

Joanna Kisielińska

Szkoła Główna Gospodarstwa Wiejskiego w Warszawie

KLASYFIKACJA GOSPODARSTW ROLNICZYCH NA PODSTAWIE DANYCH W POSTACI SZEREGÓW CZASOWYCH

1. Cele i zasady konstrukcji modeli

Celem przeprowadzonych badań było opracowanie klasyfikacyjnych modeli prognozujących sytuację finansową gospodarstw rolniczych. Zadaniem ich było podzielenie badanej zbiorowości na dwie klasy: gospodarstwa dobre, czyli takie, które rokuja rozwój i pełne wykorzystanie środków, którymi dysponują, oraz słabe, w przypadku których ewentualna pomoc może jedynie odsunąć groźbę likwidacji.

Jako kryterium przydziału do klasy przyjęto wielkość uzyskiwanego dochodu rolniczego. Dla gospodarstw określenie zysku jest zadaniem dosyć złożonym, ponieważ wymaga założenia pewnych umownych i hipotetycznych kosztów (jak wynagrodzenie za pracę własną rolnika, oprocentowanie własnego kapitału czy renta gruntowa).

Ponieważ zbiorowość ma być podzielona na dwie klasy, należy przyjąć pewną graniczną wartość dochodu rolniczego¹. Jeśli dochód gospodarstwa jest od niej mniejszy – należeć będzie ono do klasy I, w przeciwnym wypadku do klasy II. Tak określone zasady podziału oznaczają, że klasę I stanowią gospodarstwa słabsze, natomiast II – lepsze. Wartość graniczną dochodu rolniczego przyjęto na poziomie mediany, co pozwoliło zachować równoliczność klas². Podstawę konstrukcji modeli klasyfikacyjnych stanowił szeroki zestaw wskaźników finansowych uzupełniony

¹ Postępowanie takie polega w istocie na zamianie zmiennej ilościowej na jakościową i jest często stosowane w praktyce. Zawadzki i Babis [1996], klasyfikując przedsiębiorstwa, jako wartość graniczną rentowności zastosowali zero. Gruszczyński [2002], Tarczyński [1996; 2001] tworzyli klasy, dzieląc spółki giełdowe na te, dla których stopa zwrotu z akcji jest większa lub mniejsza od stopy zwrotu z Warszawskiego Indeksu Giełdowego.

² Można przyjąć również inne wartości, w zależności od tego, jaką część gospodarstw chcemy zaliczyć do słabych (może to być przede wszystkim zero). Identyfikacja gospodarstwa jako słabego wskazuje na niecelowość kierowania do niego środków finansowych. Pomoc należałoby koncentrować na jednostkach dobrych, jedynie wówczas bowiem można liczyć na restrukturyzację polskiego rolnictwa.

o pewne dodatkowe informacje o gospodarstwach. Zapewnienie modelom funkcji prognostycznej wymaga, aby cechy opisujące gospodarstwa były opóźnione względem zastosowanego kryterium klasyfikacji. Sytuacja finansowa będzie prognozowana na podstawie wartości wskaźników finansowych z przeszłości.

Modele klasyfikacyjne budowano dwiema metodami – stosując liniową analizę dyskryminacyjną oraz sieci neuronowe. Dobór taki pozwala porównać podejście liniowe z nieliniowym. Wyraźna przewaga modeli sieciowych wskazuje na nieliniowość badanych zjawisk. Stanowiąc może przesłankę do poszukiwania nieliniowych modeli funkcyjnych, jeśli istotna jest postać analityczna. W przypadku analizy dyskryminacyjnej można wówczas zastosować transformację zmiennych (tak jak w analizie regresji), można również wykorzystać modele logitowe lub probitowe.

Większość modeli klasyfikujących obiekty gospodarcze wykorzystuje informacje o jednostce z jednego roku, zwykle poprzedzającego konstruowaną prognozę. Można przypuścić, że zbudowanie modelu na podstawie danych pochodzących z większej liczby okresów (dane w postaci szeregów czasowych) pozwoliłoby poprawić jakość modeli.

