Statystyczne modele wczesnego ostrzegania – Metody formalne

Krzysztof Jajuga[1]

1. Modele wczesnego ostrzegania – systematyzacja

Jednym z podstawowych warunków podejmowania skutecznych decyzji w gospodarce jest korzystanie z dobrego modelu decyzyjnego. Modelami decyzyjnymi są również tzw. modele wczesnego ostrzegania. Jest to potoczna nazwa, ale w zasadzie została zaakceptowana przez środowisko naukowe i praktyków. Modele wczesnego ostrzegania mają na celu identyfikację różnego rodzaju zagrożeń czy też zwiększonego ryzyka. Jednym z przykładów modeli wczesnego ostrzegania są modele ryzyka upadłości przedsiębiorstwa. Wskazują one na zagrożenie przedsiębiorstwa upadłością. Innym przykładem są modele zagrożenia bezrobociem. Wskazują one na zagrożenie zwiększenia bezrobocia w określonym regionie.

Ogólna zasada konstrukcji modeli wczesnego ostrzegania jest bardzo podobna. W uproszczony sposób model wczesnego ostrzegania można przedstawić za pomocą funkcji:

$$(1) \quad Y = f(X_1, X_2, ..., X_m)$$

gdzie:

Y – zmienna charakteryzująca analizowane zjawisko zagrożone ryzykiem (np. upadłość), przy tym dla ustalenia uwagi przyjmijmy, że im wyższe wartości tej zmiennej, tym większe zagrożenie;

$X_1, X_2, ..., X_m$ – czynniki wpływające na analizowane zjawisko zagrożone ryzykiem (czasem nazywane zmiennymi objaśniającymi);

f – funkcja, często przyjmuje się, że jest to funkcja liniowa.

Procedura tworzenia modeli wczesnego ostrzegania może być przedstawiona w następujących etapach:

1. Zdefiniowanie zmiennej *Y* charakteryzującej analizowane zjawisko.
2. Identyfikacja zmiennych objaśniających.
3. Wybór metody wyznaczania funkcji *f*.
4. Określenie kryterium ostrzegania przez model.

Pierwsze dwa etapy zależą od wiedzy merytorycznej o analizowanym zjawisku. Z kolei trzeci etap ma charakter formalny. W tym opracowaniu zajmujemy się właśnie trzecim etapem.

Przedstawione są podstawowe metody formalne, statystyczne i ekonometryczne, które są wykorzystywane przy tworzeniu modeli wczesnego ostrzegania. Dodajmy jeszcze, iż w czwartym etapie określa się wartość progową, której przekroczenie generuje sygnał ostrzegawczy przez model.

W dalszych rozważaniach dla uproszczenia utożsamiać będziemy model wczesnego ostrzegania z metodą wyznaczania tego modelu.

W celu systematyzacji podstawowych modeli wczesnego ostrzegania dokonaliśmy analizy różnych metod tworzenia tych modeli. Naszym zdaniem, modele te można podzielić na dwie duże grupy:

- modele wywodzące się z analizy dyskryminacyjnej;
- modele wywodzące się z analizy regresji.

Główna różnica między tymi dwoma grupami, oprócz tego, że są w nich stosowane inne techniki statystyczne, tkwi w sposobie traktowania zmiennej *Y*. W metodach wywodzących się z analizy dyskryminacyjnej wartości tej zmiennej powstają w wyniku zastosowania konkretnej techniki – wartości te otrzymuje się w momencie wyznaczenia modelu wczesnego ostrzegania. Z kolei w metodach wywodzących się z analizy regresji wartości tej zmiennej są dane przed wyznaczeniem modelu – jego parametry są wyznaczane właśnie na podstawie wartości zmiennej Y oraz czynników na nią wpływających. Szczegółowe wyjaśnienia zawarte są przy opisie każdej grupy metod.

W dalszej części przedstawimy opis reprezentantów dwóch wyodrębnionych grup metod. Są to: klasyczna liniowa analiza dyskryminacyjna oraz modele regresji, w których zmienna objaśniana jest jakościowa.

