

Ljiljanka Kvesić<sup>1</sup><sup>1</sup>Sveučilište u Mostaru, Fakultet prirodoslovno-matematičkih i odgojnih znanosti, Matice Hrvatske b.b., 88000 Mostar, Bosna i HercegovinaUDK 519.2:336.7  
Pregledni članak

# STATISTIČKE METODE U UPRAVLJANJU KREDITNIM RIZIKOM

## SAŽETAK

Uspješne banke svoje poslovanje temelje na principima likvidnosti, profitabilnosti i sigurnosti, stoga je korektna procjena sposobnosti zajmotražitelja da izvrše ugovorne obveze od krucijalne važnosti za funkcioniranje banke. Kao podrška kreditnim analitičarima u procjeni zajmotražitelja posljednjih je nekoliko desetljeća razvijeno više modela kreditnog scoringa. U ovom su radu prezentirane tri statističke metode koje se s tom svrhom koriste u području upravljanja kreditnim rizikom: logistička regresija, diskriminativna analiza i analiza preživljenja. Njihova implementacija u bankarskom sektoru u znatnoj je mjeri potaknuta razvojem i primjenom informacijskih i komunikacijskih tehnologija. Ovim se radom nastoji ukazati na najvažnije teorijske aspekte navedenih metoda, ali također aktualizirati potrebu razvoja i primjene kreditnih scoring modela u hrvatskoj bankarskoj praksi.

**Ključne riječi:** upravljanje kreditnim rizikom, modeli kreditnog scoringa, logistička regresija, diskriminativna analiza, analiza preživljenja

## 1. Uvod

Kreditni rizik temeljno je obilježje bankarskog poslovanja. Prema Jorionu [11] kreditni rizik proizlazi iz činjenice da ugovorna strana nije voljna ili nije u mogućnosti izvršiti svoje obveze. Fabozzi i Peterson [3] izdvajaju nekoliko situacija u kojima dolazi do neispunjenja obveza kao što su npr. nevraćanje kamata i glavnice u definiranom roku ili neizvršavanje bilo koje druge stavke iz ugovora. Različiti razlozi mogu dovesti do neispunjenja ugovorne obveze, no ono uglavnom nastupa kao posljedica financijskih problema dužnika.

Budući da uspješnost poslovanja banke neposredno ovisi o sposobnosti predviđanja i kvantificiranja rizika, za nju je vrlo važno ispravno procijeniti kreditnu sposobnost zajmotražitelja. Prije pojave modela kreditnog scoringa i razvoja odgovarajućih tehnologija potrebnih za njihovu implementaciju,

takva se procjena temeljila isključivo na iskustvu i subjektivnom osjećaju bankarskog analitičara. Rezultati takvih prosudbi često nisu bili zadovoljavajući. Posljedica je toga bila da su krediti odobravani klijentima koji su kasnije zapali u poteškoće s otplatom zajma, dok su odbijani zahtjevi onih koji su bili u mogućnosti izvršavati svoje obveze.

Modeli kreditnoga scoringa razvijeni su svrhom utvrđivanja vjerojatnosti da klijent, s određenim karakteristikama, neće biti u mogućnost izvršiti svoje obveze. S vremenom su modeli kreditnoga scoringa postali snažna podrška bankama u upravljanju kreditnim rizikom. Važnost koju takvi modeli danas imaju potaknula je Thomasa, Edelmana i Crooka [19] da zakluče kako je kreditni scoring jedna od najuspješnijih primjena statistike i operacijskih istraživanja u financijama i bankarstvu.