W literaturze można znaleźć przykłady prowadzenia analizy dyskryminacyjnej na podstawie danych w postaci szeregów czasowych. Krzyżko [1990, s. 115-145] do klasyfikacji obiektów opisanych jednowymiarowymi szeregami czasowymi wykorzystuje wielomianowe funkcje trendu. W przypadku wielowymiarowych procesów stochastycznych natomiast proponuje funkcję dyskryminacyjną opartą na bayesowskiej regule klasyfikacyjnej. Szeregi czasowe reprezentujące cechy obiektów muszą być wówczas procesami autoregresyjnymi. Strahl i Markowska [2004] przedstawiają propozycje wykorzystania do klasyfikacji bezwzorcowej funkcji trendu oraz miar odległości między macierzami danych przekrojowo-czasowych. Gruszczyński zaś [2002, s. 194] proponuje wprowadzenie do modelu jako zmiennych objaśniających cech z kilku okresów wcześniejszych. Gatelny [1999, s. 55, 109, 118], budując sieci neuronowe do prognozowania finansowego, wykorzystuje średnie ruchome z różnymi wielkościami stałej wygładzania.

Oparcie klasyfikacji gospodarstw rolniczych na funkcjach trendu nie jest właściwe ze względu na wymiar problemu – kilkaset gospodarstw opisanych wieloma cechami. Dopasowywanie funkcji trendu do każdego wskaźnika finansowego dla każdego gospodarstwa jest zadaniem praktycznie niewykonalnym i niecelowym. Rodzaj funkcji trendu może być bowiem istotny w przypadku badania wskaźników makroekonomicznych, natomiast dla wskaźników finansowych jednostki gospodarczej raczej znaczenia nie ma.

Zastosowanie propozycji Gruszczyńskiego do klasyfikacji gospodarstw także wydaje się niewskazane z kilku przyczyn. Pierwsza z nich to znaczne zwiększanie wymiaru przestrzeni zmiennych niezależnych. Pojedyncza cecha bowiem będzie reprezentowana przez cały szereg zmiennych reprezentujących różne momenty z przeszłości. Po drugie, w szeregach czasowych bardzo prawdopodobne jest wystąpienie autokorelacji, dzięki czemu poszczególne zmienne niezależne mogą być

skorelowane. Zastosowanie analizy dyskryminacyjnej jest wówczas bardzo utrudnione (problem odwracania źle uwarunkowanej macierzy). Ze względu na swobodny dobór współczynników dla cech pochodzących z różnych chwil mogą się pojawić pewne dodatkowe problemy. Może się bowiem okazać, że wartość cechy z chwili wcześniejszej ma większy wpływ na uzyskaną ocenę niż wartość cechy z chwili późniejszej. Zależności takich można się spodziewać w przypadku procesów cyklicznych, wskaźniki finansowe wyznaczone dla gospodarstw z wielu okresów raczej cykliczności nie wykażą. Mogą do tego dojść problemy ze znakami współczynników. W jednych okresach znaki są dodatnie, a w innych ujemne (bez jakiegokolwiek uzasadnienia). Wymienione mankamenty tej metody dostrzec można w wynikach badań wykonanych przez Leunga, Daouka i Chena [2000] i przytoczonych przez Gruszczyńskiego [2002, s. 196].

Znacznie lepszą propozycją jest zastosowanie do klasyfikacji gospodarstw metody stosowanej przez Gatelę [1999], polegającej na agregacji szeregu czasowego metodą średniej ruchomej. Wadą takiego rozwiązania jest jednakowy wpływ wartości wskaźników finansowych z przeszłości. Oczywiście wydaje się, że to, co miało miejsce rok temu, jest bliższe chwili obecnej niż wydarzenia sprzed dwóch czy trzech lat. Gatelę wyznaczał średnie dla szeregów, w których jednostką dyskretyzacji czasu jest dzień (dziennie obroty na giełdzie, wartości wybranych indeksów). Jeżeli jednostką jest rok, wskazane jest nadanie malejących wag informacji z przeszłości, tak aby w miarę zwiększania opóźnienia czasowego udział wybranej wartości malał (tak jak ma to miejsce w metodzie ważonej średniej ruchomej). Jest to uzasadnione wynikami badań polegających na prognozowaniu bankructwa na podstawie analizy finansowej z kilkuletnim wyprzedzeniem. W miarę powiększania opóźnienia skuteczność systemu malała. Widać to wyraźnie w modelu Altmana [1968, s. 604] czy w przedstawionych przez Rogowskiego [1999, s. 66-68] modelach opracowanych przez Beermanna, Frederiklusta i Gerharda.

