UNIVERZITET U NOVOM SADU
PRIRODNO - MATEMATIČKI FAKULTET
DEPARTMAN ZA MATEMATIKU I
INFORMATIKU

Sandra Rackov

Primena Koksovog PH modela u analizi kreditnog rizika

Master rad

Mentor:
Prof. dr Zagorka Lozanov-Crvenković

Novi Sad, 2013

Sadržaj

Predgovor

U radu je predstavljena primena analize preživljavanja u analizi kreditnog rizika. Nakon upoznavanja sa osnovama anlize kreditnog rizika prikazan je istorijski razvoj kreditnih skoring modela, od njegovih začetaka pa sve do savremenih metoda izračunavanja rizičnosti klijenta.

Potom fokus rada prelazi na definisanje osnovnih pojmova i funkcija u analizi preživljavanja koja vodi do definisanja Koksovog PH modela sa vremenski zavisnim i nezavisnim promenljivama. Opisani su i problemi pri upotrebi metode maksimalne verodostojnosti prilikom ocene koeficijenata. U cilju prevazilaženja tog problema definisan je metod parcijalne verodostojnosti.

Osnovna motivacija za ovaj rad predstavlja razvoj Koksovog PH skoring modela za odobravanje plasmana koji prikazuje rizičnost klijenta u zavisnosti od roka otplate, što je predstavljeno u poslednjem poglavlju. Na samom kraju rada, radi upoređivanja dobijenih rezultata, ispitivanja efikasnosti i preciznosti, razvijen je model pomoću logističke regresije.

Izuzetnu zahvalnost dugujem svom mentoru, Prof. dr Zagorki Lozanov-Crvenković za savete, pomoć i razumevanje koje mi je ukazala tokom izrade rada i za svo znanje preneto tokom studiranja.

Takođe, želim da se zahvalim svojoj porodici i prijateljima na neprestanoj podršci koju mi ukazuju svaki dan.

1

Uvod

Značajnost statistike se ogleda u širokoj primeni u realnom životu. Nekada se statistika isključivo bavila prikupljanjem, interpretacijom i prezentacijom podataka, međutim danas se koristi i za davanje procena i merenja rizika u odnosu na faktore koji određuju posmatranu pojavu.

Analiza preživljavanja je svoje začetke imala u medicini i aktuarskoj matematici gde se modeliralo vreme do pojave posmatranog događaja u zavisnosti od konkretnih faktora rizika. Poslednjih tridesetak godina primena analize preživljavanja u finansijskom sektoru je postala veoma značajna. Primena u analizi kreditnog rizika je brojna i među najznačajnijim je primena u kreditnom skoringu. Ozbiljniji pristupi problemu ocene rizičnosti klijenta i uviđanje važnosti precizno merenog kreditnog rizika su doveli do toga da se razvijaju novi, efikasniji i stabilniji modeli. Prilikom razvijanja takvih modela moguće je uključiti, pored određenih osobina klijenta i osnovnih osobina proizvoda za koje klijent aplicira i makroekonomske faktore, kao što je promena kursa i promena osnovne kamatne stope. Kako kreditni rizik varira, između ostalog i od roka otplate kredita, neophodno je razviti dinamički model koji reflektuje sve relevantne faktore rizika ali takođe odražava i ekonomske promene koje se mogu desiti tokom otplate kredita. U ovom radu ćemo prikazati primenu Koksovog PH modela i njegovih modifikacija prilikom ocene rizičnosti klijenta u kreditnom skoringu.

U prvom delu ovog rada ćemo prikazati istorijski razvoj modela u analizi kreditnog rizika koji obuhvata:

- Tradicionalne modele,
- Standardne modele - modele koji uključuju parametarske metode, između ostalog i logističku regresiju,

- Savremene modele - modele koji uključuju semi-parametarske i neparametarske metode, između ostalog i analizu preživljavanja.

Ukoliko za posmatran događaj uzmemo da je klijent kasnio sa otplatom tri rate, cilj je da ocenimo verovatnoću takvog događaja u svakom momentu roka otplate. Standardni model koji se još uvek koristi u manje razvijenim zemljama, uključuje logističku regresiju kao kombinaciju jednostavnosti i efikasnosti. Napretkom tehnologije i ozbiljnijim pristupom problematici, Koksov model predstavlja svakako bolji izbor zbog uključivanja vremenske komponente u model kreditnog skoringa.

U drugom delu rada će biti predstavljen Koksov PH model i izvešćemo sve relevantne pokazatelje. Dobro je poznato da modele preživljavanja analiziramo posmatrajući dve fundamentalne stavke:

(1) osnovnu funkciju rizika - promenu rizika tokom vremena,

(2) efekat parametara - varijacija rizika u odnosu na nezavisne promenljive.

Koksov PH model pretpostavlja da je rizik proporcionalan, stoga je moguće oceniti efekat parametara bez određivanja same funkcionalne forme rizika.

U trećem delu ćemo pomoću realnih podataka i programskog paketa IBM SPSS Modelar odrediti rizičnost klijenta u kreditnom skoringu koristeći Koksov PH model. Dobijene rezultate tada upoređujemo sa standardnim postupkom, logističkom regresijom. Zatim ćemo diskutovati o uključivanju makro ekonomskih faktora u model i poboljšanju samog modela u vidu bolje preciznosti i stabilnosti.

Rad završavamo sumirajući dobijene rezultate i izvođenjem zaključaka o značajnosti daljeg razvoja modifikacija Koksovog modela u ekonomskom sektoru.

2

Analiza kreditnog rizika

"Kredit je ugovorni sporazum po kojem klijent - zajmotražioc (dužnik) prima novac od zajmodavca (poverioca) uz dogovor da u utvrđenom roku taj novac vrati." [1]

Pojam rizika se može definisati kao mogućnost gubitka, neizvesnost ili mogućnost bilo kog ishoda koji nije očekivan.

Kreditni rizik je rizik promene kreditne sposobnosti klijenta koja će inicirati neizvršenje obaveza iz finansijskog ugovora delimično ili u potpunosti, što će izazvati da poverilac (banka) pretrpi finansijski gubitak. Kreditni rizik banke predstavlja verovatnoću da banka neće biti u stanju da naplati svoja ukupna potraživanja po osnovu glavnice duga i po osnovu ugovorene kamate što za posledicu ima oslabljen kapital banke i negativan uticaj na finansijski rezultat banke.

Kreditni rizik je najznačajniji rizik kome je banka izložena u svom poslovanju, a upravljanje kreditnim rizikom u banci je osnova za uspešno poslovanje banke. Moglo bi se reći da je ovaj rizik star koliko i samo bankarstvo, jer pozajmljivanje novca drugoj ugovornoj strani uvek je sa sobom nosilo opasnost da pozajmljena stredstva neće biti vraćena. Još tada je prepoznata potreba za upravljanjem ovim rizikom kako bi se sprečile negativne posledice po bankarski posao, te su u skladu sa tim preduzimane različite aktivnosti. Stepen složenosti mehanizma kojima se upravljalo kreditnim rizikom tada je bio daleko jednostavniji nego što je danas. Danas, banka je dužna da kreditni rizik identifikuje, meri i procenjuje na osnovu kreditne sposobnosti dužnika i na osnovu kvaliteta založenih sredstava obezbeđenja.

[1] http://www.nbs.rs/internet/cirilica/glossary.html

Pod kreditnom sposobnošću dužnika podrazumeva se njegova sposobnost da uredno servisira obaveze prema banci i vrati ih u ugovorenom roku. Sredstvo obezbeđenja osim što služi kao pokriće poverioca (banke) u slučaju neizvršenja obaveza, takođe predstavlja i podsticaj dužniku da kredit otplaćuje prema ugovoru. Postojanje sredstava obezbeđenja umanjuje kreditni rizik, ukoliko se sredstvo obezbeđenja može lako realizovati, odnosno prodati. Sredstva obezbeđenja, kao instrumenti zaštite od kreditnog rizika, su široko prisutni u praksi.

Metodologija utvrđivanja kreditnog rizika nalazi se u središtu kreditne analize, ali je osnovna ideja kreditne politike banke minimizacija kreditnog rizika i maksimizacija profita. Jedan od osnovnih zahteva u banci je sposobnost u pokrivanju nastalih gubitaka na osnovu plasiranih kredita. Cilj je da se na adekvatan način kontroliše izloženost riziku i prate sva eventualna pogoršanja kao i da se preduzmu preventivne mere kako bi se nepovoljne situacije sprečile. Efikasan sistem upravljanja kreditnim rizikom utiče na eliminaciju većine problema prisutnih u bankama. Upravljanje kreditnim rizikom vrši se primenom limita, selekcijom kreditnih zahteva, diversifikacijom plasmana i primenom adekvatnih sredstava obezbeđenja plasmana.

Kreditni rizik se posmatra na nivou pojedinačnih plasmana, na nivou klijenta/dužnika i na nivou celokupnog portfolia. Na nivou plasmana kreditni rizik se može umanjiti definisanjem:

- vremenskog perioda kreditiranja;
- kreditne sposobnosti dužnika;
- kreditnog limita dužnika;
- kontrole korišćenja kredita;
- obezbeđenje povrata kredita.

Na nivou portfolia, umanjenje rizika se postiže:

- limitiranjem veličine kredita prema vrsti korisnika kredita;
- restrikcijom odobravanja kredita za pojedina regionalna područja;
- polaganjem depozita sa ciljem smanjenja broja kreditnih zahteva.

Parametri koji utiču na kreditni rizik izražavaju se kroz formulu:

$$EL = EAD * PD * LGD$$

pri čemu su oznake sledeće:

EL (expected loss) - očekivani gubitak;

EAD (exposure at default) - izloženost u trenutku neizvšenja obaveza, što predstavlja iznos glavnice sa pripadajućom kamatom, koje banka još nije naplatila kroz proces amortizacije kredita u trenutku neizvršenja obaveza klijenta;

PD (probability of default) - verovatnoća neizvršenja obaveza je verovatnoća da klijent neće otplatiti tri dospela mesečna anuiteta;

LGD (loss given default) - gubitak usled neizvršenja obaveza.

Neizvršenje kreditnih obaveza dužnika predstavlja kašnjenje klijenta od 3 meseca sa značajnim iznosom duga ili jednostavnije rečeno, 90 dana docnje. Pod terminom neizvršenja kreditnih obaveza dužnika smatra se i ako dužnik prekrši neku od zaštitnih klauzula u kreditnom ugovoru, što automatski pokreće pregovore između banke i dužnika, u protivnom banka zahteva da dužnik vrati celokupan dug.

Kreditni proces počinje sa kreditnim zahtevima od strane kompanija ili stanovništva (fizičkih lica). Ovi kreditni zahtevi ulaze u proceduru koja ima za cilj da se izvrši adekvatna analiza kreditnog rizika. Izvori informacija koji su relevantni za kreditnu ocenu se zasnivaju na informacijama koje podnosi tražilac kredita, na osnovu baze podataka kojom raspolaže banka i na osnovu spoljnih informacija koje banka prikuplja. Baze podataka u samoj banci su veoma značajne. Osim što pružaju istorijski uvid u karakteristike klijenata istog ili sličnog kreditnog proizvoda, može se desiti da postoje određene informacije o tražiocu kredita, na osnovu dotadašnje saradnje sa njim. Na osnovu svih raspoloživih informacija banka donosi pozitivnu ili negativnu kreditnu odluku. Suština donošenja kreditnih odluka je da banka proceni stepen kreditnog rizika.

Kreditna analiza je analiza koju banka obavlja pri dodeljivanju kredita tražiocu sa ciljem da utvrdi njegovu kreditnu sposobnost i na taj način ustanovi

stepen kreditnog rizika. Svaki kredit koji banka odobri sadrži u određenoj meri kreditni rizik, ali banka u svojoj kreditnoj politici mora da utvrdi stepen rizika koji može da prihvati. U savremenom bankarstvu proces kreditiranja je postao znatno složeniji nego u ranijim uslovima. Važnost korišćenja višestrukih izvora informacija tražiocima kredita, posebno kompanijama, je u ukrštanju informacija, odnosno podataka, kako bi se proverila njihova tačnost.

Kreditna sposobnost se može ispitati na dva načina:

1. Primenom klasične kreditne analize (kvalitativno) - daje se opisna ocena rizičnosti klijenta analizom dostavljenih finansijskih izveštaja;

2. Primenom kreditnog skoring/rejting modela (kvantitativno) - računa se verovatnoća neizvršenja obaveza klijenta primenom kreditnog skoring/rejting modela.

Klasična kreditna analiza, koja se još uvek primenjuje pri odobravanju kredita velikim kompanijama, predstavlja detaljnu analizu dostavljenih finansijskih izveštaja i kao rezultat daje opisnu ocenu rizičnosti tražioca kredita. Klasična kreditna analiza obuhvata opis delatnosti kojom se tražilac kredita bavi, kratak istorijat poslovanja, broj zaposlenih, analiza povezanih lica, ocena ponuđenih sredstava obezbeđenja, analiza platnog prometa, saradnja sa drugim bankama, analiza strukture prihoda i ključnih finanskijskih indikatora, analiza vlasničke strukture tražioca kredita i rukovodstva i imovine u tražiocevom vlasništvu, i promenu svih navedenih faktora u bar dva poslednja finansijska izveštaja. Klasična kreditna analiza predstavlja subjektivnu ocenu kreditnog analitičara koji upotrebljavaju svoje znanje i iskustvo radi ocene kreditne sposobnosti tražioca kredita. Međutim, klasična kreditna analiza se pokazala kao vremenski veoma zahtevan i skup proces i zavistan od subjektivnih stavova kreditnih analitičara pa su iz tog razloga banke pokušale da usavrše, ubrzaju i skrate proces donošenja odluka, što je predstavljalo početak razvoja kreditnih skoring modela.