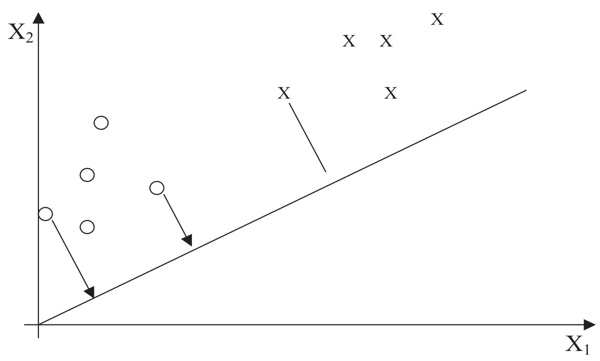
Analiza dyskryminacyjna

Metoda analizy dyskryminacyjnej jest to jedna z najprostszych metod stosowanych w zagadnieniu dyskryminacji. Przedstawimy teraz to zagadnienie w najprostszej wersji, w której mamy do czynienia z dyskryminacją na dwie klasy. Odnosząc to do zagadnienia modeli wczesnego ostrzegania, można powiedzieć, że jedna z tych klas odpowiada zagrożeniu ryzykiem, a druga brakowi tego zagrożenia.

Metoda analizy dyskryminacyjnej polega na wyznaczeniu funkcji, której wartości umożliwiają przydzielenie rozpatrywanego obiektu do jednej z dwóch klas: zagrożonej ryzykiem lub niezagrożonej ryzykiem. Przy tym wyznaczenie funkcji ma miejsce na podstawie danych dotyczących obiektów, o których wiadomo, do której z dwóch klas należą. Innymi słowy, część obiektów w przeszłości była zagrożona ryzykiem (np. przedsiębiorstwa, które upadły), zaś część nie. Wynika z tego, że funkcja dyskryminacyjna powinna „dobrze separować" obiekty należące do dwóch różnych klas. Z kolei wyznaczona funkcja służy do podjęcia decyzji o zakwalifikowaniu do dwóch klas rozpatrywanych obiektów.

Idea analizy dyskryminacyjnej przedstawiona jest na rysunku 1.

[1] Akademia Ekonomiczna we Wrocławiu, Katedra Inwestycji Finansowych i Zarządzania Ryzykiem.



Rys 1. Ilustracja analizy dyskryminacyjnej.

Na rysunku tym w dwuwymiarowym układzie współrzędnych przedstawione są dwa zbiory punktów. Są to pewne obiekty (np. przedsiębiorstwa, regiony), analizowane ze względu na dwa czynniki. Punkty oznaczone krzyżykami odpowiadają obiektom zagrożonym ryzykiem. Z kolei punkty oznaczone kółkami odpowiadają obiektom niezagrożonym ryzykiem. Klasyczna analiza dyskryminacyjna polega na wyznaczeniu funkcji liniowej transformującej obserwacje wielowymiarowe (w przedstawionej ilustracji są to obserwacje dwuwymiarowe) w punkty leżące na prostej. Odpowiednia linia jest zaznaczona na rysunku 1. Na rysunku przedstawiony jest również punkt C, w odniesieniu do którego ma być podjęta decyzja, czyli wystąpi sygnał ostrzegawczy. Z rysunku wynika, iż taki sygnał powinien zostać wygenerowany, gdyż punkt C znalazł się po „zagrożonej stronie linii prostej". Należy dodać, iż w praktyce często zdarza się tak, iż zbiór punktów nie jest liniowo separowalny, tzn. nie da się jednoznacznie rozdzielić obu zbiorów punktów. Wtedy linia jest wyznaczona w taki sposób, aby jak najmniej było błędnych decyzji o przydzieleniu obiektów do klas.

Przedstawimy teraz w sposób formalny metodę analizy dyskryminacyjnej.

Załóżmy, że dysponujemy zbiorem, liczącym *n* obserwacji (dotyczących n rozpatrywanych obiektów). Przy tym każda obserwacja jest *m*-wymiarowa, to znaczy zawiera wartości m czynników wpływających na analizowane zjawisko. Przez analogię oznaczymy je jako $X_1, X_2, \ldots, X_m$

W ten sposób dysponujemy tzw. macierzą obserwacji, o wymiarach *n* x *m*. Jej element, jest to wartość zmiennej o numerze j dla obiektu o numerze *i*. Przy tym wiadomo również, że *n1* obserwacji należy do pierwszej klasy (jest to klasa zagrożona ryzykiem), zaś *n2* obserwacji należy do drugiej klasy.

Zadaniem analizy dyskryminacyjnej jest wyznaczenie liniowej funkcji dyskryminacyjnej:

$$(2)\quad Y = a_1 X_1 + a_2 X_2 + \ldots + a_m X_m$$

Oznacza to, iż w odniesieniu do każdej obserwacji otrzymujemy wartości funkcji dyskryminacyjnej:

$$(3)\quad y_i = a_1 x_{i1} + a_2 x_{i2} + \ldots + a_m x_{im} \qquad i = 1, \ldots, n$$

Idea wyznaczenia tej funkcji jest prosta (wynika to z omówionej już ilustracji graficznej): obserwacje należące do dwóch różnych klas powinny mieć zróżnicowane wartości. W przypadku dwóch klas oznacza to również, że czynniki wpływające na analizowane zjawisko (zmienne objaśniające) powinny mieć jednoznaczną interpretację, tzn. im wyższe wartości, tym większe zagrożenie ryzykiem.