Zastosowanie zwykłej lub ważonej średniej ruchomej w modelach dyskryminacyjnych służy agregacji informacji z przeszłości w celu uzyskania wartości pojedynczej (skalarnej). Zauważmy jednak, że zabieg ten można również interpretować w kategorii tworzenia prognozy wskaźnika na okres przyszły. Średnie ruchome, poza wygładzaniem szeregów czasowych, są stosowane również do generowania prognoz.

Przedstawione w artykule badania przeprowadzono na podstawie danych z 705 gospodarstw rolniczych prowadzących rachunkowość rolną pod kierunkiem Instytutu Ekonomiki Rolnictwa i Gospodarki Żywnościowej (IERiGŻ) w latach 1999-2002³. Dochód rolniczy wyznaczający przynależność do klas pochodził z roku 2002, wskaźniki natomiast stanowiące podstawę budowy modeli dyskryminacyjnych – z lat 1999-2001. Do obliczeń wykorzystany został program STATISTICA, wszystkie operacje przekształcania danych wykonano zaś w EXCELU.

³ Analizy ograniczono do lat 1999-2002, ponieważ IERiGŻ w roku 2002 wprowadził istotne zmiany w sposobie gromadzenia danych o gospodarstwach. Począwszy od 2003 r. zaś, nie są udostępniane dane indywidualne.

2. Wskaźniki finansowe i ich prognozy

W badaniach uwzględniono szeroki zestaw wskaźników finansowych dobrany na podstawie prac [Kulawik 1995; Zięta i in. 1994; Wyszowska 1996]. Zestaw ten objął początkowo 29 wskaźników, w tym wskaźniki płynności finansowej, rentowności, obrotowości, sprawności gospodarowania, wspomagania finansowego oraz stanowiące charakterystykę majątku trwałego. Dodatkowo uwzględniono powierzchnię użytków rolnych (wyrażoną w ha przeliczeniowych), wiek osoby prowadzącej gospodarstwo, poziom intensywności produkcji oraz poziom intensywności organizacji produkcji. Wskaźniki finansowe dla gospodarstw rolniczych wyznaczane są według nieco innych formuł niż dla przedsiębiorstw, co wynika z wyróżniania innych kategorii dochodów i produkcji. Poza przedstawioną wcześniej literaturą, również prace Kisielińskiej [2003, s. 51; 2004, s. 72] zawierają reguły ich obliczania.

Agregacja informacji zawartej w szeregu czasowym do pojedynczej wartości metodą ważonej średniej ruchomej wymaga założenia stałej wygładzania oraz wag dla poszczególnych lat. Dobór tych wartości można przeprowadzić niestety jedynie metodą prób i błędów.

W badaniach zastosowano dwie stałe wygładzania 2 i 3. Dla każdej z nich dobrano po trzy warianty wag. Dodatkowo uwzględniono przypadek stałej wygładzania równej 1, czyli prognozę opartą na wskaźnikach pochodzących tylko z jednego roku. W tab. 1 przedstawiono oznaczenia różnych wariantów wyznaczania prognoz cech. Oznaczenia te będą dalej utożsamiane z oznaczeniami modeli klasyfikacyjnych zbudowanych na ich podstawie. Prognozy budowano dla wszystkich wskaźników finansowych, poziomu intensywności produkcji oraz poziomu intensywności organizacji produkcji. Wiek osoby kierującej gospodarstwem i powierzchnię użytków rolnych przyjęto na poziomie roku poprzedzającego prognozę.

Tabela 1. Oznaczenia różnych wariantów wyznaczania prognoz wskaźników finansowych dla gospodarstw rolniczych metodą ważonej średniej ruchomej

Oznaczenie różnych wariantów wyznaczania prognoz wskaźników finansowych	Stała wygładzania	Wagi dla kolejnych lat		
		W_{i-1} **	W_{i-2}	W_{i-3}
W1L1	1	1	-	-
W1L2	2	0,8	0,2	-
W2L2	2	0,7	0,3	-
W3L2	2	0,6	0,4	-
W1L3	3	0,7	0,2	0,1
W2L3	3	0,6	0,3	0,1
W3L3	3	0,5	0,3	0,2

* Numer po literze W oznacza wariant zastosowanych wag, natomiast po literze L stałą wygładzania.

** Gdzie i jest rokiem, dla którego budowana jest prognoza wskaźnika.

Źródło: opracowanie własne.