Kreditni rejting je zajednički naziv za kriterijume na osnovu kojih se određuje kreditna sposobnost dužnika da redovno servisira kredit. Kreditni rejting uključuje kako formalne kriterijume utemeljene na kreditnoj istoriji neke osobe ili kompanije, tako i apstraktne kriterijume koji mogu biti reputacija ili životne navike fizikčkog lica ili kompanije, odnosno politička nestabilnost kada su u pitanju države. Ukoliko primenjujemo rejting modele, potrebno je razlikovati eksterne i interne kreditne rejtinge. Eksterne kreditne rejtinge

obezbeđuju Agencije za kreditni rejting od kojih su najpoznatije: "Moody's Investors Service", "Standard & Poor's", "Fitch Group" i "Dominion Bond Rating Service". Interni rejting modeli su proizvod samih banaka i zavise od kriterijuma koje su banke same postavile.

Skoring modeli dodeljuju kvantitativnu meru potencijalnom klijentu predstavljajući na taj način verovatnoću budućeg ponašanja u otplati kredita za koji klijent aplicira. Kreditni skoring modeli se razvijaju radi prevazilaženja subjektivne ocene kreditnog analitičara pri ocenjivanju kreditne sposobnosti tražioca kredita i finansijske institucije su neprestano u procesu pronalaženja sve boljih kreditnih skoring modela. Kao polaznu pretpostavku uzimamo da su ocene kreditne sposobnosti potencijalnog klijenta primenom kreditnog skoring modela znatno preciznije od subjektivnih procena kreditnih analitičara. Istorijski razvoj i primena kreditnog skoring modela ukazuje da je kreditni skoring model zaista precizniji od subjektivnih ocena kreditnih analitičara pri oceni kreditne sposobnosti potencijalnih klijenata što znači da u većoj meri prepoznaje kreditno sposobne klijente i one koji to nisu. Na taj način banka odobrava više plasmana kreditno sposobnim klijentima a manje kreditno nesposobnim klijentima. Primenom kreditnog skoring modela se identifikuju sve one karakteristike komitenata koje najbolje predviđaju otplatu kredita i kao takvi služe kao automatizovano, efikasno i konzistentno sredstvo pri donošenju odluka o odobravanju kredita. Kreditni skoring model predstavlja sistem dodeljivanja bodova potencijalnom klijentu u cilju dobijanja numeričke vrednosti koja pokazuje verovatnoću neizvršenja obaveza po kreditu za koji klijent aplicira.

Kreditni skoring modeli omogućavaju bankama koje ih primenjuju da ponude povoljnije kreditne uslove dobrim klijentima (onima sa značajnom imovinom i dobrom kreditnom istorijom) i po nižim troškovima za banku, u poređenju sa bankama koje primenjuju tradicionalan način odobravanja kredita. Postoje razlozi za i protiv kreditnog skoring modela. Generalno, sa jedne strane skoring je efikasniji i čini kreditni proces bržim, a i ne postoji pristrasnost koja se može pojaviti kod kreditnih referenata, ipak, s druge strane, mogu se pojaviti nekakve jedinstvene karakteristike koje skoring neće uzeti u obzir, a što bi kreditni referent svakako primetio.

Ako se posmatraju prednosti koje kreditni skoring ima prilikom odobravanja kredita, generalno bi se moglo navesti sledeće:

- Kreditni skoring modeli su objektivni, konzistentni i efikasni - kako je proces automatizovan, kreditni analitičari ne moraju manuelno da pregledaju finansijske izveštaje, već svoje vreme mogu efikasnije iskoristiti;

- Smanjuje se operativni rizik - rizik pri samom radu kreditnih analitičara se minimizuje jer je proces automatizovan;
- Kreditni skoring modeli su relativno jeftini - smanjuju se troškovi osoblja što utiče na smanjivanje cene kredita;
- Kreditni skoring modeli su relativno jednostavni i lako se interpretiraju;
- Metodologija upotrebljena u izgradnji takvih modela je uobičajena i shvatljiva;
- Banke ostvaruju bolje usluge potencijalnim klijentima svojom sposobnošću da brzo odgovore na njihove zahteve;
- Kako kreditni skoring određuje verovatnoću hoće li komitent kasniti ili ne sa većom preciznošću, moguće je cenu kredita prilagoditi riziku. Na taj način se povećava profitabilnost i smanjuje cena kredita što je poželjno za potencijalne klijente;
- U zavisnosti od uslova na tržištu, banka je u mogućnosti da oceni količinu kredita koju će ponuditi u skladu sa kreditnom politikom banke. Povećavanjem i snižavanjem granične vrednosti banke mogu kontrolisati tržišnu aktivnost ili iznos plasiranih kredita;
- Osigurava se bolja kontrola kreditnog portfolia i njegovih karakteristika putem nadgledanja i revidiranja kreditnog skoring modela.

Nedostaci kreditnog skoring modela su sledeći:

- Modeli degradiraju tokom vremena - ako se populacija na kojoj se kreditni skoring modeli primenjuju promeni u odnosu na originalan uzorak prema kome je model razvijen, model neće biti prediktivan;
- Razvijeni kreditni skoring modeli se primenjuju samo na tipove klijenata koji su bili uključeni pri razvoju modela. Na primer, ako se banka odluči za izdavanje kreditnih kartica studentima, ali koristi kreditni skoring model koji je razvijen na temelju uzorka koji nije sadržao studente, model neće dobro razlikovati dobre i loše klijente;
- Verovatnoće neizvršenja obaveza izračunate na osnovu interno razvijenog kreditnog skoring modela ne mogu se posmatrati na tržištu. Ukoliko je predmet interesovanja rizičnost klijenta na tržištu, postoje javno dostupne informacije o kreditnom rejtingu od strane Agencija

za rejting ili generički kreditni skoring modeli o kojima će biti reči u nastavku. Verovatnoća neizvršenja obaveza je interni podatak koji prikazuje verovatnoću da će klijent kasniti sa otplatom po određenom plasmanu za koji je aplicirao;

- Ako ne postoji dovoljan obim uzorka ili u sistemu banke ne postoje istorijski podaci, model se ne može razvijati.

Postoje dve vrste kreditnog skoring modela ako posmatramo podatke koji se koriste u njihovom razvoju:

(1) Generički kreditni skoring modeli

Bazirani su isključivo na podacima kreditnih biroa koji raspolažu sa ogromnom bazom podataka o kreditnoj istoriji klijenata koji imaju tekuće račune. Na temelju takve baze podataka primenom različitih metoda kreiraju se kreditni skoring modeli koji obuhvataju one karakteristike potencijalnih klijenata koje najbolje predviđaju buduće ponašanje u otplati kredita. Prvi generički skoring model bio je "Prescore" koji je dizajnirao "Fair Isaac and Co." u periodu od 1984. do 1985. godine. U razvijanju generičkih skoring modela ili kako se još nazivaju, skoring modeli kreditnog biroa, analitičari identifikuju one karakteristike klijenata koje najbolje predviđaju hoće li on otplatiti svoj kredit u celini i na vreme. Za svaku karakteristiku se određuje numerička vrednost tako da kreditni sistem ocenjuje značajnost date karakteristike u preciznom predviđanju otplate kredita.

(2) Kreditni skoring modeli prilagođeni korisniku

Bazirani su na podacima o klijentima konkretne finansijske institucije. Dakle, razvijaju se zasebno za svaku finansijsku instituciju. Procedure podržane statističkim i drugim metodama se primenjuju na podatke kojima raspolaže određena finansijska institucija pa se izdvajaju one karakteristike klijenta koje su značajne za otplatu kredita.

U praksi, banke koriste korisniku prilagođene kreditne skoring modele razvijene na osnovu baze podataka koju banka poseduje i na osnovu podataka iz kreditnog biroa. Tako formirani skoring modeli, u praksi su se pokazali veoma precizni. Pri samom odabiru promenljivih, za kreditni skoring prilagođen korisniku, kreditori učestvuju i definišu neophodne promenljive u zavisnosti od kvaliteta i strukture dostupnih podataka.

3

Istorijski razvoj kreditnih skoring modela

Kreditni skoring predstavlja način prepoznavanja različitih grupa u populaciji kada nije moguće direktno uočiti karakteristike koje ih razdvajaju. Ideju i pojam diskriminacije među grupama u okviru jedne populacije u statistici je uveo Fisher 1936. godine, želeći da napravi razliku između dve sorte irisa. David Durand (1941) je prvi uočio da se mogu koristiti iste tehnike pri razlikovanju dobrih i loših kredita. To saznanje je ostalo zabeleženo u sklopu istraživačkog projekta za američki nacionalni biro za ekonomska istraživanja i tada se nije koristio za ocenu i predviđanje kvaliteta kredita. U isto vreme nekoliko finansijskih institucija je imalo problema sa svojim kreditnim menadžmentom, jer su njihovi kreditni analitičari koji su donosili odluke o plasiranju kredita i potencijalnoj rizičnosti klijenta, prešli u službu vojske. Pre nego što su promenili službu, finansijske institucije su tražile od kreditnih analitičara da napišu pravila po kojima su se vodili u oceni rizičnosti klijenta. Nedugo zatim se uvidela prednost statističkog skoringa. Prva konsultanska kuća je nastala u San Francisku od strane Bill Fair i Earl Isaac oko 1950-te godine. Neophodnost nastanka kreditnog skoringa i automatizacija odobravanja plasmana postala je realnost pojavljivanjem kreditnih kartica 60-tih godina prošlog veka. Osim što se proces odobravanja plasmana znatno ubrzao, pokazalo se da su kreditni skoring modeli znatno precizniji i stopa loših klijenata se umanjila za 50%. Događaj koji je presudno uticao na prihvatanje kreditnog skoringa je Akt o jednakosti mogućeg kreditiranja, koji je donet 1975. godine u Americi. Postignut uspeh pri odobravanju kreditnih kartica je imao za posledicu da su banke počele da koriste kreditne skoring modele pri odobravanju i ostalih proizvoda što je prouzrokovalo

konstantno usavršavanje kreditnih skoring modela. Razvijanjem računara i primenom u svakodnevnom životu, omogućili su 80-tih godina prošlog veka implementaciju linearnog programiranja i logističke regresije u skoring modele, dok je poslednjih godina zastupljena kombinacija standardnih metoda sa tehnikama veštačke inteligencije i neuronskih mreža. Godinama unazad, cilj razvoja i primene kreditnih skoring modela je bio razlikovanje dobrih i loših plasmana radi minimizacije plasiranja sredstava u loše klijente. Međutim, taj cilj se danas modifikovao u maksimizaciju profita koji ostvaruju finansijske institucije na svakom plasiranom plasmanu.

3.1 Pregled metoda kreditnog skoringa

3.1.1 Tradicionalni modeli

"5C" model

Tradicionalni model koji nalaže da je kreditnom analizom potrebno ispitati "5 Cs of Credit" [2]

- Karakteristike tražioca kredita (Character)

Procena karaktera dužnika podrazumeva analizu ličnih osobina, poslovnog ugleda i rukovodstva tražioca kredita. Predstavlja subjektivnu ocenu finansijske institucije i iz tog razloga se detaljno analizira poslovna reputacija, vrsta delatnosti i pravni status tražioca kredita u cilju utvrđivanja odgovornosti, integriteta i tačnosti u izmirivanju svojih obaveza i doslednosti u vođenju poslovnih knjiga. Stoga zaključujemo da se karakter utvrđuje u cilju procene spremnosti i želje tražioca kredita da servisira dug na osnovu odobrenog mu plasmana.

- Kapacitet ili sposobnost otplate (Capacity)

Finansijska institucija pre momenta odobravanja plasmana mora utvrditi najmanje dva jasno identifikovana, međusobno nezavisna izvora otplate kredita. Izvori otplate kredita mogu biti: dobit, prihod ostvaren prodajom aktive, prihod ostvaren prodajom akcija ili sredstava dobijena od drugih finansijskih

[2]Vunjak N., Kovačević Lj., s.561

institucija. Pri analizi kapaciteta tražioca kredita potrebno je analizirati dobit, buduću dobit (obim posla i njegovu prirodu), postojeći dug i strukturu troškova. Stoga zaključujemo da se kapacitet tražioca kredita utvrđuje u cilju procene sposobnosti otplate kreditnih obaveza na osnovu tekućeg prihoda koji će se generisati u periodu ugovorene otplate kredita.

- Kapital ili imovina dužnika (Capital)

Kapital predstavlja jedan od izvora otplate kredita, tako da je kreditni analitičar u procesu finansijske analize u obavezi da adekvatno proceni njegovu realnu vrednost. Kapital tražioca kredita prikazuje njegovu neto imovinu, što predstavlja jedan od pokazatelja finansijskog stanja u prethodnom periodu. Neto imovinu dobijamo kao razliku ukupnih sredstava i ukupnih obaveza. Na osnovu kapitala tražioca kredita, banka limitira iznos kredita koji može da odobri i pod kojim uslovima. Veći iznos stalnog kapitala, za banku predstavlja manji kreditni rizik. Da bi se zaštitila od rizika koji nosi pogrešno prikazani finansijski izveštaji, neophodno je da revizori koji su izvršili pregled, daju pozitivno mišljenje.