W klasycznym zadaniu analizy dyskryminacyjnej funkcja dyskryminacyjna powstaje jako rozwiązanie zagadnienia maksymalizacji następującej funkcji kryterium:

$$(4)\quad S = \frac{n1(\bar{y}_1 - \bar{y})^2 + n2(\bar{y}_2 - \bar{y})^2}{\sum_{i \in C1} (y_i - \bar{y}_1)^2 + \sum_{i \in C2} (y_i - \bar{y}_2)^2}$$

gdzie:
C1, C2 – *numery klas.*

Wyrażenie (4) jest to iloraz, w którym licznik jest to wariancja międzyklasowa, zaś mianownik jest to wariancja wewnątrzklasowa, wyznaczone na podstawie transformowanych obserwacji otrzymanych według wzoru (3). Ponieważ, jak wskazywaliśmy, obserwacje należące do dwóch klas powinny mieć zróżnicowane wartości, zatem relacja wariancji międzyklasowej do wariancji wewnątrzklasowej powinna być jak największa. Uzasadnia to konieczność maksymalizacji wyrażenia (4) w celu wyznaczenia współczynników funkcji dyskryminacyjnej

Można dowieść, że optymalne rozwiązanie powstaje poprzez wzięcie wektorów własnych macierzy, będącej iloczynem macierzy odwrotnej do macierzy rozrzutu wewnątrzklasowego i macierzy rozrzutu międzyklasowego. Obie te macierze są odpowiednio skonstruowanymi macierzami kowariancji. Metoda analizy dyskryminacyjnej jest to podstawowa metoda statystycznej analizy wielowymiarowej, która jest oprogramowana w standardowych pakietach statystycznych.

W praktyce stosowania analizy dyskryminacyjnej do rozstrzygnięcia pozostaje jeszcze kwestia wartości progowej funkcji dyskryminacyjnej. Pozwala ona na zakwalifikowanie rozpatrywanego obiektu do jednej z dwóch klas: zagrożonej ryzykiem (wtedy generowany jest sygnał ostrzegawczy; lub niezagrożonej ryzykiem).

2. Modele regresji z jakościową zmienną objaśnianą

Modele regresji z jakościową zmienną objaśnianą jest to druga grupa modeli, które mogą być stosowane w charakterze modeli wczesnego ostrzegania. Obecnie przedstawimy je w sposób syntetyczny. Nasze rozważania rozpoczniemy od przedstawienia ogólnej idei tych modeli. Do tego celu zastosujemy najprostszy (ale nie najlepszy) model zaliczany do tej grupy, tzw. **liniowy model prawdopodobieństwa.**

Rozważmy zwykły model regresji liniowej:

(5) $$Y = \beta_1 X_1 + \beta_2 X_2 + ... + \beta_m X_m + \beta_0 + \varepsilon$$

gdzie:

Y – zmienna objaśniana (jest to zmienna losowa);

X_j – zmienna objaśniająca o numerze j; zakładamy tutaj, że wartości zmiennych objaśniających to wielkości ustalone (nielosowe);

β_j – parametr modelu (j=0,1,2,…,m);

ε – składnik losowy modelu (jest to zmienna losowa, której wartość oczekiwana wynosi 0).

W liniowym modelu prawdopodobieństwa przyjmuje się, że zmienna objaśniana może przyjąć dwie wartości z określonymi prawdopodobieństwami:

(6) $$Y = \begin{Bmatrix} 1, & P(1) = p \\ 0, & P(0) = 1 - p \end{Bmatrix}$$

Przy tym przyjmujemy, że wartość zmiennej objaśnianej równa 1 oznacza wystąpienie pewnego zdarzenia. W przypadku modeli wczesnego ostrzegania oznacza to wystąpienie sytuacji, stanowiącej zagrożenie ryzykiem, czyli generującej sygnał ostrzegawczy.
Jak wiadomo, przy estymacji modelu regresji liniowej wyznacza się wartość oczekiwaną zmiennej objaśnianej. Można dowieść, że w przypadku liniowego modelu prawdopodobieństwa zachodzi właściwość:

(7) $$E(Y) = p$$

Oznacza to, że wartość oczekiwana zmiennej objaśnianej równa się prawdopodobieństwu zajścia zdarzenia, czyli zagrożenia ryzykiem. Biorąc pod uwagę model (5) i zastępując parametry modelu ich ocenami, otrzymujemy następującą postać liniowego modelu prawdopodobieństwa:

(8) $$p = b_1 X_1 + b_2 X_2 + ... + b_m X_m + b_0$$

Wynika z tego, że liniowy model prawdopodobieństwa przedstawia prawdopodobieństwo zdarzenia ryzykownego jako liniową funkcję czynników wpływających na to zdarzenie. Jest to sensowna interpretacja, która powoduje, iż idea, stojąca u podstaw liniowego modelu prawdopodobieństwa, jest dość atrakcyjna.

Niestety, model ten ma jeden niedostatek, który powoduje, iż częściej stosowane są inne modele, które przedstawimy w dalszej części. Niedostatek ten wynika z faktu, iż po podstawieniu do wzoru (8) możemy otrzymać wartości prawdopodobieństwa, które znajdują się poza przedziałem [0;1], co w znacznym stopniu osłabia zalety interpretacyjne modelu, czyniąc ten model mniej użytecznym.

Stwarza to konieczność poszukiwania innych modeli, będących w pewnym sensie modyfikacjami liniowego modelu prawdopodobieństwa. Zaliczamy do nich:

- **model probitowy;**
- **model logitowy.**

W celu ich przedstawienia zwróćmy uwagę, że prawdopodobieństwo formalnie może być interpretowane jako wartość funkcji dystrybuanty.

W związku z tym pojawił się pomysł, aby przy określaniu prawdopodobieństwa wyznaczyć dystrybuantę rozkładu. Ponieważ dystrybuanta przyjmuje wartości z przedziału [0;1], jest zatem gwarancja, iż wyznaczone prawdopodobieństwa znajdą się w przedziale [0;1]. Zauważmy, iż w liniowym modelu prawdopodobieństwa tej gwarancji nie ma z uwagi, iż prawdopodobieństwo jest określone jako liniowa funkcja zmiennych objaśniających, a nie jako funkcja dystrybuanty.
Wymienione powyżej dwa modele, probitowy i logitowy, zakładają określoną funkcję dystrybuanty, mianowicie:

- model probitowy – funkcja dystrybuanty standaryzowanego rozkładu normalnego, dana wzorem:

(9) $$p = F(x) = \Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-0,5t^2} dt$$

- model logitowy – funkcja dystrybuanty rozkładu logistycznego, dana wzorem:

(10) $$p = F(x) = L(x) = \frac{1}{1 + e^{-x}}$$

Dla porównania przedstawimy również funkcję dystrybuanty rozkładu jednostajnego na przedziale [0;1] – dana jest wzorem:

(11) $$p = F(x) = U(x) = \begin{cases} 0, & x < 0 \\ x, & 0 \le x \le 1 \\ 1, & x > 1 \end{cases}$$

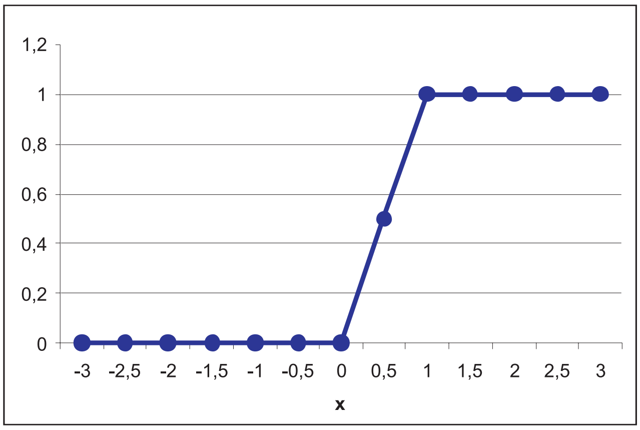
Tablica 1 przedstawia porównanie wartości dystrybuanty dla trzech rozkładów, będących podstawą trzech modeli: liniowego prawdopodobieństwa, probitowego i logitowego.

Tablica 1. Wartości dystrybuanty dla trzech rozkładów.