Zastosowanie siedmiu przedstawionych w tab. 1 wariantów wyznaczania prognoz wskaźników finansowych daje siedem różnych zbiorów danych, wykorzystanych dalej do budowy modeli klasyfikacyjnych. Dla każdego zbioru wstępnie przeprowadzono analizę korelacji w celu wyeliminowania z dalszych obliczeń cech silnie skorelowanych ze sobą. Jako wartość graniczną przyjęto współczynnik korelacji o wartości powyżej 0,85. W dalszych badaniach uwzględniono tę cechę ze skorelowanej pary, która była silniej skorelowana ze zmienną określającą przynależność do klasy. Należy podkreślić, że w większości zbiorów silnie skorelowane ze sobą były zwykle te same cechy, choć wystąpiły drobne różnice.

Zrezygnowano z podziału zbioru na uczący i testowy. Celem niniejszej publikacji jest prezentacja pewnego sposobu agregacji danych w modelach klasyfikacyjnych, a nie gruntowna weryfikacja modeli. Podział zbioru na dwie części komplikowałby porównywanie wyników klasyfikacji. Gdy te same dane są używane do tworzenia modeli dyskryminacyjnych oraz ich oceny, stosowane testy są obciążone. Obciążenie to maleje jednak ze wzrostem liczebności próby [Frank i in. 1965, s. 253]. Ponieważ analizowany zbiór obiektów jest dosyć pokaźny, można przypuszczać, że obciążenie nie jest zbyt duże.

Opierając się na zestawach wskaźników pozostałych po wstępnej eliminacji, zbudowano dla wszystkich wariantów klasyfikacyjne modele funkcyjne i sieciowe. W przypadku analizy dyskryminacyjnej stosowano analizę krokową. Sieci neuronowe natomiast same dobierają najlepsze kombinacje zmiennych wejściowych. Ostatecznie w najlepszym modelu funkcyjnym i sieciowym uwzględnionych zostało 20 cech, których prezentacja zostanie pominięta ze względu na ograniczoną objętość artykułu.

3. Wyniki klasyfikacji

W tab. 2 przedstawione zostały wyniki klasyfikacji uzyskane przy użyciu liniowych funkcji dyskryminacyjnych wyznaczonych na podstawie prognoz wskaźników finansowych uzyskanych metodą ważonej średniej ruchomej dla różnych wariantów wag i zastosowanych stałych wygładzania.

Tabela 2. Wyniki klasyfikacji gospodarstw rolniczych przeprowadzonej za pomocą liniowych funkcji dyskryminacyjnych wyznaczonych na podstawie prognoz wskaźników finansowych uzyskanych metodą ważonej średniej ruchomej dla różnych wariantów wag i zastosowanych stałych wygładzania

Oznaczenie różnych wariantów wyznaczania prognoz wskaźników finansowych	Procent poprawnych klasyfikacji w klasie I	Procent poprawnych klasyfikacji w klasie II	Procent poprawnych klasyfikacji
W1L1	79,55	76,13	77,84
W1L2	80,11	77,27	78,69
W2L2	80,40	77,27	78,84
W3L2	80,68	76,70	78,69
W1L3	79,55	77,55	78,55
W2L3	79,83	77,84	78,83
W3L3	79,83	78,13	78,98

Źródło: opracowanie własne.

Procent poprawnych klasyfikacji dla utworzonych modeli jest zbliżony. Najgorsze wyniki uzyskano, opierając prognozę na wskaźnikach tylko z jednego roku (wariant danych oznaczony jako W1L1), najlepsze natomiast dla stałej wygładzania równej 3 i kombinacji wag 0,5, 0,3 i 0,2 (W3L3). Procent poprawnych klasyfikacji uzyskany dla wariantu W1L1 wyniósł 77,84%, dla W3L3 zaś był równy 78,98%. Różnica nie jest duża – nieco ponad 1%, wskazuje jednak, że uwzględnienie danych z większej liczby lat pozwala na poprawienie prognozy. Wszystkie klasyfikacje opierające się na szeregach czasowych cechuje wyższa jakość niż zbudowane dla danych jedynie z jednego roku.

Wszystkie modele wykazują niewielką asymetrię w rozpoznawaniu klas. Lepiej rozpoznawana jest klasa I niż II. Należy zauważyć, że model W3L3 cechuje znacznie mniejsza asymetria niż W1L1.