- Kolaterali i ostala sredstva obezbeđenja (Collateral)

Zaloga ili kolateral predstavlja realno sredstvo obezbeđenja za banku od kreditnog rizika, odnosno predstavlja sekundarni izvor otplate kredita. Kolateral uglavnom predstavlja uslov odobravanja kredita, pri tome je neophodno dostaviti jasne dokaze o vlasništvu sredstava obezbeđenja, jedinstvenu identifikaciju i dokazanu utrživu vrednost. Zbog rizika da li će banka uspeti da se zaštiti od eventualnog gubitka, praksa je da se uzima određen procenat tržišne vrednosti kolaterala, koji je jasno definisan kreditnom politikom banke.

- Uslovi u okruženju i poslovanju (Conditions)

Ekonomski uslovi okruženja mogu imati veliki uticaj na poslovanje kako tražioca kredita tako i banke. Nepovoljni ili promenljivi makroekonomski uslovi uzrokovaće veći gubitak i time ugroziti mogućnost otplate kreditnih obaveza. Važno je da se na početku proceni tržište i uslovi na tržištu kojima je izložen tražilac kredita i identifikuju potencijalni konkurenti, ili prilike zapošljavanja u preduzeću. Budući uslovi poslovanja tražioca kredita se sagledavaju i u zavisnosti od rokova vraćanja kredita. Za kratkoročne kredite relativno je lakše sagledati trendove budućih promena i njihove efekte na poslovanje preduzeća. Što je period kreditiranja duži, to je manje moguće da se realno sagleda buduće tržišno kretanje.

Od svih karakteristika kreditne sposobnosti najveći značaj se daje prvom karakteru, volji tražioca kredita da redovno servisira dospelu obevezu prema banci. Banka može odobriti kredit i komitentu čija je kreditna sposobnost na granici prihvatljivog, ali na taj način se preuzima veći rizik i ima prava da od tražioca kredita zatraži plaćanje veće kamate nego što je uobičajeno. U praksi, banka u tim situacijama uzima jaka sredstva obezbeđenja, da bi se umanjio kreditni rizik, koliko je to moguće.

Beaver-ov model

Prvi statistički modeli su bili univarijantni, obično bazirani na računovodstvenim podacima. Beaver [3] (1966) je prezentovao svoj model koji koristi kombinacije finansijskih pokazatelja i upoređuje takve pokazatelje tražioca kredita sa standardima u industriji u kojoj tražilac kredita pripada.

Beaver je svoj model za procenu finansijskog neuspeha bazirao na sledeća tri pokazatelja:

- $\frac{tok\ novca}{ukupna\ imovina}$
- $\frac{čist\ prihod}{ukupni\ dugovi}$
- $\frac{tok\ novca}{ukupni\ dugovi}$

Za svaki pokazatelj, pojedinačno, Beaver je izračunao graničnu vrednost i u odnosu na nju, tražioce kredita je smeštao u grupu potencijalno uspešnih, ako je finansijski pokazatelj bio iznad propisane granične vrednosti, i u grupu potencijalno neuspešnih, ako je finansijski pokazatelj bio ispod granične vrednosti. Budući da su univarijantni modeli uključivali samo jednu promenljivu, vrlo teško se zaključivalo na temelju takve analize.

Z-skor model

Z-skor model je kreirao Edward Altman, koristeći klasičnu linearnu regresiju. Kao nezavisne promenljive, Altman je koristio 30 finansijskih pokazatelja, ali usavršavanjem modela došao je do finalnih pet indikatora:

[3] Beaver W."Financial ratios as predictors of failure", Empirical Research in Accounting, 1966

$X_1 = \frac{tekuća\ aktiva}{ukupna\ aktiva}$

$X_2 = \frac{zadržani\ dobitak}{ukupna\ aktiva}$

$X_3 = \frac{operativni\ dobitak}{ukupna\ aktiva}$

$X_4 = \frac{tržišna\ vrednost\ glavnice}{knjigovodstvena\ vrednost\ ukupnog\ duga}$

$X_5 = \frac{prihodi\ od\ prodaje}{ukupna\ aktiva}$

Zavisnu promenljivu, Altman je nazvao Z-skor, što predstavlja skor kreditnog rizika i meri uspeh, tj. neuspeh tražioca kredita da otplati odobren kredit. Ocenjene koeficijente za svaku promenljivu dobio je na osnovu uzorka od 33 uspešna i 33 neuspešna preduzeća. Opšta formula Z-skor modela je sledeća:

$$Z = 0.012X_1 + 0.014X_2 + 0.033X_3 + 0.006X_4 + 0.010X_5$$

Granice za promenljivu Z su sledeće:

- $Zona\ bankrotstva \Rightarrow Z < 1.81$
- $Siva\ zona \Rightarrow 1.81 \leq Z < 2.99$
- $Bezbedna\ zona \Rightarrow Z \geq 2.99$

U zavisnosti od delatnosti tražioca kredita postoje i posebne Altmanove formule, koje u ovom radu nećemo navoditi.

Nakon Z-skor modela Altman, Haldeman i Harayanan su kreirali ZETA model. Cilj ovog modela je bilo da se analizira i testira klasifikacija preduzeća na ona koja će bankrotirati i na ona koja neće. Na početku se analiziralo 27 promenljivih, ali finalni model je sadržao svega 7 promenljivih. U poređenju sa Z-skor modelom, ZETA model je pokazivao preciznije ocene neuspešnih preduzeća od 2 do 5 godina pre bankrotstva, dok je tačnost za prvu godinu gotovo jednaka za oba modela.

3.1.2 Standardni modeli

Statistička teorija nudi različite metode razvoja kreditnih skoring modela. Fokus je na statističkim metodama koje uključuju parametarske metode, kao što su linearna regresiona analiza, analiza diskriminante, analiza binarnog odgovora i metode vremenski diskretnih panela.

Analiza diskriminante

Analiza diskriminante je tehnika klasifikacije koja se obično primenjuje na korporativne klijente i predviđanje njihovih bankrota. Linearna analiza diskriminante se zasniva na proceni linearne funckije diskriminante u cilju odvajanja individualnih grupa (grupe dobrih i loših klijenata) na osnovu specifičih karakteristika. Funkcija diskriminante je predstavljena linearnom funkcijom, pri čemu su nezavisne promenljive svi dostupni podaci koje imamo o klijentima. Dobijeni rezultat predstavlja očekivanje da klijent neće izvršiti dospele obaveze u zavisnosti od karakteristika i ocenjenih koeficijenata. Dobijeni rezultat se naziva i skor i predstavlja promenljivu diskriminante. Cilj ove metode je maksimizacija razlika između grupa dobrih i loših klijenata a u isto vreme i minimizacija razlika unutar svake pojedinačne grupe. Nedostatak analize diskriminante je u tome da je rezultat proizvoljna veličina i ne može se interpretirati kao verovatnoća. Može poslužiti u poređenju predviđanja za različite klijente, pri čemu viši rezultat ukazuje na viši rizik.

Logistička regresija

Označimo zavisnu promenljivu sa Y a nezavisne promenljve sa $X_1, X_2, ..., X_n$. Regresiona analiza omogućava ocenu srednje odnosno očekivane vrednosti kao i predviđanje očekivane vrednosti zavisne promenljive, pri čemu su date vrednosti nezavisnih promenljivih, ali isto tako i testiranje hipoteza o prirodi zavisnosti između promenljivih. Za višestruku regresiju imamo sledeću jednačinu:

$$Y_i = \alpha + \beta_1 X_{1i} + \beta_2 X_{2i} + ... + \beta_k X_{ki} + \epsilon_i, \ i = 1, 2, ..., n$$

tj.

$$Y_i = \alpha + \sum_{j=1}^{k} \beta_j X_{ji} + \epsilon_i, \quad i = 1, 2, ..., n$$

gde je α odsečak koji predstavlja očekivanu vrednost za Y ako su X_i, $i = 1, ..., n$ jednaki nuli; β_i parcijalni koeficijenti regresije koji predstavljaju promenu vrednost u Y za jediničnu promenu u X_i pri čemu se pretpostavlja da su svi ostali X-evi nepromenjeni i ϵ grešku regresije. Koeficijente regresije α i β je neophodno oceniti za n datih zapažanja za X i Y. Osnovne pretpostavke koje važe, prilikom ocene parametara:

(1) *sredina je nula* - očekivanje greške regresije je 0, tj. $E(\epsilon_i) = 0, \ \forall i$,

(2) *homoskedastičnost* - disperzija greške regresije je konstanta, tj.

$$Var(\epsilon_i) = \sigma^2, \ \forall i$$

(3) *normalnost* - greška regresije ϵ_i ima normalnu raspodelu za $\forall i$,tj. $\epsilon_i : \mathcal{N}(0, \sigma^2)$

(4) *odsustvo autokorelacije* - greške regresije su nezavisne, tj.

$$cov(\epsilon_i, \epsilon_j) = 0, \ \forall i, j$$

(5) $cov(\epsilon_i, X_i) = 0$,

(6) *nestohastičnost promenljive* X_i - nezavisne promenljive X_i nisu stohastične i važi

(6.1) $\sum_{i=1}^{n}(X_i - \bar{x})^2 \neq 0, \ \forall i$,
(6.2) $\sum_{i=1}^{n}(X_i - \bar{x})^2 < \infty, \ \forall i$.

(7) Među nezavisnim promenljivama nema linearne zavisnosti.

Dobija se da je stohastiĭka jednačina višestruke regresije data na sledeći način:

$$Y_i = \widehat{\alpha} + \widehat{\beta}_1 X_{1i} + \widehat{\beta}_2 X_{2i} ... + \widehat{\beta}_k X_{ki} + \widehat{\epsilon_i}, \ i = 1, 2, ..., n$$

tj.

$$Y_i = \widehat{\alpha} + \sum_{j=1}^{k} \widehat{\beta}_j X_{ji} + \widehat{\epsilon_i}, \quad i = 1, 2, ..., n$$

gde je sa kapicom označena ocena parametara α i β, a $\widehat{\epsilon_i}$ se naziva rezidual. Metoda najmanjih kvadrata daje ocene koeficijenata α,β_i,...,β_k koje su nepristrasne i imaju najmanju varijansu u klasi svih linearnih nepristrasnih ocenjenih parametara.

Logistička regresija omogućava da se modelira odnos nezavisnih promenljivih ako je zavisna promenljiva binarna i kojom su predstavljeni dobri i loši klijenti. Klijent se definiše kao loš ukoliko nije platio tri uzastopne rate, tj. u kašnjenju je 90 dana ili više.

Pretpostavimo da je Y Bernulijeva slučajna promenljiva koja uzima vrednosti 0 ili 1, u zavisnosti da li je klijent dobar ili loš. Verovatnoća da će klijent biti loš u zavisnosti od datih nezavisnih promenljivih se definiše kao $\pi = P(Y = 1|X = x)$, a verovatnoća da je klijent dobar $1 - \pi = P(Y = 0|X = x)$. Posmatramo odnos ove dve verovatnoće:

$$odds(x) = \frac{P(Y = 1|X = x)}{P(Y = 0|X = x)} = \frac{\pi}{1 - \pi}$$

Logaritmovanjem se dobija jednačina logističke regresije kao funkcija nezavisnih promenljivih x_i, $i = 1, 2, ..., n$:

$$ln(odds(x)) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 ... + \beta_n X_n$$

tj.

$$ln(\frac{\pi}{1 - \pi}) = \beta_0 + \beta_1 X_1 + \beta_2 X_2 ... + \beta_n X_n$$

odnosno

$$\pi = \frac{e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 ... + \beta_n X_n}}{1 + e^{\beta_0 + \beta_1 X_1 + \beta_2 X_2 ... + \beta_n X_n}}$$

Pretpostavke koje važe za logističku regresiju su sledeće:

(1) Y ima Bernulijevu raspodelu sa parametrom $\pi(x)$:

$$Y : \begin{pmatrix} 0 & 1 \\ 1 - \pi(x) & \pi(x) \end{pmatrix},$$

(2) Nijedna promenljiva od značaja nije izostavljena dok nijedna promenljiva koja nema značaja nije uključena,

(3) Logaritam nezavisnih promenljivih i zavisna promenljiva su linearno nezavisne,

(4) Nema značajne korelacije među nezavisnim promenljivama.