Prema Hand i Henley [6] kreditni scoring termin

je koji se koristi za opisivanje formalnih statističkih metoda klasifikacije aplikirane za kredit u dobru i lošu rizičnu skupinu. Dakle kreditni se scoring svodi na klasifikacijski problem, gdje inpute predstavljaju podaci koji se odnose na podnositelja zahtjeva za kredit. Pri kreiranju modela kreditnoga scoringa potrebno je biti svjestan činjenice da se neće moći točno klasificirati svaki analizirani klijent. Savršena je klasifikacija nemoguća, budući da i dobri i loši klijenti ponekad imaju iste ili slične karakteristike. Zbog toga se pri modeliranju nastoji determinirati pravilo koje rezultira najmanjim mogućim brojem netočnih klasifikacija.

U ovom su radu teorijski prezentirane tri statističke metode koje su našle primjenu u području upravljanja kreditnim rizikom: logistička regresija, diskriminativna analiza i analiza preživljenja.

## 2. Logistička regresija

Logistička regresija (Christensen [2], Harrell [7], Tabachnick, Fidell [18], Vittinghoff, Glidden, Shiboski, McCulloch [20]) pogodna je za modeliranje binarnih ishoda, a dobiveni scor jednostavno se može konvertirati u procjene vjerojatnosti. Zbog toga se ova tehnika široko koristi u izgradnji scoring modela. Osim toga, ona ne pretpostavlja linearan odnos između zavisne i nezavisnih varijabli, a niti da je zavisna varijabla ili pogreška relacije normalno distribuirana.

Zavisna varijabla u modelu binarne logističke regresije mora biti dihotomna, dok nezavisne varijable mogu biti numeričke i kategorijalne. U modelu logističke regresije pretpostavlja se da su nezavisne varijable, pomnožene s njihovim pridruženim koeficijentima, linearno povezane ne samo sa  $Y$ , već i s prirodnim logaritmom šansi da će se  $Y$  dogoditi (Mays [15]).

Za specifikaciju vjerojatnosti može biti korištena logistička funkcija, koja je definirana kao:

$$f(z) = \frac{e^z}{1 + e^z}$$

U modelu kreditnoga scoringa varijabla  $z$  predstavlja izloženost skupu faktora rizika, dok  $f(z)$  predstavlja vjerojatnost određenog rezultata s obzirom na taj skup. Dakle varijabla  $z$  je mjera ukupnog doprinosa svih faktora rizika koji su korišteni u modelu.

Varijable  $z$  obično se definira na sljedeći način:

$$z = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_j x_j + \dots + \beta_k x_k$$

Slobodan član  $\alpha$  predstavlja vrijednost varijable  $z$  u slučaju da su sve nezavisne varijable jednake 0, odnosno da nema rizičnih faktora. Svaki od koeficijentata regresije  $\beta_i$  iskazuje doprinos faktora rizika na koji se odnosi. Njegova pozitivna vrijednost znači da faktor rizika povećava vjerojatnost ishoda, a negativna da ga smanjuje.

Neka je  $Y=1$  ako klijent ne ispunjava obveze, a 0 inače. U logističkoj regresiji vjerojatnost  $\pi$  je modelirana kao

$$\pi = P(Y = 1 | X_1 = x_1, \dots, X_k = x_k)$$

Dakle, vrijedi:

$$\pi = \frac{e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}$$

Dobivena funkcija logističke regresije je nelinearna u parametrima, a može se linearizirati pomoću logit transformacije. Tada je

$$1 - \pi = \frac{1}{1 + e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}}$$

Ako je  $\pi$  vjerojatnost neispunjenja obveze, tada se omjer

$$\frac{\pi}{1 - \pi}$$

naziva odnos šansi. Odnos šansi dan je s

$$\frac{\pi}{1 - \pi} = e^{\alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k}$$

Logaritmiranjem objiju strana jednadžbe dobivamo:

$$\log\left(\frac{\pi}{1 - \pi}\right) = \alpha + \beta_1 x_1 + \beta_2 x_2 + \dots + \beta_k x_k$$

Uvođenjem logit funkcije možemo pisati

Razmotrimo sada neke posebne slučajeve modela multiple logističke regresije. Ako je u modelu samo jedan prediktor i ako je on binarnoga karaktera, model se može zapisati na sljedeći način:

Iz navedenog se može zaključiti da razlika između  $\log odds$  kada je  $X=1$  u usporedbi kada je  $X=0$  iznosi upravo  $\beta_1$ .

$$\log it (Y = 1 | X = 0) = \alpha$$

$$\log it (Y = 1 | X = 1) = \alpha + \beta_1$$

Neka je

$$P^0 = P(Y = 1 | X = 0)$$

Tada se regresijski  $P^1 = P(Y = 1 | X = 1)$  parametri

mogu interpretirati na sljedeći način:

Prema Halmi [5], kako bi model logističke

$$\beta_0 = \log it (P^0) = \log\left[\frac{P^0}{1 - P^0}\right]$$

$$\beta_1 = \log it (P^1) - \log it (P^0)$$

$$\beta_1 = \log\left[\frac{P^1}{1 - P^1}\right] - \log\left[\frac{P^0}{1 - P^0}\right]$$

$$\beta_1 = \log\left[\frac{P^1 / (1 - P^1)}{P^0 / (1 - P^0)}\right]$$

regresije bio valjan, treba biti zadovoljeno nekoliko preduvjeta. Prva je pretpostavka da su slučajne kategorijalne varijable definirane kao dihotomne varijable koje poprimaju vrijednosti 1 i 0. Nadalje, opažanja moraju biti statistički nezavisna, odnosno neautokorelirana, model mora sadržavati sve relevantne prediktore, a slučajevi koji su predmet analize ne mogu biti u više izlaznih kategorija.

## 3. Diskriminativna analiza

Diskriminativna analiza (Härdle, Simar [8], Izenman [10], Rencher [17], Yang, Trewn [21]) multivarijantna je tehnika koja se koristi s ciljem determiniranja seta varijabli koje prave najveću razliku između definiranih skupina (npr. klijenata koji uredno i neuredno izvršavaju obveze iz ugovora). Pomoću diskriminativne analize ispitanici se klasificiraju u jednu od dvije ili više grupa na temelju skupa obilježja.

Neka se u kontekstu kreditnoga scoringa zamisli kreditni službenik koji odlučuje o odobravanju kredita klijentu. Kreditni će službenik donijeti odluku određujući jesu li karakteristike klijenta (npr. dob, njegovi prihodi, bračni status, dugovanja i imovina) navedene u zahtjevu za kredit, više slične klijentima koji su u prošlosti uredno vršili otplatu kredita ili onima koji nisu uredno ispunjavali ugovorne obveze. Postupak rješavanja ovog problema počinje definiranjem uzorka klijenata koji će se promatrati tijekom određenog razdoblja, snimanjem njihovih karakteristika i utvrđivanju statusa u ispunjavanju obveza na kraju testnog razdoblja. Ako klijent ne izvrši svoje obveze na kraju razdoblja, klasificira se kao loš, a ako klijent uredno izvrši ugovorne obveze, klasificira se kao dobar. S ovako prikupljenim podacima problem kreditnog odlučivanja može se rješavati primjenom dvogrupne linearne diskriminativne analize (Kvesić [13]). U tom je slučaju moguće definirati samo jednu diskriminativnu funkciju.

Diskriminativna funkcija predstavlja linearnu kombinaciju diskriminativnih varijabli:

$$Z_i = a_1 X_1 + a_2 X_2 + \dots + a_m X_m$$

Cilj je odrediti koeficijente koji maksimiziraju razliku između grupa. To je moguće ostvariti maksimiziranjem međugrupne varijance u odnosu prema unutargrupnoj. Taj odnos dan je sljedećim izrazom:

$$\lambda_j = \frac{A^T S_B A}{A^T S_W A},$$

gdje je:

$A = [a_1, a_2, \dots, a_m]$  vektor diskriminativnih koeficijenata  
 $S_B$ - matrica varijacija i kovarijacija između grupa  
 $S_W$ - matrica varijacija i kovarijacija unutar grupa

Izrazi za matrice varijacija i kovarijacija  $S_B$  i  $S_W$  su:

$$S_B = \sum_{k=1}^c n_k (\bar{x}_k - \bar{\bar{x}})(\bar{x}_k - \bar{\bar{x}})^T$$

$$S_W = \sum_{k=1}^c \sum_{i=1}^{n_k} (x_i - \bar{x}_k)(x_i - \bar{x}_k)^T$$

U prethodnim je izrazima:

$$\bar{x}_k = \frac{1}{n_k} \sum_{i=1}^{n_k} x_i$$

$$\bar{\bar{x}}_k = \frac{1}{n} \sum_{i=1}^{n_k} n_k \bar{x}_k$$

Da bi se riješio problem, potrebno je izračunati niz parcijalnih derivacija. Na taj način dobiveni vektori mogu se napisati u sljedećem obliku:

$$(S_B - \lambda S_W) A = 0$$

Problem se nadalje svodi na rješavanje karakteristične jednadžbe:

$$(S_W^{-1} S_B - \lambda I) A = 0$$

Nakon izračuna maksimalnog  $\lambda$ , omjera on se uvrštava u gore navedenu jednadžbu kako bi se odredio vektor diskriminativnih koeficijenata.

#### 4. Analiza preživljenja

Analiza preživljenja (Foster, Barkus, Yavorsky [4], Huber [9], Lee, Wenyuwang [14], Narain [16]) grupa je statističkih metoda i tehnika posebno namijenjenih ispitivanju trajanja nekog događaja na temelju prikupljenih informacija u vremenu. Činjenica da kreditni status ima dinamički element mijenja pitanje pojave prestanaka ili propuštanja izvršenja ugovornih obveza u pitanje vremena pojave statusa neispunjavanja obveza. Kod ovoga pristupa kreditnom bodovanju pokušava se procijeniti koliko će dugo proći do pojave nekog događaja, iako u mnogim slučajevima takav događaj neće ni nastupiti.

Neka je  $T$  dužina vremena do pojave statusa neotplaćivanja kredita. Prema Banasiku, Crooku i Thomasu [1] postoje tri standardna načina opisivanja slučajnog odabira  $T$  u analizi preživljenja:

- funkcija distribucije  $F(t)$  gdje je  $F(t) = P(T \leq t)$ , a  $S(t) = 1 - F(t)$  funkcija preživljenja;

- funkcija gustoće  $f(t)$  gdje je  $p(t \leq T \leq t + \delta t) = f(t) \delta t$ ;

- funkcija hazarda  $h(t)$  gdje je  $h(t) = f(t) / (1 - F(t))$  pa je prema tome  $h(t) \delta t = P(t \leq T \leq t + \delta t | T \geq t)$

Najčešće se u analizi preživljenja pretpostavlja da razdioba trajanja slijedi eksponencijalnu i Weibullovu distribuciju.

Za eksponencijalnu distribuciju s parametrom  $\lambda$  vrijedi:

$$F(t) = 1 - e^{-\lambda t}, f(t) = \lambda e^{-\lambda t}, h(t) = \lambda$$

s očekivanjem  $1/\lambda$ .

Za Weibullovu distribuciju s parametrima  $\lambda$  i  $k$  vrijedi:

$$F(t) = 1 - e^{-(\lambda t)^k}, f(t) = k \lambda^k t^{k-1} e^{-(\lambda t)^k}, h(t) = k \lambda^k t^{k-1}$$

s očekivanjem  $\Gamma(1/k)/k\lambda$ , gdje je  $\Gamma$  – gama funkcija.