Łączny procent poprawnych klasyfikacji na poziomie 78,98% świadczy o średniej jakości modelu. Podniesienie jej umożliwiłoby prawdopodobnie podzielenie badanej zbiorowości na grupy obejmujące podobne gospodarstwa (np. o podobnych typach produkcji), uzupełnienie opisu o dodatkowe cechy (np. region, w jakim usytuowane jest gospodarstwo, liczba osób w nim pracujących, wykształcenie kierownika).

Dla siedmiu zbiorów danych różniących się sposobem wyznaczenia prognozy wskaźników (przedstawionych w tab. 1) skonstruowano sieci neuronowe. Do budowy modeli wykorzystano narzędzie zwane automatycznym projektantem, które testuje wiele sieci, dobierając ich strukturę i stopień złożoności. Obliczenia dla każdego zbioru danych powtarzane były wielokrotnie i spośród wyznaczonych sieci wybierano najlepszą. Wyznaczone tak modele dyskryminacyjne zostały następnie wykorzystane do klasyfikacji badanego zbioru gospodarstw. Wyniki tej klasyfikacji zawiera tab. 3.

Podobnie jak w przypadku modeli funkcyjnych, jakość modeli sieciowych jest również zbliżona. Najmniejszy procent poprawnych klasyfikacji otrzymano dla modelu W1L1 (dane z jednego roku), największy natomiast dla modelu W3L2 (stała wygładzania równa 2 z wagami 0,6 i 0,4).

Tabela 3. Wyniki klasyfikacji gospodarstw rolniczych przeprowadzonej za pomocą sieci neuronowych opracowanych na podstawie prognoz wskaźników finansowych uzyskanych metodą ważonej średniej ruchomej dla różnych wariantów wag i zastosowanych stałych wygładzania

Oznaczenie różnych wariantów wyznaczenia prognoz wskaźników finansowych	Procent poprawnych klasyfikacji w klasie I	Procent poprawnych klasyfikacji w klasie II	Procent poprawnych klasyfikacji
W1L1	81,25	78,13	79,69
W1L2	83,24	80,97	82,10
W2L2	79,28	80,97	80,11
W3L2	84,38	80,40	82,39
W1L3	79,55	83,81	81,68
W2L3	80,97	83,52	82,24
W3L3	78,98	81,25	80,11

Źródło: badania własne.

Model oparty na danych z jednego roku (W1L1) był w przypadku sieci również modelem najgorszym. Różnica w liczbie poprawnie zaklasyfikowanych gospodarstw między najlepszym i najgorszym modelem sieciowym jest równa 2,6%. Jest to ponad dwa razy więcej niż w modelach funkcyjnych.

Wszystkie modele sieciowe okazały się lepsze niż modele funkcyjne (dla kolejnych modeli poprawa odsetka poprawnie zaklasyfikowanych gospodarstw była równa kolejno 1,9, 3,4, 1,3, 3,7, 3,1, 3,4 oraz 1,1%). Podobnie jak modele funkcyjne, cechuje je wprawdzie niewielka asymetria w rozpoznawaniu klas, choć w tym wypadku jedynie trzy modele rozpoznają lepiej klasę I, natomiast pozostałe klasę II.

4. Wnioski

Przedstawione powyżej wyniki obliczeń dokonanych dla 704 gospodarstw rolniczych prowadzących rachunkowość pod kierunkiem IERiGŻ pozwalają na wyciągnięcie następujących wniosków.

1. Wszystkie modele, zarówno funkcyjne, jak i sieciowe, zbudowane na podstawie danych w postaci szeregów czasowych były lepsze od modeli opartych na danych wyłącznie z jednego roku – poprzedzającego prognozę. Różnice nie były jednak duże. W przypadku najlepszego modelu funkcyjnego poprawa klasyfikacji była rzędu 1%, dla sieciowego zaś 2,6%.

2. Wszystkie modele sieciowe były lepsze od wszystkich modeli funkcyjnych. Różnice dla poszczególnych wariantów wyznaczania prognoz wahały się w zakresie od nieco ponad 1% do prawie 4%.

3. Najlepszy model funkcyjny zbudowany w celu prognozowania sytuacji finansowej gospodarstw rolniczych prawidłowo zaklasyfikowało 78,98% gospodarstw (79,83% należących do klasy I i 78,13% z klasy II). Najlepszy model sieciowy poprawnie rozpoznało 82,39% gospodarstw (84,38% należących do klasy I i 80,40% z klasy II).