Opisan metod je najrasprostranjeniji metod koji se koristi pri razvoju kreditnog skoring modela. Za nezavisne promenljive se uzimaju sve dostupne karakteristike klijenata kako iz baze podataka kojom raspolaže banka i kreditni biro tokom prethodne saradnje sa klijentom, tako i iz samog zahteva za kredit. Prednosti kreditnog skoring modela razvijenog putem logističke regresije se ogleda u tome da se značajnost modela kao i individualni koeficijenti mogu testirati a dobijeni rezultati se mogu direktno interpretirati kao verovatnoće neizvršenja dospelih obaveza klijenta. Veliki nedostatak ovog modela je upravo u definisanju zavisne promenljive. Neophodno je definisati period u kom se istorijski posmatra da li je klijent postao delikventan i na taj način se predviđa verovatnoća neizvršenja obaveza u već fiksiranom vremenskom periodu. U praksi se najčešće uzima period od 12 meseci. Međutim, metod logističke regresije ne daje odgovor na pitanje šta se dešava u narednih 18, 24 ili 36 meseci otplate kredita niti koja je verovatnoća neizvršenja obaveza u tim periodima. Još jedan problem koji se javlja u praksi je upravo niska preciznost modela ako promenljive nisu linearno povezane. Semiparametarske i neparametarske metode upravo prevazilaze navedene probleme.

3.1.3 Savremeni modeli

Standardni modeli podrazumevaju parametarske modele, međutim, za razvoj kreditnih skoring modela moguće je koristiti i neparametarske metode, kao što su neuronske mreže i stabla odluke i semi-parametarske metode, kao što su hazard modeli, što je predmet istraživanja ovog rada. Jednim imenom oni se nazivaju i savremeni modeli.

Neuronske mreže

Neuronske mreže predstavljaju jednu od altrnativnih metoda parametarskim metodama. One nude fleksibilniji dizajn, objektivne su i netransparentne i predstavljaju spoj između nezavisnih i zavisnih promenljivih. Međutim, sistem neuronskih mreža definišu tri faktora: ulazni, skriveni i izlazni parametri. Ulazni parametri prvo procesuiraju ulazne karakteristike do skrivenih parametara. Tada skriveni parametri izračunavaju adekvatan ponder koristeći funkcije aktivacije, kao što su funkcije hiperbolične tangente ili logističke funkcije, pre nego što proslede informacije do izlaznih parametara. Kombinacijom mnogih novostvorenih čvorova može se detektovati nelinearna veza

među podacima. Na *Figuri 3.1.* je prikazan jednostavan grafik koji je nastao percepcijom tri opisana faktora.

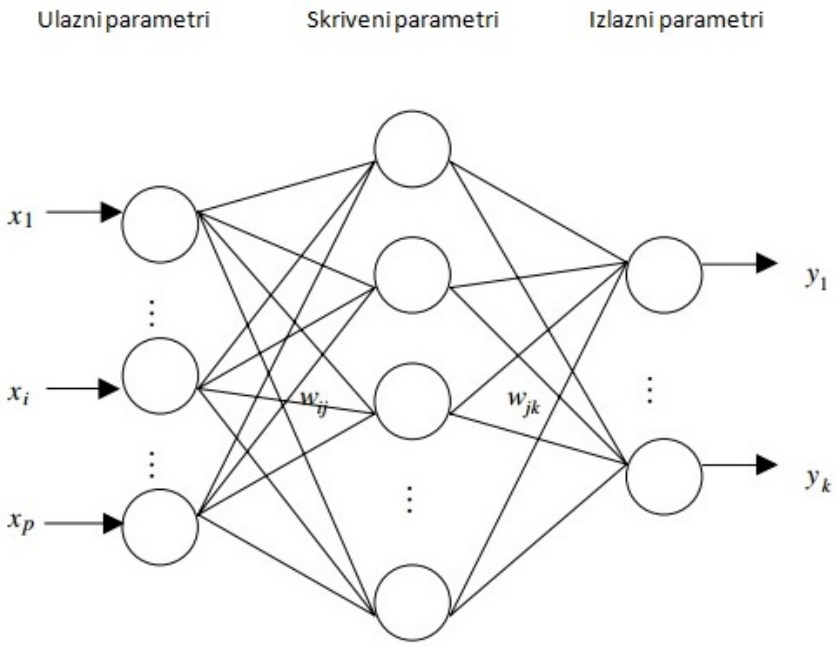


Figure 3.1: Primer neuronske mreže

Kao što je već napomenuto, neuronske mreže pripadaju klasi neparametarskih metoda. Nastanak neuronskih mreža se vezuje za rad nervnog sistema i procesom obrade informacija neurotransmitera. One se tipično sastoje od mnogo čvorova koji emituju određenu izlaznu promenljivu u slučaju da prime specifične ulazne parametre od drugog čvora u mreži sa kojom su povezani. Na sličan način, kao i parametarske metode i neuronske mreže se obučavaju putem uzorka za razvoj koji korektno klasifikuje dužnike. Finalna mreža se ostvaruje prepodešavanjem veza među čvorovima u zavisnosti od ulaznih i izlaznih promenljivih i bilo kojih potencijalnih veza među čvorovima.

Neuronske mreže, pored toga što se koriste za razvijanje kreditnog skoring modela, imaju široku primenu i u analizi i detekciji prevara.

Jedan od glavnih nedostataka neuronskih mreža je dužina trajanja procesuiranja podataka za uzorke velikog obima kao i slaba preciznost ukoliko je uzorak manjeg obima ili ako u samom uzorku postoje karakteristike koje nisu relevantne.

Stabla odlučivanja

Još jedan metod koji ima široku primenu u razvoju kreditnog skoring modela su stabla odlučivanja. U praksi je poznato i pod nazivom drvo raspodele ili stabla klasifikacije. Stabla predstavljaju modele koji se sastoje iz skupa ako-onda uslova deljenja kod slučajeva klasifikacije na dve ili više različitih grupa. Ovom metodom osnovni uzorak se deli na grupe u skladu sa nezavisnim promenljivama. U slučaju binarne klasifikacije, na primer, svaki čvor stabla je dodeljen pravilu odlučivanja koje opisuje uzorak i deli ga na dve podgrupe. Sada se proces posmatranja razvija naniže preko drveta u skladu sa pravilom donošenja odluka sve do krajnjeg čvora u razgranatoj šemi stabla, koji tada predstavlja klasifikaciju ovog posmatranja. U zavisnosti od izabranog algoritma, cilj je kategorizacija neprekidnih promenljivih i rekategorizacija već postojećih kategorijalnih promenljivih. Zaključujemo da je jedna od ključnih razlika u odnosu na parametarske modele ta da su sve nezavisne promenljive tretirane kao kategorijalne promenljive.

Pored do sada navedenih metoda koja se mogu koristiti pri izradi kreditnog skoring modela, postoje još mnoge druge, međutim u ovom radu smo naveli najzastupljenije. Potrebno je naglasiti da se većina banaka ne ograničava samo na jednu od navedenih metoda, već se često koriste njihove kombinacije.

3.2 Tipovi kreditnog skoring modela

U zavisnosti od namene kreditnog skoring modela, razlikuju se sledeći tipovi:

- **Kreditni skoring modeli za odobravanje plasmana na osnovu podataka iz zahteva za plasman** - koriste se za određivanje verovatnoća neizvršenja obaveza na osnovu čega se odlučuje da li će se kredit odobriti ili ne i pod kojim uslovima. Promenljive koje se najčešće koriste su promenljive o prihodima klijenta, podaci o dužini radnog odnosa, mesto stanovanja itd. Klijent ove podatke dostavlja dostavlja prilikom podnošenja zahteva za plasman.
- **Kreditni skoring modeli za predviđanje klijenata koji žele da prekinu saradnju sa bankom**-koriste se za identifikaciju klijenata koji žele da zatvore ili smanje promet preko svojih računa ili da prevremeno otplate postojeće plasmane. Identifikacija takvih klijenata dozvoljava menadžmentu da preduzme proaktivne mere u cilju sprečavanja nastanka opisanih situacija. Kao cilj se nameće zadržavanje dobrih klijenata u portfoliu banke.

- **Kreditni skoring modeli za predviđanje bankrota**-koriste se za anlizu i predviđanje klijenata koji će najverovatnije bankrotirati, radi sprovođenja kontrola i dodatnih mera u cilju minimizacije potencijalnih gubitaka.

- **Kreditni skoring modeli o opštem ponašanju klijenta**-koriste se za interno rangiranje klijenata koje se dobija na osnovu istorijskih podataka dostupnih za pojedinačnog klijenta iz prethodne saradnje sa bankom. Pomažu pri oceni kreditnog rizika jer smo već upoznati sa ranijim ponašanjem klijenta. Ovi podaci se mogu koristiti i pri odobravanju plasmana kao i u procesu naplate.

- **Kreditni skoring modeli za naplatu**-koriste se za rangiranje klijenata po verovatnoći neizvršenja ovabeza. Klijenti se svrstavaju u nekoliko kategorija i u zavisnosti od kategorije u kojoj se nalaze, sprovode se različite strategije naplate. Ovi rezultati se koriste u srednjim i kasnim stupnjevima delikvencije.

- **Kreditni skoring modeli za analizu i detekciju prevara**-koriste se za identifikaciju računa sa potencijalno spornim aktivnostima. Prevare su najviše zastupljene kod kreditnih kartica, pa njihova detekcija pomaže u identifikaciji i kontrolisanju potencijalnih gubitaka kao i asistencija menadžmentu banke u razvoju novih kontrola za prevenciju prevara.

- **Kreditni skoring modeli za projekciju plaćanja**- koriste interne podatke u banci u cilju rangiranja klijenata obično po relativnom procentu duga koji će se otplatiti. Neki modeli predviđaju procenat otplate duga, dok drugi klasifikuju verovatnoću otplate duga. Rezultati ovog modela se obično koriste u početnim i srednjim stupnjevima delikvencije.

- **Kreditni skoring modeli odziva**-koriste se za upravljanje troškova akvizicije novih klijenata. Banke su u mogućnosti da prilagode svoje marketinške kampanje identifikacijom ciljne grupe potencijalnih klijenata minimizirajući troškove akvizicije.

- **Kreditni skoring modeli za optimizaciju prihoda**-koristi se za maksimizaciju prihoda u odnosu na rizik koji klijent nosi, tako da se za klijente sa niskim rizikom neizvršenja obaveza smanjuje kamatna stopa, a za klijente sa visokim rizikom kamatne stope se povećavaju.

4

Osnovni pojmovi u analizi preživljavanja

Analiza preživljavanja ima svoju primenu u raznim oblastima. Iako je osnovna funkcija analize preživljavanja uvek ista, u inženjerstvu koristimo termin *Analiza pouzdanosti*; u sociologiji je poznata pod imenom *Analiza istorijskog događaja*; u ekonomiji pod imenom *Analiza trajanja* dok je u medicini poznat termin *Analiza preživljavanja*.

Analiza preživljavanja predstavlja skup statističkih metoda za analizu podataka pri čemu je promenljiva od interesa vreme dok se posmatran događaj ne realizuje. Vreme se može iskazati satima, danima, nedeljama ili godinama i ono predstavlja neprekidnu slučajnu promenljivu koja prima pozitivne realne vrednosti.

Označimo sa:

T - vreme dok se posmatran događaj ne pojavi; predstavlja slučajnu promenljivu koja ujedno označava i vreme preživljavanja subjekta;

t - označava bilo koju specifičnu vrednost od interesa za promenljivu T.

U zavisnosti od primene analize preživljavanja mogu se posmatrati razni događaji. U medicini, posmatrani događaj je obično smrt pacijenta, povratak bolesti ili oboljenje. U sociologiji posmatran događaj može biti vreme trajanja brakova, vreme do napuštanja škole ili vreme do izvršenja zločina. U analizi kreditnog rizika posmatran događaj je default klijenta što predstavlja nemogućnost klijenta da redovno izmiruje svoje dospele obaveze. Default događaj predstavlja 90 dana kašnjenja sa otplatom kredita, odnosno klijent nije platio tri dospela mesečna anuiteta po barem jednom aktivnom plasmanu.

Kako u ovom radu proučavamo primenu analize preživljavanja u analizi kreditnog rizika, posmatran događaj predstavlja vreme do pojavljivanja default događaja klijenta, što označava neuspeh.

Veoma bitno je napomenuti da se tokom jedne analize predpostavljamo da je samo jedan događaj, nad posmatranim subjektima, nama od interesa ali se može posmatrati i više od jednog događaja. Kada se posmatra više od jednog događaja tada se statistički problem karakteriše kao problem višestrukog rizika, što je poznato kao *Teorija prebrojavanja* ali je ona van domena ovog rada.

Cenzurisanje predstavlja značajan pojam u analizi preživljavanja. Osnovni razlog za cenzurisanje jeste ograničavanje vremena posmatranog perioda za mogućnost pojavljivanja događaja od interesa. Cenzurisanje se pojavljuje kada imamo delimičnu informaciju o vremenu pojavljivanja događaja. Ukoliko postoji informacija o vremenu ulaska subjekta u period posmatranja ali vreme pojavljivanja posmatranog događaja ostaje nepoznato, samim tim i ukupno vreme preživljavanja ostaje nepoznato, takva situacija predstavlja jedan klasičan primer cenzurisanja. Ono što je poznato je $T > t$. Cilj je da se iskoristi posmatrano vreme pojavljivanja događaja kod ostalih subjekata kako bi se izveli zaključci o tačnom vremenu preživljavalja subjekata kod kojih je ta informacija nepoznata.

Postoje više razloga zbog kojih se pojavljuje cenzurisanje:

- u posmatranom periodu kod subjekata se nije pojavio događaj od interesa;
- u posmatranom periodu subjekat je refinansirao svoje obaveze ili je restruktuirao plasman i tada kažemo da je subjekat izgubljen tokom posmatranog perioda;
- subjekat se isključuje iz analize zbog smrtnog ishoda ili nekog drugog razloga;
- subjekat je ranije otplatio plasman u posmatranom periodu.