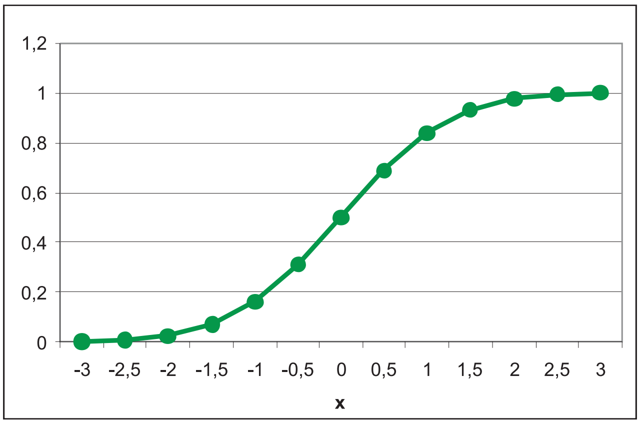
x	Wartość dla rozkładu jednostajnego	Wartość dla standaryzowanego rozkładu normalnego	Wartość dla rozkładu logistycznego
-3,0	0	0,001	0,047
-2,5	0	0,006	0,076
-2,0	0	0,023	0,119
-1,5	0	0,067	0,182
-1,0	0	0,159	0,269
-0,5	0	0,309	0,378
0	0	0,5	0,5
0,5	0,5	0,691	0,622
1,0	1	0,841	0,731
1,5	1	0,933	0,818
2,0	1	0,977	0,881
2,5	1	0,994	0,924
3,0	1	0,999	0,953

Źródło: obliczenia własne

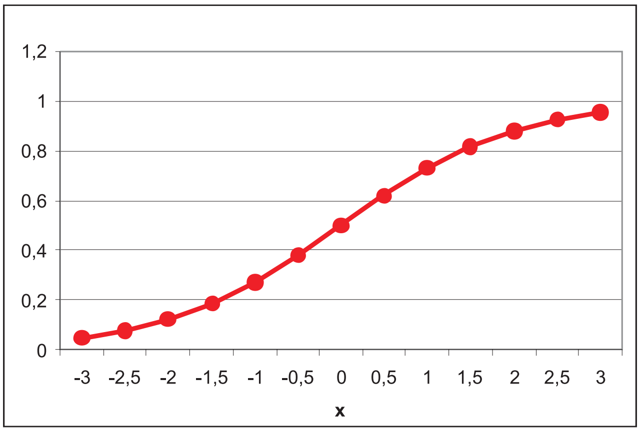
Tablica 1 pokazuje dość wyraźnie różnice między modelem probitowym i logitowym z jednej strony, a modelem liniowego prawdopodobieństwa z drugiej strony. Zilustrowane to jest również na rysunkach 2-4.



Rys 2. Wartość dystrybuanty dla rozkładu jednostajnego



Rys 3. Wartość dystrybuanty dla standaryzowanego rozkładu normalnego



Rys 4. Wartość dystrybuanty dla rozkładu logistycznego

Należy jeszcze raz podkreślić, iż model liniowego prawdopodobieństwa – w przeciwieństwie do modelu logitowego i probitowego – nie jest modelem, w którym prawdopodobieństwo jest określone jako funkcja dystrybuanty określona na wartości liniowej funkcji zmiennych objaśniających, lecz jako sama wartość liniowej funkcji zmiennych objaśniających. To właśnie implikuje możliwość otrzymania wartości spoza przedziału [0;1]. Jednak można oczywiście zmodyfikować model liniowego prawdopodobieństwa poprzez wzięcie dystrybuanty rozkładu jednostajnego. W praktyce tak się czasem czyni, gdyż w przypadku otrzymania wartości wyższej od 1 – przyjmuje się 1, zaś w przypadku otrzymania wartości niższej od 0 – przyjmuje się 0. Jednak klasyczny model liniowego prawdopodobieństwa takiej możliwości nie uwzględnia.

Z praktycznego punktu widzenia istotne jest oszacowanie modeli regresji z jakościową zmienną objaśnianą. Są to modele regresji, więc można tu stosować klasyczne metody estymacji. Jedną z nich jest metoda najmniejszych kwadratów. Jednak z uwagi na zjawisko heteroskedastyczności składnika losowego powinna być tu zastosowana ważona metoda najmniejszych kwadratów.

Oszacowany model może być zastosowany w odniesieniu do każdego analizowanego obiektu w celu wyznaczenia oszacowania prawdopodobieństwa zagrożenia ryzykiem. Oczywiście użytkownik modelu (decydent) powinien przyjąć wartość prawdopodobieństwa progowego, którego przekroczenie powoduje generowanie sygnału ostrzegawczego.

Literatura:

1. Jajuga K. (1990), Statystyczna teoria rozpoznawania obrazów, Wydawnictwo Naukowe PWN, Warszawa.
2. Bartosiewicz S. (red.) (1990), Estymacja modeli ekonometrycznych, PWE, Warszawa.