4. Obydwa modele – funkcyjny i sieciowy – prawidłowo zaklasyfikowało 76% gospodarstw, błędnie zaś 14%. Pozostałe 10% zostało prawidłowo rozpoznane przez jeden, a błędnie przez drugi model. Oznacza to, że więcej gospodarstw zaklasyfikowano błędnie obydwoma metodami niż tylko jedną. Można więc przypuszczać, że błędne rozpoznanie wynika bardziej z cech obiektów niż niedostatków metod. Celowe jest więc uzupełnienie zestawu danych, stanowiącego podstawę budowy modeli, o pewne dodatkowe informacje.

Literatura

Altman E.I. (1968), *Financial Ratios, Discriminant Analysis and the Prediction of Corporate Bankruptcy*, „Journal of Finance”, vol. 23 nr 4, s. 589-609.

- Frank R.E., Massy W.F., Morrison D.G. (1965), *Bias in Multiple Discriminant Analysis*, „Journal of Marketing Research”, vol. II, s. 250-258.
- Gately E. (1999), *Sieci neuronowe. Prognozowanie finansowe i projektowanie systemów transakcyjnych*, WIG-Press, Warszawa.
- Gruszczyński M. (2002), *Modele i prognozy zmiennych jakościowych w finansach i bankowości*, Szkoła Główna Handlowa, Warszawa.
- Kisielińska J. (2003), *Klasyfikacja gospodarstw rolniczych siecią neuronową Kohonena w oparciu o wybrane wskaźniki finansowe*, „Zagadnienia Ekonomiki Rolnej” nr 2, s. 49-64.
- Kisielińska J. (2004), *Wykorzystanie liniowej funkcji dyskryminacyjnej oraz sieci neuronowych do prognozowania kondycji finansowej gospodarstw rolniczych*, „Przegląd Statystyczny” nr 2, s. 69-83.
- Krzyśko M. (1990), *Analiza dyskryminacyjna*, Wydawnictwo Naukowo-Techniczne, Warszawa.
- Kulawik J. (1995), *Wskaźniki finansowe i ich systemy w zarządzaniu gospodarstwami rolniczymi*, „Studia i monografie IERiGŻ”, z. 72.
- Leung M.T., Daouk H., Chen A.S. (2000), *Forecasting Stock Indices: a Comparison of Classification and Level Estimation Models*, „International Journal of Forecasting”, vol. 16, nr 2, s. 173-190.
- Rogowski W. (1999), *Możliwość wczesnego rozpoznawania symptomów zagrożenia zdolności płatniczej przedsiębiorstwa*, „Bank i Kredyt” nr 6, s. 56-72.
- Strahl D., Markowska M. (2004), *Klasyfikacja obiektów w ujęciu dynamicznym*, [w:] Taksonomia 11, red. K. Jajuga, M. Walesiak, Prace Naukowe Akademii Ekonomicznej we Wrocławiu nr 1022, AE, Wrocław, s. 556-565.
- Tarczyński W. (1996), *Analiza dyskryminacyjna na giełdzie papierów wartościowych*, „Przegląd Statystyczny” nr 1-2, s. 49-66.
- Tarczyński W. (2001), *Rynki kapitałowe: metody ilościowe*, vol. 1, Agencja Wydawnicza Placet, Warszawa.
- Wyszowska Z. (1996), *Wybrane elementy wskaźnikowej analizy finansowej na przykładzie przedsiębiorstw rolniczych*, ART, Bydgoszcz.
- Zawadzki J., Babis H. (1996), *Próba zastosowania analizy dyskryminacyjnej do badania kondycji finansowej przedsiębiorstw*, [w:] Taksonomia 3, s. 119-127.
- Ziętara W., Kosiorek M., Tchorzewska E., Kondraszuk T. (1994), *Rachunek ekonomiczny i analiza finansowa w przedsiębiorstwie rolniczym*, Centrum Doradztwa i Edukacji w Brwinowie, Warszawa.

CLASSIFICATION OF FARMS BASED ON TIME SERIES DATA

Summary

In the paper the construction of classification models for farms based on time series data is proposed. In the suggested approach data concerning farms were aggregated using the weighted moving average method. The proposed method was tested on a sample of 704 farm households for various smoothing constants and various combinations of weights. Two types of models were built in the form of linear discriminant functions and obtained by neural networks which enabled a comparison of linear and nonlinear approach to the classification problem. It turned out that all function and network models built based on data representing several preceding periods were better than models built based on data representing one year. Besides all network models were better than their function counterparts but the differences were slight.