Neka je $\delta \in (0, 1)$ indikator promenljiva koja predstavlja ili cenzurisanje ili default događaj (neuspeh). Ukoliko se događaj od interesa pojavio tokom posmatranog perioda δ uzima vrednost 1 i imamo neuspeh. Ako se događaj od interesa nije pojavio u posmatranom periodu δ uzima vrednost 0 i subjekat je cenzurisan iz nekog od gore navedenih razloga.

$$\delta = \begin{cases} 1 & \text{neuspeh} \\ 0 & \text{cenzurisanje} \end{cases}$$

Pri modeliranju kreditnog rizika, postoje dve osnovne podele karakteristika nezavisnih promenljivih. Prva podela razdvaja karakteristike koje se odnose isključivo na posmatrane subjekte nasuprot socio-ekonomskim karakteristikama. Druga podela razdvaja karakteristike koje su nepromenjene tokom vremena nasuprot vremenski zavisnim promenljivama. Vremenski zavisne promenljive su promenljive čije se promene mogu direktno zapisati kao funkcije vremena preživljavanja.

Cilj analize preživljavanja predstavlja modeliranje vremena preživljavanja indirektno, preko stope rizika, koja opisuje šansu prelaska iz jednog stanja u drugo u posmatranom vremenskom periodu, pod uslovom preživljavanja do tog momenta. Analiza preživljavanja pored toga što omogućava ocenu parametara modela, omogućava interpretaciju i zaključivanje svih bitnih karakteristika.

4.1 Opšti prikaz podataka

U tabeli u nastavku je dat opšti prikaz podataka koji se koriste u razvoju jednog modela. U prvoj koloni je dat uzorak i broj subjekata u uzorku je n. U drugoj koloni je prikazano vreme do pojave događaja od interesa ili cenzurisanja. Treća kolona prikazuje idikator promenljivu δ. Ostatak tabele prikazuje vrednosti nezavisnih promenljivih i pretpostavimo da u uzorku ima p nezavisnih promenljivih.

Uzorak	$\mathbf{t}$	$\boldsymbol{\delta}$	$\mathbf{X_1}$	$\mathbf{X_2}$	...	$\mathbf{X_p}$
1	t_1	δ_1	X_{11}	X_{12}	...	X_{1p}
2	t_2	δ_2	X_{21}	X_{22}	...	X_{2p}
...	...	...	...	...	...	...
n	t_n	δ_n	X_{n1}	X_{n2}	...	X_{np}

Tabela 1. Opšti prikaz podataka

4.2 Osnovne funkcije u analizi preživljvanja

Osnovni pojmovi u analizi preživljavanja su vreme preživljavanja koje se opisuje svojom funkcijom gustine i kumulativnom funkcijom raspodele, funkcija preživljavanja i funkcija rizika.

Neka je T neprekidna slučajna promenljiva koja predstavlja vreme preživljvanja, tada se definiše $F(t)$ kao kumulativna funkcija raspodele a $f(t)$ funkcija gustine vremena preživljvanja. Funkcija preživljavanja se prema tome definiše kao $S(t) \equiv 1 - F(t)$, pri čemu je izražena verovatnoća da proizvoljna promenljiva T prekorači specifično vreme t. Tada je funkcija raspodele, koja se još naziva i funkcija neuspeha, izražena sa $F(t) = P(T \leq t)$, što implicira da je funkcija preživljvanja $P(T > t) = 1 - F(t) \equiv S(t)$.

Važi sledeća veza

$$f(t) = \lim_{\Delta \to 0} \frac{P(t \leq T \leq t + \Delta t)}{\Delta t} = \lim_{\Delta \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{\partial F(t)}{\partial t} = -\frac{\partial S(t)}{\partial t} \tag{4.1}$$

pri čemu Δ predstavlja veoma mali vremenski interval.

Funkcija preživljvanja $S(t)$ i funckija neuspeha $F(t)$ su obe verovatnoće. Funckija preživljavanja je strogo opadajuća funkcija po t i predstavlja verovatnoću da subjekat nije doživeo neuspeh do određenog vremenskog trenutka. Za $t \in (0, \infty)$, važi da je $S(t) \in [0, 1]$, pri čemu je $S(0) = 1$ i $\lim_{t \to 0} S(t) = 0$, kao i $\frac{\partial S(t)}{\partial t} < 0$. Teorijski posmatrano, $S(t)$ je opadajuća glatka kriva.

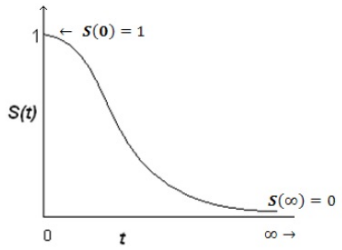


Figure 4.1: Funkcija preživljavanja

U praksi, funkcije preživljavanja su više stepenaste funkcije, pre nego glatke krive. Štaviše, zbog toga što period posmatranja nikad nije beskonačno dug

moguće je da se događaj od interesa neće desiti svim posmatranim subjektima.

Za funkciju gustine važi $f(t) \geq 0$ i ona je poznata pod terminom bezuslovna stopa neuspeha.

Funkcija rizika se označava sa $h(t)$ i predstavlja trenutni potencijal po jedinici vremena da se događaj pojavi, ako se zna da se nije pojavio do momenta t. Suprotno od funkcije preživljavanja koja se fokusira na pozitivan događaj, tj. da se događaj od interesa ne pojavi, funkcija rizika se fokusira na neuspeh, tj. da se posmatran događaj pojavi. Drugim rečima, kada $S(t)$ raste onda $h(t)$ opada i obrnuto. Rizik je stopa, a ne verovatnoća i funkcija rizika se ponekad naziva i uslovna stopa preživljavanja. Vrednost funkcije rizika se nalazi između 0 i ∞.

Stopa rizika se definiše

$$h(t) = \frac{f(t)}{1 - F(t)} = \frac{f(t)}{S(t)} \tag{4.2}$$

Ako pretpostavimo da je $P(A)$ verovatnoća napuštanja stanja u intervalu između t i $t + \Delta t$; $P(B)$ verovatnoća preživljavanja do momenta t, onda možemo izvesti verovatnoću napuštanja stanja u intervalu $[t, t + \Delta t]$, pod uslovom preživljavanja do vremena t, na sledeći način

$$P(A|B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)} = \frac{P(A)}{P(B)},$$

jer je $P(B|A) = 1$.

Ako se iskoriste ranije uvedene oznake, imamo $\frac{P(A)}{P(B)} = \frac{f(t)\Delta t}{S(t)}$. Ako uporedimo sa jednačinom (4.2), dobija se uslovna verovatnoća rizika u intervalu Δt.

$$h(t)\Delta t = \frac{f(t)\Delta t}{S(t)}$$

Za neprekidnu promenljivu T funkcija rizika $h(t)$ predstavlja stopu, međutim ako se posmatra T kao diskretni vremenski intervali, tada funckija rizika $h(t)$ predstavlja verovatnoću. Jedino ograničenje za stopu rizika, koje proizilazi iz osobina $f(t)$ i $S(t)$ jeste

$$h(t) \geq 0.$$

Bez obzira na to koja se funkcija preferira $S(t)$ ili $h(t)$, postoji jasna veza između njih. Ako se zna forma $S(t)$ tada se može izvesti i odgovarajuće $h(t)$ i obrnuto.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t \leq T < t + \Delta t | T \geq t)}{\Delta t} \tag{4.3}$$

Već smo napomenuli da je funkcija rizika stopa, a ne verovatnoća za neprekidno T. U jednačini (4.3) funkcije rizika, izraz sa desne strane posle znaka za graničnu vrednost daje odnos dve vrednosti, brojilac - koji predstavlja uslovnu verovatnoću i imenilac - što označava mali vremenski interval. Ovim deljenjem se dobija verovatnoća u jedinici vremena, što više nije verovatnoća već stopa. Skala za taj odnos nije od 0 do 1, kao za verovatnoću već je u opsegu od 0 do ∞. Uslovom $\Delta t \to 0$ funkcija hazarda želi da kvantifikuje trenutan rizik da će se posmatran događaj desiti u vremeu t, pod uslovom da se posmatran događaj nije dogodio do momenta t.

Za specifičnu vrednost t, rizik $h(t)$ ima sledeće karakteristike

- uvek je nenegativan;
- ne postoji gornja granica.

Od ove dve funkcije, $S(t)$ i $h(t)$, funkcija preživljavanja je pogodnija za analiziranje podataka o preživljavanju zato što funkcija preživljavanja direktno opisuje iskustvo preživljavanja iz posmatranog perioda. Međutim i funkcija rizika je od interesa zbog sledećih razloga

- pruža uvid o uslovnim stopama preživljavanja;
- može se koristiti za identifikaciju određenog oblika modela, kao što je eksponencijalna ili lognormalna kriva koja odgovara podacima;
- model preživljavanja se obično izračunava u uslovima funkcije rizika.

Sada želimo da izvedemo vezu između funkcije preživljavanja i funkcije rizika.

$$\begin{aligned} h(t) &= \frac{f(t)}{1 - F(t)} = \frac{-\frac{\partial S(t)}{\partial t}}{S(t)} = \frac{-\partial(1 - F(t))}{\partial t} \frac{1}{1 - F(t)} \\ &= \frac{\partial[-ln(1 - F(t))]}{\partial t} = \frac{\partial[-lnS(t)]}{\partial t} \end{aligned} \tag{4.4}$$

$$\Longrightarrow \qquad h(t) = \frac{\partial[-lnS(t)]}{\partial t} \qquad / \int_0^t$$

$$\int_0^t h(u)du = -ln[S(t)]|_0^t = -ln[S(t)] \qquad /(-1)$$

jer je $S(0) = 1$ i $ln(1) = 0$.

$$\Longrightarrow \qquad -\int_0^t h(u)du = ln[S(t)] \qquad /e$$

$$\boxed{S(t) = e^{-\int_0^t h(u)du}} \tag{4.5}$$

Ako definišemo kumulativnu funkciju rizika

$$H(t) = \int_0^t h(u)du \tag{4.6}$$

jednačinu (4.5) zapisujemo kao $S(t) = e^{-H(t)}$ i dobijamo da je

$$H(t) = -lnS(t) \tag{4.7}$$

Najznačajnije osobine $H(t)$ su $H(t) \geq 0$ i važi $h(t) = \frac{\partial H(t)}{\partial t}$.

Kako smo se upoznali sa osnovnim pojmovima analize preživljavanja, možemo reći da su ciljevi:

- procena i tumačenje funkcije preživljavanja i/ili rizika iz podataka;
- poređenje funkcije preživljavanja i rizika;
- pronalaženje veze između opisanih promenljivih sa vremenom preživljavanja.

4.3 PH pretpostavka

Do sada smo definisali funkciju rizika kao funkciju vremena preživljavanja. Međutim, sada želimo da postavimo dopunske uslove tako da stopa rizika varira među subjektima u zavisnosti od njihovih karakteristika. Postoje dve vrste modela

(1) Proporcionalni hazard modeli

(2) Modeli ubrzanog vremena neuspeha (koji u ovom radu neće biti razmatrani)

Posmatrana neprekidna stopa rizika nije uzimala u obzir bilo koje potencijalne razlike u stopi rizika između subjekata. Neka su X_1, X_2,...,X_p vektori nezavisnih promenljivih za posmatrane subjekte. Definišimo matricu $\mathbf{X}$ preko vektora karakteristika

$$\mathbf{X} = \begin{bmatrix} \mathbf{X_1} \\ \mathbf{X_2} \\ \vdots \\ \mathbf{X_n} \end{bmatrix}$$

pri čemu je $\mathbf{X_i}$, $i = 1, ..., n$ definisan vektor karakteristika subjekta i.

Neka je $h(t, \mathbf{X})$ - funkcija rizika i $S(t, \mathbf{X})$ - funkcija preživljavanja. Heterogenost i-tog subjekta se uključuje u model na sledeći način. Definišemo linearnu kombinaciju karakteristika

$$\mathbf{X_i}\boldsymbol{\beta^T} = \beta_0 + \beta_1\mathbf{X_{i1}} + \beta_2\mathbf{X_{i2}} + ... + \beta_p\mathbf{X_{ip}}$$

Posmatramo p vektora promenljivih, koje nisu vremenski zavisne, već su nepromenjene, pri čemu je neophodno oceniti $\boldsymbol{\beta}$-ta koeficijente.

Proporcionalni hazard modeli se karakterišu na sledeći način

$$h(t, \mathbf{X}) = h_0(t)e^{\mathbf{X}\boldsymbol{\beta}^T} = h_0(t)\lambda \tag{4.8}$$

gde je

- $h_0(t)$ - bazna hazard funkcija koja zavisi samo od t i za koju pretpostavljamo da je uniformna u celoj populaciji rezultirajući proporcionalnim hazardom;

- $\lambda = e^{\mathbf{X}\boldsymbol{\beta}^T}$ - nenegativna funkcija nezavisnih promenljivih koja je specifična za svakog subjekta zasebno i ne zavisi od t.

