnings management*) i zwiększyło skalę kształtowania zysków typu realnego (*real earnings management*). Czy modele detekcji manipulacji księgowych, takie jak M-score czy proponowany C-score, są odporne na praktyki kształtowania zysków w ujęciu realnym?

6. Podstawowe wartości recenzowanej rozprawy

Recenzowana dysertacja ma wiele zalet, jednak do najważniejszych spośród nich zaliczyłbym trzy poniżej wymienione:

- 1) Po pierwsze, rozprawa koncentruje się na niezmiernie aktualnym zagadnieniu, jakim jest rachunkowość kreatywna. Motywy i konsekwencje ingerencji zarządzających w procesy raportowania wyników – podobnie jak proponowany wskaźnik C-score – mają charakter wielowymiarowy. Właśnie z tego powodu potencjalnymi beneficjentami uzyskanych danych może być szerokie grono użytkowników informacji finansowych, tj. dostarczyciele kapitału, tzw. gatekeepers (audytorzy, analitycy czy instytucje regulujące rynek), a także kontrahenci, pracownicy czy także organy administracji podatkowej.
- 2) Po drugie, współczesne badania nad praktykami kreatywnej rachunkowości bazują na dużych próbach i metodologii statystycznej. Z tym wiążą się określone problemy z identyfikacją tych procesów oraz osób za nie odpowiedzialnych. Praktycy – jak potwierdza niniejsza rozprawa – są w stanie obserwować indywidualne przypadki manipulacji księgowych. Badacze skoncentrowali się na dużych zbiorach danych i statystycznie istotnych determinantach, które nie są specjalnie interesujące dla praktyki. Poszukiwanie syntetycznych, intuicyjnych miar,



takich jak wskaźnik C-score, wydaje się wyjściem naprzeciw oczekiwaniom praktyków i stwarza szansę na ich upowszechnienie.

- 3) Wreszcie po trzecie, zrealizowana praca badawcza dobrze wpisuje w nurt tego typu badań prowadzonych na świecie, w szczególności – mimo pewnych uproszczeń – warstwa metodologiczna pracy generalnie może być uznana za aktualną. Jednocześnie Doktorant ma świadomość, że wybrany ostatecznie zbiór predyktorów jest „przypisany do próby”, a sprawność modelu w dużej mierze zależy od uwarunkowań gospodarczych i prawnych, w jakich firmy działają.

*

Na zakończenie pozostaje mi formalnie stwierdzić, że przedstawiona do recenzji dysertacja pana Arkana Thomasa Kłoczy'ego pod tytułem „Rachunkowość kreatywna: badanie empiryczne z wykorzystaniem wskaźnika C-score” spełnia wymogi Ustawy z dnia 14 marca 2003 roku o stopniach naukowych i tytule naukowym oraz o stopniach i tytule w zakresie sztuki stawiane rozprawom doktorskim w dziedzinie nauk ekonomicznych w dyscyplinie finanse. W związku z tym wnioskuję do Rady Wydziału Nauk Ekonomicznych i Zarządzania Uniwersytetu Szczecińskiego o jej przyjęcie i dopuszczenie do publicznej obrony.

Poznań, dnia 18 lutego 2019 r.

Michał Kaldowski