Pri procjenjivanju funkcije preživljenja i funkcije hazarda javlja se problem broja ispravno cenzuriranih promatranja. Oni se pojavljuju kada neki predmet ostane nakon određenog vremena u uporabi. Stoga je potrebno evidentirati koliko dugo je klijent bio zajmoprimalatelj i je li to završilo sa statusom neotplaćivanja ugovornih obveza ili ne.

Moguće je da postoje eksplanatorne varijable koje utječu na vrijeme preživljenja promatranog klijenta. Ove eksplanatorne varijable imaju istu ulogu kao i svojstva klijenta u normalnim pristupima kreditnog bodovanja. Kalbfleisch i Prentice [12] navode da postoje dva modela koja povezuju eksplanatorne varijable s vremenom neuspjeha u analizi preživljenja - modeli proporcionalnog hazarda i modeli ubrzanog života, no oni u ovom radu nisu posebno analizirani.

#### 5. Zaključak

Snažan rast bankarstva, a poglavito kreditne industrije, potaknuo je posljednjih desetljeća razvoj sofisticiranih modela kreditnoga scoringa, koji su danas od velike pomoći analitičarima u procesu odlučivanja o prihvaćanju, odnosno odbijanju zahtjeva zajmotražitelja. Naime u uvjetima velikog porasta potražnje za kreditima i ponude različitih kreditnih proizvoda, tradicionalni pristup procjeni kreditne sposobnosti nikako se nije mogao smatrati zadovoljavajućim. Rješenje problema bili su automatski sustavi kvantifikacije rizika temeljeni na relevantnim kriterijima važnosti određenih karakteristika klijenata. Nedavna financijska kriza, koja još uvijek nije prebrođena, kao i promjene u bankarskoj regulativi, dodatno su utjecale na podizanje svijesti o važnosti modela kreditnoga scoringa. Stoga se nedvojbeno može zaključiti da su oni neizostavna podrška u upravljanju kreditnim rizikom.

U ovom su radu prezentirane tri statističke metode za modeliranje kreditnog rizika: logistička regresija, diskriminativna analiza i analiza preživljenja. Svaka od njih posjeduje određene prednosti i ograničenja. U znatnoj se mjeri efikasnost navedenih statističkih metoda može unaprijediti primjenom simulacija. Preciznost procjene moguće je dodatno poboljšati uvođenjem različitih relevantnih varijabli u model, uključujući i onih makroekonomskega karaktera. Nezaobilazno je pitanje u izgradnji scoring modela dostupnost adekvatnih podataka i pokazatelja, stoga je posebnu pozornost potrebno posvetiti i problematici modeliranja i izgradnje baze podataka.

Pored logističke regresije, diskriminativne analize i analize preživljenja, u upravljanju kreditnim rizikom koriste se i druge tehnike, poput neuronskih mreža, stabla odlučivanja i metode k-najbližih susjeda, no one, zbog ograničenog prostora, u ovom radu nisu obrađene.