Ako posmatramo dva subjekta i i j sa vektorima karakteristika $\mathbf{X_i} = [\mathbf{X_{i1}}, \mathbf{X_{i2}}, ..., \mathbf{X_{ip}}]^T$ i $\mathbf{X_j} = [\mathbf{X_{j1}}, \mathbf{X_{j2}}, ..., \mathbf{X_{jp}}]^T$ i za vremenski trenutak t^*, dobijamo

$$\frac{h(t^*, \mathbf{X_i})}{h(t^*, \mathbf{X_j})} = \frac{\cancel{h_0(t^*)} e^{\mathbf{X_i}\boldsymbol{\beta^T}}}{\cancel{h_0(t^*)} e^{\mathbf{X_j}\boldsymbol{\beta^T}}} = e^{(\mathbf{X_i} - \mathbf{X_j})\boldsymbol{\beta^T}} \qquad /log$$

$$log \frac{h(t^*, \mathbf{X_i})}{h(t^*, \mathbf{X_j})} = (\mathbf{X_i} - \mathbf{X_j})\boldsymbol{\beta^T} \tag{4.9}$$

Primetimo da desna strana jednačine (4.9) ne zavisi od t; početna pretpostavka je da su nezavisne promenljive nezavisne tokom vremena. Sledi da proporcionalna razlika rezultira konstantnim hazardom.

Koeficijenti $\boldsymbol{\beta}$ u PH modelima se interpretiraju na sledeći način. Neka koeficijenat $\boldsymbol{\beta}_k$ uz $\mathbf{X_k}$ nezavisnu promenljivu ima osobinu

$$\boldsymbol{\beta}_k = \frac{\partial \log h(t, \mathbf{X})}{\partial \mathbf{X_k}}$$

što nam govori da u PH modelu, svaki koeficijent sumarizuje proporcionalni efekat rizika u odnosu na apsolutne promene karakteristika nezavisnih promenljivih. Ovaj efekat ne zavisi od vremena preživljavanja.

Ako je dato $h(t, \mathbf{X}) = h_0(t)\lambda$, pri čemu λ ne zavisi od vremena preživljvanja, važi PH pretpostavka, tada

$$S(t, \mathbf{X}) = e^{-\int_0^t h(u)du} = e^{-\lambda \int_0^t h_0(u)du} = S_0(t)^\lambda$$

$$S(t, \mathbf{X}) = S_0(t)^\lambda \tag{4.10}$$

pri čemu je bazna funkcija preživljavanja data sa

$$S_0(t) \equiv e^{-\int_0^t h_0(u)du} \tag{4.11}$$

Sledi, $\log S(t, \mathbf{X}) = \lambda \log S_0(t)$.

Uzimajući u obzir vezu između funckije preživljavanja i funkcije gustine vremena preživljvanja važi sledeća jednakost

$$f(t) = f_0(t)\lambda[S_0(t)]^{\lambda-1}$$

gde je $f_0(t)$ bazna funkcija gustine za koju važi $f_0(t) = f(t|\mathbf{X} = 0)$.

Alternativno, može se zapisati i preko kumulativne hazard funkcije

$$H(t) = \lambda H_0(t) \tag{4.12}$$

za $H_0(t) = -lnS_0(t)$.

Ako osnovna hazard funkcija ne zavisi od vremena preživljvanja, već je konstantna, imamo

$$h_0(u) = h, \forall u \qquad \Longrightarrow \qquad H_0(t) = \int_0^t h_0(u)du = h\int_0^t du = ht$$

Stoga u slučaju konstantne stope rizika grafik kumulativne hazard funkcije u odnosu na vreme preživljvanja daje pravu liniju. Ako je kumulativna hazard funkcija konkavnog tipa, to ukazuje da stopa rizika opada sa vremenom preživljavanja. Ako je kumulativna hazard funkcija konveksnog tipa to ukazuje da stopa rizika raste sa vremenom preživljavanja.

Veza između funkcije preživljavanja i bazne funkcije preživljavanja implicira

$$\ln[-\ln(S(t))] = \ln\lambda + \ln(-\ln(S_0(t))) = \mathbf{X}\boldsymbol{\beta}^T + \ln(-\ln(S_0(t)))$$

Ako zapišemo u obliku kumulativne hazard funkcije

$$\ln(H(t)) = \mathbf{X}\boldsymbol{\beta}^T + ln(H_0(t))$$

4.4 Uključivanje vremenski zavisnih promenljivih

Ukoliko su nezavisne promenljive vremenski zavisne, dobijamo

$$h(t, \mathbf{X}, \mathbf{Y_t}) = h_0(t)e^{\mathbf{X}\boldsymbol{\beta}^{\boldsymbol{T}}+\mathbf{Y_t}\boldsymbol{\beta}_{\boldsymbol{Y}}^{\boldsymbol{T}}} \tag{4.13}$$

Primetimo da za bilo koje dato vreme preživljavanja $t = t^*$, apsolutna razlika u promenljivama odgovara proporcionalnoj razlici hazarda. Međutim, faktor proporcionalnosti sada varira sa vremenom preživljavanja, umesto da je konstantan.

Funkcija preživljavanja je oblika

$$S(t, \mathbf{X}, \mathbf{Y_t}) = e^{-\int_0^t h(u)du} = e^{-\int_0^t h_0(u)e^{\mathbf{X}\boldsymbol{\beta}^{\boldsymbol{T}}+\mathbf{Y_t}\boldsymbol{\beta}_{\boldsymbol{Y}}^{\boldsymbol{T}}}du} \tag{4.14}$$

Funkcija preživljvanja se više ne može jednostavno faktorisati, kao u slučaju kada su nezavisne promenljive bile konstantne. Međutim, može se znatno olakšati dalji rad ako pretpostavimo da je svaka nezavisna promenljiva konstantna u određenom i definisanom vremenskom intervalu.

Ako pretpostavimo da postoji jedna vremenski zavisna promenljiva Y koja uzima dve vrednosti u zavisnosti da li je vreme preživljavanja pre ili posle nekog datuma, tj.

$$\begin{aligned} Y &= Y_1 \quad \text{ako je} \quad t < s \\ Y &= Y_2 \quad \text{ako je} \quad t \geq s \end{aligned}$$

Funkcija preživljavanja je oblika

$$\begin{aligned} S(t, \mathbf{X}, \mathbf{Y_t}) &= e^{-\int_0^s h_0(u)e^{\mathbf{X}\boldsymbol{\beta}^{\boldsymbol{T}}+\mathbf{Y_1}\boldsymbol{\beta}_{\boldsymbol{Y}}^{\boldsymbol{T}}}du-\int_s^t h_0(u)e^{\mathbf{X}\boldsymbol{\beta}^{\boldsymbol{T}}+\mathbf{Y_2}\boldsymbol{\beta}_{\boldsymbol{Y}}^{\boldsymbol{T}}}du} \\ &= e^{-\lambda_1\lambda\int_0^s h_0(u)du-\lambda_2\lambda\int_s^t h_0(u)du} \\ &= e^{-\lambda_1\lambda\int_0^s h_0(u)du}e^{-\lambda_2\lambda\int_s^t h_0(u)du} \\ &= [S_0(s)]^{\lambda_1\lambda}\frac{[S_0(t)]^{\lambda_2\lambda}}{[S_0(s)]^{\lambda_2\lambda}} \\ &= [S_0(s)]^{\lambda_1\lambda}[\frac{S_0(t)}{S_0(s)}]^{\lambda_2\lambda} \end{aligned}$$

pri čemu je $\lambda = e^{\mathbf{X}\boldsymbol{\beta}^{\boldsymbol{T}}}$; $\lambda_1 = e^{\mathbf{Y_1}\boldsymbol{\beta}_{\boldsymbol{Y}}^{\boldsymbol{T}}}$ i $\lambda_2 = e^{\mathbf{Y_2}\boldsymbol{\beta}_{\boldsymbol{Y}}^{\boldsymbol{T}}}$.

Zaključujemo da je verovatnoća preživljavanja do momenta t proizvod verovatnoća preživljavanja do vremena s i verovatnoća preživljavanja do t, pod uslovom preživljavanja do s.

4.5 Ocenjivanje parametara metodom maksimalne verodostojnosti

Za ocenjivanje parametara u analizi preživljavanja najčešće se koristi metoda maksimalne verodostojnosti. Potrebno je napomenuti da metoda najmanjih kvadrata nije pogodna iz razloga što ne podržava cenzurisanje niti vremenski zavisne promenljive.

Podsetimo se forme funkcije verodostojnosti

$$L = \prod_{i=1}^{n} L_i \qquad \Longleftrightarrow \qquad \log L = \sum_{i=1}^{n} \log L_i$$

Ako imamo uzorak koji se sastoji od

- subjekata kod kojih se desio događaj u posmatranom periodu, pri čemu je $j = 1, ..., J$ gde je t_j takvo da je $t_j \leq t^*$: $L_j = f(t_j)$;
- subjekata koji su cenzurisani iz bilo kog razloga (ili se nije pojavio posmatran događaj ili je klijent izašao iz analize), pri čemu je $k = 1, ..., K$ za t_k.

Sledi

$$L = \prod_{j=1}^{J} f(t_j) \prod_{k=1}^{K} S(t_k)$$

Ako želimo da napišemo funkciju maksimalne verodostojnosti koristeći funkciju rizika, imamo

$$
\begin{aligned}
\log L &= \sum_{j=1}^{J} \log f(t_j) + \sum_{k=1}^{K} \log S(t_k) \\
&= \sum_{j=1}^{J} \log[(\frac{f(t_j}{S(t_j}) S(t_j]) + \sum_{k=1}^{K} \log S(t_k) \\
&= \sum_{j=1}^{J} \log[h(t_j] + \sum_{j=1}^{J} \log[S(t_j] + \sum_{k=1}^{K} \log S(t_k) \\
&= \sum_{j=1}^{J} \log[h(t_j] + \sum_{i=1}^{N} \log[S(t_i] \\
&= \sum_{i=1}^{N} \delta_i \log[h(t_i] + \log[S(t_i]
\end{aligned}
$$

gde je δ_i indikator promenljiva koja takođe prestavlja i status cenzurisanja

$$
\delta_i = \begin{cases} 1 & \text{nije cenzurisano posmatranje} \\ 0 & \text{cenzurisno posmatranje} \end{cases}
$$

Pozivajući se na jednačine (4.6) i (4.7) dobijamo

$$
\log L_i = \delta_i \log h(t_i) - H(t_i) = \delta_i \log h(t_i) - \int_0^{t_i} h(u)du
$$

Obzirom da je logaritamska funkcija monotona, ponekad je lakše naći ekstremnu vrednost funkcije rešavajući sistem jednačina

$$
\frac{\partial \log L(\boldsymbol{\beta}_i)}{\partial \boldsymbol{\beta}_i} \equiv 0, \qquad\qquad i = 1, ..., N
$$

Veoma bitno je napomenuti da se model ne može oceniti ako su sva vremena preživljavanja cenzurisana ($\delta_i = 0, \forall i$).

4.6 Ocenjivanje parametara metodom maksimalne verodostojnosti u uslovima vremenski zavisnih parametara

Pretpostavili smo da je vreme preživljavanja neprekidno. Uključujući vremenski zavisne promenljive potrebno je stvoriti vremenske intervale u kojima su takve promenljive konstantne.

Pretpostavimo da subjekat i uzima sledeće vrednosti za vremenski zavisnu promenljivu Y

$$\begin{aligned} Y &= Y_1 \quad \text{ako je} \quad t < s \\ Y &= Y_2 \quad \text{ako je} \quad t \geq s \end{aligned}$$

Tabelarni prikaz je sledeći

Subjekat	**Indikator cenzurisanja**	**Vreme preživljavanja**	**Vreme ulaska**	**Vrednost vremenski zavisne promenljive**
1	$\delta_i = 0/1$	t_i	0	$-$

Tabela 2. Originalni podaci

Subjekat	**Indikator cenzurisanja**	**Vreme preživljavanja**	**Vreme ulaska**	**Vrednost vremenski zavisne promenljive**
1	$\delta_i = 0$	u	0	Y_1
2	$\delta_i = 0/1$	t_i	u	Y_2

Tabela 3. Modifikovani podaci

Na ovaj način podaci su višestruko prikazani, ali vremenski zavisne promenljive su sada konstantne.

Logaritam funkcije verodostojnosti smo zapisali u obliku

$$\log L_i = \delta_i \log[h(t_i)] + \log[S(t_i)]$$

Ako postoji promena u vremenskom trenutku u, imamo

$$\log[S(t_i)] = \log[S(u)\frac{S(t_i)}{S(u)}] = \log[S(u)] + \log[\frac{S(t_i)}{S(u)}]$$

Sledi da logaritam verovatnoće preživljavanja do t jednak zbiru logaritma verovatnoće preživljavanja do trenutka u i logaritam verovatnoće preživljavanja do t_i, pod uslovom da je preživeo do u.

Na ovaj način stvaramo nov podatak gde je $\delta_i = 0$; $t = u$ i još jedan podatak sa odloženim ulaskom u momentu u i indikatorom cenzurisanja δ_i originalnog podatka. Vremeski zavisna promenljiva kod prvog podatka uzima vrednost Y_1 a kod drugog podatka uzima vrednost Y_2.

5

Koksov PH model

Posebno mesto u klasi statističkih modela preživljavanja imaju modeli sa proporcionalnim rizikom. Kao što smo naveli, modele preživljavanja analiziramo posmatrajući dve fundamentalne stavke, a to su osnovna funkcija rizika i efekat parametara. Osnovna funkcija rizika opisuje kako se menja rizik tokom vremena dok efekat parametara opisuje kako rizik varira u odnosu na nezavisne promenljive. Koksov model sa proporcionalnim rizikom pretpostavlja da je rizik proporcionalan te je moguće oceniti efekat parametara bez određivanja same funkcionalne forme rizika.

Koksova proučavanja iz 1972 godine su promenila pristup standardnoj parametarskoj analizi preživljavanja i proširila metod neparametarskih Kaplan Mejerovih ocena na argumente oblika regresije za analizu životnih tablica. Koks je unapredio predviđanje vremena preživljavanja subjekta bez pretpostavki o osnovnoj funkciji rizika subjekta ali pretpostavljajući da funkcija rizika različitih subjekata ostaje proporcionalna i konstantna tokom vremena.

Ključni razlog za popularnost Koksovog modela leži u činjenici da iako je funkcija osnovnog rizika neodređena mogu se izvesti dobre ocene koeficijenata regresije, hazard količnika i prilagođene krive preživljavanja za širok spektar podataka. Koksov model je stabilan model. Rezultati dobijeni upotrebom Koksovog modela su veoma približni rezultatima tačnog parametarskog modela. To znači da se Koksovim modelom, uz minimum pretpostavki mogu dobiti primarne informacije o osnovnim funkcijama analize preživljavanja.

Još jedna bitna činjenica zbog koje je Koksov model popularan je upravo to što on ima prioritet nad logističkom regresijom kada imamo informacije o vremenu preživljavanja i kada postoji cenzurisanje. Koksov model koristi

više informacija nego logistički model koji ignoriše vreme preživljavanja i cenzurisanje.

Osnovna pretpostavka Koksovog modela je mogućnost ocene veza između stope rizika i nezavisnih promenljivih bez uvođenja pretpostavki o obliku bazne funkcije rizika. Upravo je to razlog, zažto se Koksov model smatra i semi-parametarskim modelom.

Faktički, proz poglavlje 4.3 smo uveli funkciju rizika ako pri tome važi PH pretpostavka, što suštinski definiše Koksov model. U poglavlju 4.3 smo definisali funkciju rizika, funkciju preživljavanja i veze među tim funkcijama. Podsetimo se

$$h(t, \mathbf{X}) = h_0(t)e^{\mathbf{X}\boldsymbol{\beta}^T}$$

$$S(t, \mathbf{X}) = e^{-\int_0^t h(u)du} = e^{-\int_0^t h_0(u)e^{\mathbf{X}\boldsymbol{\beta}^T}du} = S_0(t)^{e^{\mathbf{X}\boldsymbol{\beta}^T}}$$

$$H(t) = -e^{\mathbf{X}\boldsymbol{\beta}^T} \ln S_0(t)$$

Osnovna razlika se pak ogleda u primenjenoj metodi za ocenu koeficijenata u modelu, sa čim ćemo se upoznati u nastavku.

5.1 Metod parcijalne verodostojnosti

Pretpostavimo da je vreme preživljavanja T neprekidno, međutim prilikom razvoja modela, istraživačima je tačno poznato vreme preživljavanja pojedinog subjekta i označimo ga sa t. Funkciju verodostojnosti možemo zapisati i na sledeći način ako pretpostavimo da je sa X_i označen vektor karakteristika za i-tog subjekta, odnosno $\mathbf{X_i} = [X_{i1}, X_{i2}, ..., X_{ip}]^T$, pri čemu važi $i = 1, ..., n$.

$$L(\boldsymbol{\beta}) = \prod_{i=1}^{n} f(t_i, \boldsymbol{\beta}, \mathbf{X_i})^{\delta_i} S(t_i, \boldsymbol{\beta}, \mathbf{X_i})^{1-\delta_i} \tag{5.1}$$

Ako logaritmujemo jednačinu (5.1), dobijamo

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} [\delta_i \ln f(t_i, \boldsymbol{\beta}, \mathbf{X_i}) + (1 - \delta_i) \ln S(t_i, \boldsymbol{\beta}, \mathbf{X_i})] \tag{5.2}$$

Kako je ova funkcija monotona, maksimum se postiže kada prvi izvod izjednačimo sa nulom.

$$\frac{\partial \ln L(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \equiv 0$$

Funkcija gustine se može zapisati kao proizvod funkcije rizika i funkcije preživljavanja, $f(t_i, \boldsymbol{\beta}, \mathbf{X_i}) = h(t_i, \boldsymbol{\beta}, \mathbf{X_i}) S(t_i, \boldsymbol{\beta}, \mathbf{X_i})$, stoga dobijamo

$$\ln L(\boldsymbol{\beta}) = \sum_{i=1}^{n} \delta_i \ln h_0(t_i, \boldsymbol{\beta}, \mathbf{X_i}) + \delta_i \mathbf{X_i} \boldsymbol{\beta}^T + e^{\mathbf{X_i}\boldsymbol{\beta}^T} \ln S_0(t_i, \boldsymbol{\beta}, \mathbf{X_i}) \tag{5.3}$$

jer je poznato da je $h(t_i, \boldsymbol{\beta}, \mathbf{X_i}) = h_0(t_i) e^{\mathbf{X_i}\boldsymbol{\beta}^T}$ i $S(t_i, \boldsymbol{\beta}, \mathbf{X_i}) = S_0(t_i)^{e^{\mathbf{X_i}\boldsymbol{\beta}^T}}$.

Metod maksimalne verodostojnosti zahteva da se maksimizira jednačina (5.3) u odnosu na koeficijente $\boldsymbol{\beta}$, nepoznatu baznu funkciju rizika i preživljavanja. Pozivajući se na radove Kalbfleisch i Prentice (2002), nije moguće maksimizirati jednačinu (5.3).

Koks je predložio alternativni način, u cilju prevazilaženja ovog problema, koji je nazvao metod parcijalne verodostojnosti. Funkcijom parcijalne verodostojnosti se ocenjuju samo $\boldsymbol{\beta}$ koeficijenti, ali Koks je tvrdio da se ovom metodom dobijaju iste osobine raspodele kao što bi se dobilo funkcijom maksimalne verodostojnosti, što je i potvrđeno kasnijih godina.

Funkcija parcijalne verodostojnosti je data sledećom jednačinom

$$L_p(\boldsymbol{\beta}) = \prod_{i=1}^{n} \left[\frac{e^{\mathbf{X}_i \boldsymbol{\beta}^T}}{\sum_{j \in R(t_i)} e^{\mathbf{X}_j \boldsymbol{\beta}^T}} \right]^{\boldsymbol{\delta}_i} \tag{5.4}$$

pri čemu je $R(t_i)$ skup svih subjekata čije je vreme preživljavanja ili cenzurisanja veće ili jednako od specifično zadatog vremena. Logaritam parcijalne funkcije verodostojnosti je

$$\ln L_p(\boldsymbol{\beta}) = \sum_{i=1}^{n} \boldsymbol{\delta}_i \mathbf{X_i} \boldsymbol{\beta}^T - \sum_{i=1}^{n} \boldsymbol{\delta}_i \ln \sum_{j \in R(t_i)} e^{\mathbf{X}_j \boldsymbol{\beta}^T} \tag{5.5}$$

Vektor efikasnih rezultata u oznaci $U(\boldsymbol{\beta})$ dobijamo na sledeći način

$$U(\boldsymbol{\beta}) = \frac{\partial \ln L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \equiv 0$$

Odnosno

$$\begin{aligned} U(\boldsymbol{\beta}) = \frac{\partial \ln L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} &= \boldsymbol{\delta}^T \mathbf{X} - \sum_{i=1}^{n} \frac{\delta_i \sum_{j \in R_{(t_i)}} \mathbf{X_j} e^{\mathbf{X_j}\boldsymbol{\beta}^T}}{\sum_{j \in R_{(t_i)}} e^{\mathbf{X_j}\boldsymbol{\beta}^T}} \\ &= \boldsymbol{\delta}^T \mathbf{X} - \sum_{i=1}^{n} \delta_i \sum_{j \in R_{(t_i)}} w_{ij}(\boldsymbol{\beta}) \mathbf{X_j} = \boldsymbol{\delta}^T \mathbf{X} - \sum_{i=1}^{n} \delta_i \overline{\mathbf{X}}_{w_i} \end{aligned} \tag{5.6}$$

pri čemu je $\mathbf{X}$- matrica nezavisnih promenljivih dimenzija $(n \times p)$; $\boldsymbol{\delta}^T$ - vektor statusne promenljive; $\mathbf{X}_{(\mathbf{j},\cdot)} = \mathbf{X_j}$ - j-ti red u matrici X koji sadrži nezavisne promenljive j-tog subjekta.

Radi pojednostavljivanja izraza uzeli smo da je

$$w_{ij}(\boldsymbol{\beta}) = \frac{e^{\mathbf{X_j}\boldsymbol{\beta}^T}}{\sum_{l \in R_{(t_i)}} e^{\mathbf{X_l}\boldsymbol{\beta}^T}} \tag{5.7}$$

$$\overline{\mathbf{X}}_{w_i} = \sum_{j \in R_{(t_i)}} w_{ij}(\boldsymbol{\beta}) \mathbf{X_j} \tag{5.8}$$

5.1.1 Različita vremena preživljavanja

U zavisnosti od tipa vremena preživljavanja, razlikujemo dva slučaja.

1. Kod svih posmatranih subjekata u uzorku su različita vremena preživljavanja;

2. Postoje subjekti u uzorku koji imaju ista vremena preživljvanja.

Ako pretpostavimo da su sva vremena preživljvanja različita i iz originalnog uzorka izbacimo cenzurisane subjekte, tada jednačinu (5.4) zapisujemo

$$L_p(\boldsymbol{\beta}) = \prod_{i=1}^{n} \frac{e^{\mathbf{X}_{(i)}\boldsymbol{\beta}^T}}{\sum_{j \in R(t_{(i)})} e^{\mathbf{X}_j\boldsymbol{\beta}^T}} \tag{5.9}$$

Na ovaj način dobijamo proizvod m različitih vremena preživljvanja, $t_1 < t_2 < ... < t_n$, pri čemu $X_{(i)}$ označava vektor karakteristika za subjketa sa vremenom preživljavanja $t_{(i)}$.

Logaritam parcijalne verodostojnosti je

$$\ln L_p(\boldsymbol{\beta}) = \sum_{i=1}^{m} \left(\mathbf{X_{(i)}} \boldsymbol{\beta}^T - \ln \left(\sum_{j \in R(t_{(i)})} e^{\mathbf{X}_j \boldsymbol{\beta}^T} \right) \right) \tag{5.10}$$

Ekstremna vrednost parcijalne funkcije verodostojnosti se postiže parcijalnim diferenciranjem po komponentama vektora $\boldsymbol{\beta}$ i izjednačavanjem sa 0. Odnosno

$$\frac{\partial \ln L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k} = \sum_{i=1}^{m} \left(\mathbf{X_{(ik)}} - \frac{\sum_{j \in R_{(t_{(i)})}} \mathbf{X_{jk}} e^{\mathbf{X_j} \boldsymbol{\beta}^T}}{\sum_{j \in R_{(t_{(i)})}} e^{\mathbf{X_j} \boldsymbol{\beta}^T}} \right) = \sum_{i=1}^{m} \left(\mathbf{X_{(ik)}} - \overline{\mathbf{X}}_{w_i k} \right) \tag{5.11}$$

gde je $w_{ij}(\boldsymbol{\beta}) = \frac{e^{\mathbf{X_j} \boldsymbol{\beta}^T}}{\sum_{l \in R_{(t_{(i)})}} e^{\mathbf{X_l} \boldsymbol{\beta}^T}}$ i $\overline{\mathbf{X}}_{w_i k} = \sum_{j \in R_{(t_{(i)})}} w_{ij}(\boldsymbol{\beta}) \mathbf{X_{jk}}$.

Za rešavanje ovog sistema nelinearnih jednačina se najčešće koristi Newton - Raphson algoritam i tada se dobijaju ocene koeficijenata $\boldsymbol{\beta}$, u oznaci $\widehat{\boldsymbol{\beta}}$. Maksimum je postignut ako je matrica drugih parcijalnih izvoda negativno definitna.