## LITERATURA

1. Banasik, J., Crook, J. N., Thomas, L. C.: "Not if but when will borrowers default", *The Journal of the Operational Research Society*, Vol. 50, No. 12, 1999, pp. 1185-1190.
2. Christensen, R.: *Log-Linear Models and Logistic Regression, Second Edition*, Springer, New York, 1997.
3. Fabozzi, F. J., Peterson, P. P.: *Financial Management and Analysis, Second Edition*, John Wiley & Sons, Inc., Hoboken, 2003.
4. Foster, J., Barkus, E., Yavorsky, C.: *Understanding and Using Advances Statistics*, SAGE Publications, London, 2006.
5. Halmi, A.: *Multivarijantna analiza u društvenim znanostima*, Alinea, Zagreb, 2003.
6. Hand, D. J., Henley, W.E.: *Statistical Classification Methods in Consumer Credit Scoring: a Review*, *Journal of the Royal Statistical Society: Series A (Statistics in Society)*, Vol. 160, No. 3, 1997, pp 523–541.
7. Harrell, F. E. Jr., *Regression Modeling Strategies - With Applications to Linear Models, Logistic Regression, and Survival Analysis*, Springer, New York, 2001.
8. Härdle, W., Simar, L.: *Applied Multivariate Statistical Analysis, Second Edition*, Springer, Berlin, 2007.
9. Huber, C.: "Robust Versus Nonparametric Approaches and Survival Data Analysis", in Nikulin, M.S., Limnios, N., Balakrishnan, N., Kahle, W. Huber-Carol, C. (Eds.): *Advances in Degradation Modeling: Applications to Reliability, Survival Analysis, and Finance*, Birkhäuser, Boston, 2010, pp. 323-337.
10. Izenman, A. J.: *Modern Multivariate Statistical Techniques: Regression, Classification, and Manifold Learning*, Springer, New York, 2008.
11. Jorion, P.: *Financial Risk Manager Handbook, Second Edition*, John Wiley & Sons, Inc., Hoboken, 2003.
12. Kalbfleisch, J. D., Prentice, R. L.: *The Statistical Analysis of Failure Time Data, Second Edition*, John Wiley & Sons, Inc., Hoboken, 2002.
13. Kvesić, Lj.: *Modeli kreditnih rizika za predviđanje statusa neispunjavanja obveza, doktorska disertacija*, Sveučilište Josipa Jurja Strossmayera u Osijeku, Ekonomski fakultet, Osijek, 2010.
14. Lee, E. T., Wenyuwang, J.: *Statistical Methods for Survival Data Analysis, Third Edition*, John Wiley & Sons, Inc., Hoboken, 2003.
15. Mays, E.: "The Basics of Scorecard Development and Validation", in Mays, E. (Ed.): *Handbook of Credit Scoring*, Glenlake Publishing Company, Ltd., Chicago, 2001, pp. 89-106.
16. Narain, B.: "Survival Analysis and the Credit-Granting Decision", in Thomas, L. C., Edelman, D. B., Crook, J. N. (Eds.): *Readings in Credit Scoring: Recent Developments, Advances, and Aims*, Oxford University Press, Oxford, 2004, pp. 235-245.
17. Rencher, A. C.: *Methods of Multivariate Analysis, Second Edition*, John Wiley & Sons, Inc., New York, 2002.
18. Tabachnick, B. G., Fidell, L. S.: *Using Multivariate Statistics, Fifth Edition*, Pearson Education, Inc., Boston, 2007.
19. Thomas, L. C., Edelman, D. B., Crook, J. N.: *Credit Scoring and Its Applications*, Society for Industrial and Applied Mathematics, Philadelphia, 2002.
20. Vittinghoff, E., Glidden, D. V., Shiboski, S. C., McCulloch, C. E.: *Regression Methods in Biostatistics: Linear, Logistic, Survival, and Repeated Measures Models*, Springer, New York, 2005.
21. Yang, K., Trewn, J.: *Multivariate Statistical Methods in Quality Management*, McGraw-Hill, New York, 2004.

Ljiljanka Kvesić

## STATISTICAL METHODS IN CREDIT RISK MANAGEMENT

### ABSTRACT

Successful banks base their operations on the principles of liquidity, profitability and safety. Therefore, the correct assessment of the ability of a loan applicant to carry out certain obligations is of crucial importance for the functioning of a bank. In the past few decades several credit scoring models have been developed to provide support to credit analysts in the assessment of a loan applicant. This paper presents three statistical methods that are used for this purpose in the area of credit risk management: logistical regression, discriminatory analysis and survival analysis. Their implementation in the banking sector was motivated to a great extent by the development and application of information and communication technologies. This paper aims to point out the most important theoretical aspects of these methods, but also to actualise the need for the development and application of the credit scoring model in Croatian banking practice.

**Keywords:** credit risk management, credit scoring models, logistical regression, discriminatory analysis, survival analysis