Ocenu disperzije za ocenjene koeficijente koji ujedno i predstavljaju i elemente na glavnoj dijagonali matrice informacija $\mathbf{I}(\boldsymbol{\beta}) = -\frac{\partial^2 \ln L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}^2}$ se dobijaju na sledeći način

$$\begin{aligned} \frac{\partial^2 \ln L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k^2} &= -\sum_{i=1}^{m} \left(\frac{\left[\sum_{j \in R_{(t_{(i)})}} \mathbf{X_{jk}^2} e^{\mathbf{X_j} \boldsymbol{\beta}^T} \right] \left[\sum_{j \in R_{(t_{(i)})}} e^{\mathbf{X_j} \boldsymbol{\beta}^T} \right] - \left[\sum_{j \in R_{(t_{(i)})}} \mathbf{X_{jk}} e^{\mathbf{X_j} \boldsymbol{\beta}^T} \right]^2}{\left[\sum_{j \in R_{(t_{(i)})}} e^{\mathbf{X_j} \boldsymbol{\beta}^T} \right]^2} \right) \\ &= -\sum_{i=1}^{m} \sum_{j \in R(t_{(i)})} w_{ij}(\boldsymbol{\beta}) \left[X_{jk} - \overline{X}_{w_i k} \right]^2 \end{aligned} \tag{5.12}$$

[2]matrica informacija se naziva u literaturi i Fisher-ova matrica podataka ili Hesijan

Elementi matrice informacija van glavne dijagonale se definišu

$$\frac{\partial^2 \ln L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_k^2} = -\sum_{i=1}^{m} \sum_{j \in R(t_{(i)})} w_{ij}(\boldsymbol{\beta}) \left[X_{jk} - \overline{X}_{w_i k}\right] \left[X_{jl} - \overline{X}_{w_i l}\right] \tag{5.13}$$

Ocena disperzije i ocena standardne greške ocenjenih koeficijenata je data u nastavku

$$\widehat{Var}(\widehat{\boldsymbol{\beta}}) = \mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1}$$

$$\widehat{SE}(\widehat{\boldsymbol{\beta}}) = \sqrt{\widehat{Var}(\widehat{\boldsymbol{\beta}})} = \sqrt{\mathbf{I}(\widehat{\boldsymbol{\beta}})^{-1}}$$

5.1.2 Ista vremena preživljavanja

U praksi, veoma retko se dešava da su vremena preživljavanja kod svih posmatranih subjekata različita. Stoga je bilo neophodno modifikovati parcijalnu funkciju verodostojnosti. Dve najpoznatije aproksimacije su date od strane Breslow-a (1974) i Efron-a (1977).

Breslow-a aproksimacija

Pretpostavimo da postoji d_i istih vremena preživljavanja u i-tom različitom vremenu preživljavanja, tada važi

$$L_p(\boldsymbol{\beta}) = \prod_{i=1}^{m} L_{p_i}(\boldsymbol{\beta}) = \prod_{i=1}^{m} \frac{e^{\sum_{j \in D_{(t_{(i)})}} \mathbf{X_j} \boldsymbol{\beta}^T}}{\left[\sum_{j \in R(t_{(i)})} e^{\mathbf{X_j} \boldsymbol{\beta}^T}\right]^{d_i}} \tag{5.14}$$

gde je sa $D_{t_{(i)}}$ označen skup subjekata sa vremenom preživljavanja $t_{(i)}$.

Ako postoji značajna razlika u skupovima $D_{(t_{(i)})}$ i $R_{(t_{(i)})}$, Breslow-a aproksimacija daje ocene koje su veoma precizne. Međutim, ako to nije slučaj, koristi se Efron-ova aproksimacija.

Efron-ova aproksimacija

Neka važe pretpostavke ranije uvedene za $D_{(t_{(i)})}$ i $R_{(t_{(i)})}$ kao i d_i. Efron-ova aproksimacija je data sa

$$\begin{aligned} L_p(\boldsymbol{\beta}) &= \prod_{i=1}^{m} L_{p_i}(\boldsymbol{\beta}) \\ &= \prod_{i=1}^{m} \frac{e^{\sum_{j \in D_{(t_{(i)})}} \mathbf{X_j}\boldsymbol{\beta}^T}}{\prod_{k=1}^{d_i} \left[\sum_{j \in R(t_{(i)})} e^{\mathbf{X_j}\boldsymbol{\beta}^T} - \frac{k-1}{d_i} \sum_{j \in D(t_{(i)})} e^{\mathbf{X_j}\boldsymbol{\beta}^T} \right]} \end{aligned} \tag{5.15}$$

Primetimo da za $d_i = 1$ dobijamo da su (5.14) i (5.15) identične sa (5.9).

5.2 Newton - Raphson algoritam

U zavisnosti da li imamo različita vremena preživljavanja, koristimo odgovarajuću $L_p(\boldsymbol{\beta})$. Prilikom određivanja ocena koeficijenata $\boldsymbol{\beta}$, u ozaci $\widehat{\boldsymbol{\beta}}$, za maksimizaciju funkcije parcijalne verodostojnosti ćemo koristiti Newton-Raphson algoritam. Newton-Raphson algoritam predstavlja iterativni postupak za rešavanje nelinearnih jednačina.

Neka je $\mathbf{u}^T = \left(\frac{\partial L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_1}, \frac{\partial L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_2}, ..., \frac{\partial L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_p} \right)$, a sa $\mathbf{I}$ smo označili Hesijan matricu koja ima sledeće elemente

$$I_{ab} = \frac{\partial^2 L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}_a \partial \boldsymbol{\beta}_b}.$$

Neka su sa $\mathbf{u^{(s)}}$ i $\mathbf{I^{(s)}}$ označeni $\mathbf{u}$ i $\mathbf{I}$ u $\boldsymbol{\beta}^{(s)}$, pri čemu je $\boldsymbol{\beta})^{(s)}$ s-ti pokušaj za aproksimaciju $\widehat{\boldsymbol{\beta}}$. Korak s iterativnog procesa ($s = 0, 1, 2, ...$) aproksimira $L_p(\boldsymbol{\beta})$ u blizini $\boldsymbol{\beta}^{(s)}$ Tejlorovim polinomom drugog reda što se može zapisati na sledeći način

$$L_p(\boldsymbol{\beta}) \approx L_p(\boldsymbol{\beta}^{(s)}) + \mathbf{u^{(s)^T}}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)}) + \frac{1}{2}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)})^T \mathbf{I^{(s)}}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)})$$

Rešavajući

$$\frac{\partial L_p(\boldsymbol{\beta})}{\partial \boldsymbol{\beta}} \approx \mathbf{u^{(s)}} + \mathbf{I^{(s)}}(\boldsymbol{\beta} - \boldsymbol{\beta}^{(s)}) = 0$$

po $\boldsymbol{\beta}$ dolazimo do

$$\boldsymbol{\beta}^{(s+1)} = \boldsymbol{\beta}^{(s)} - \left(\mathbf{I^{(s)}}\right)^{-1} \mathbf{u^{(s)}} \tag{5.16}$$

pri čemu važe pretpostavke da $\mathbf{I^{(s)}}$ nije singularna matrica.

Nakon izračunavanja iteracija $s = 0, 1, 2, ...$ neophodno je utvrditi kriterijum konvergencije. Za Newton-Raphson metod postoje tri kriterijuma konvergencije[1]

1. Najveća razlika između apsolutne vrednosti odnosa aproksimacije parametara u dve uzastopne iteracije;
2. Apsolutna razlika logaritma funkcije verodostojnosti između dve uzastopne iteracije podeljena sa logaritmnom funkcije verodostojnosti iz prethodne iteracije;
3. Maksimalni broj iteracija.

Nakon odabira kriterijuma konvergencije dobija se željena aproksimacija koja predstavlja ocenu parametra $\boldsymbol{\beta}$, u oznaci $\widehat{\boldsymbol{\beta}}$.

Napomenimo samo da za ocenu $\widehat{\boldsymbol{\beta}}$ je poznato da prati asimptotsku p-dimenzionalnu normalnu distribuciju, tj.

$$\sqrt{\mathbf{I}(\boldsymbol{\beta})}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \underset{n\to\infty}{\longrightarrow} \mathcal{N}(0, 1)$$

Inverz matrice informacija, $\mathbf{I^{-1}}(\widehat{\boldsymbol{\beta}})$ je konzistentna ocena matrice kovarijanse za $\widehat{\boldsymbol{\beta}}$. Može se koristiti za konstruisanje intervala poverenja za komponente $\boldsymbol{\beta}$.

[1]IBM SPSS Statistics 20 Algorithms; IBM Corporation 1989, 2011.

5.3 Ocena funkcije hazarda i funkcije preživljavanja

Pod PH pretpostavkom smo definisali funkciju preživljvanja na sledeći način

$$S(t, \mathbf{X}, \boldsymbol{\beta}) = [S_0(t)]^{e^{\mathbf{X}\boldsymbol{\beta}^T}} \tag{5.17}$$

Nakon ocene koeficijenata $\boldsymbol{\beta}$ neophodno je oceniti baznu funkciju preživljavanja, preko koje ćemo oceniti funkciju preživljavanja i funkciju rizika.

Definišimo uslovnu baznu funkciju preživljavanja sa

$$\alpha_{0i} = \frac{S_0(t_{(i)})}{S_0(t_{(i-1)})} \tag{5.18}$$

pri čemu je $t_{(i)}$ vreme preživljavanja definisano ranije.

Verovatnoća preživljavanja je

$$\begin{aligned} \frac{S(t_{(i)}, \mathbf{X}, \boldsymbol{\beta})}{S(t_{(i-1)}, \mathbf{X}, \boldsymbol{\beta})} &= \frac{\left[S_0(t_{(i)})\right]^{e^{\mathbf{X}\boldsymbol{\beta}^T}}}{\left[S_0(t_{(i-1)})\right]^{e^{\mathbf{X}\boldsymbol{\beta}^T}}} \\ &= \left[\frac{S_0(t_{(i)})}{S_0(t_{(i-1)})}\right]^{e^{\mathbf{X}\boldsymbol{\beta}^T}} \\ &= \alpha_{0i}^{e^{\mathbf{X}\boldsymbol{\beta}^T}} \end{aligned} \tag{5.19}$$

Ako iskoristimo ranije uvedenu oznaku $\widehat{\lambda}_l = e^{\mathbf{X_l}\boldsymbol{\beta}^T}$, ocena uslovne bazne verovatnoće preživljvanja se dobija rešavanjem jednačine

$$\sum_{l \in D_{(t_{(i)})}} \frac{\widehat{\lambda}_l}{1 - \alpha_{0i}\widehat{\lambda}_l} = \sum_{l \in R_{(t_{(i)})}} \widehat{\lambda}_l \tag{5.20}$$

gde je sa $R_{(t_{(i)})}$ označen skup rizičnih subjekata u trenutku $t_{(i)}$ a $D_{(t_{(i)})}$ skup klijenata čije je vreme preživljavanja $t_{(i)}$.

Ako su u uzorku vremena preživljavanja kod svih subjekata različita tada skup $D_{(t_{(i)})}$ uvek sadrži tačno jednog subjekta pa jednačinu (5.20) zapisujemo

$$\widehat{\alpha}_{0i} = \left[1 - \frac{\widehat{\lambda}_i}{\sum_{l \in R_{(t_{(i)})}} \widehat{\lambda}_l}\right]^{\widehat{\lambda}_i^{-1}} \tag{5.21}$$

Međutim, ukoliko postoje ista vremena preživljvanja, onda se jednačina (5.19) rešava iterativnim postupkom. Ocena bazne funkcije preživljavanja je proizvod pojedinačnih ocena uslovne bazne verovatnoće preživljavanja, odnosno

$$\widehat{S}_0(t) = \prod_{t_{(i)} \leq t} \widehat{\alpha}_{0i} \tag{5.22}$$

pri čemu je $\widehat{\alpha}_{0i}$ rešenje jednačine (5.20). Breslow je ponudio alternativnu ocenu, ako u jednačini (5.20) aproksimiramo $\widehat{\alpha}_{0i}^{\widehat{\lambda}_l}$ sa $\widehat{\alpha}_{0i}^{\widehat{\lambda}_l} \approx 1 + \widehat{\lambda}_l \ln(\alpha_{0i})$. Rešenje je oblika $\tilde{\alpha_{0i}} = e^{-\frac{d_i}{\sum_{l \in R_{(t_{(i)})}} \widehat{\lambda}_l}}$, gde je d_i broj neuspeha u periodu $t_{(i)}$ i uvrštavanjem u jednačinu (5.22) se dobija ocena bazne funkcije preživljavanja.

Ocena dobijena iterativnim postupkom odgovara Kaplan-Mejerovoj oceni a aproksimativna ocena odgovara Nelson-Aalen oceni.

Bez obzira da li imamo ista ili različita vremena preživljavanja u uzorku, ocenu za baznu funkciju preživljavanja možemo dobiti na gore opisane načine.

Ako iskoristimo poznate ocene bazne funkcije preživljavanja i ocene $\boldsymbol{\beta}$ koeficijenata, u jednačini (5.17), dobijamo ocenu funkcije preživljavanja.

Sledi ocena za baznu funkciju rizika

$$\widehat{h}_0(t_{(i)}) = 1 - \widehat{\alpha}_{0i} \tag{5.23}$$

Pokazalo se da je ovakva ocena veoma nestabilna, pa se pre uzima kumulativna bazna funkcija rizika

$$\widehat{S}_0(t) = e^{-\widehat{H}_0(t)} \qquad \Longrightarrow \qquad \widehat{H}_0(t) = -\ln\left[\widehat{S}_0(t)\right] \tag{5.24}$$

Odnosno, ocena kumulativne funkcije rizika je

$$\widehat{H}(t, \mathbf{X}, \widehat{\boldsymbol{\beta}}) = -\ln\left[\widehat{S}(t, \mathbf{X}, \widehat{\boldsymbol{\beta}})\right] = e^{-\mathbf{X}\widehat{\boldsymbol{\beta}}^T} \ln\left[\widehat{S}_0(t)\right] \tag{5.25}$$

5.4 Selekcioni kriterijumi

Ukoliko postoji velik broj nezavisnih promenljivih koje mogu a i ne moraju biti relevantne za donošenje pretpostavki o vremenu preživljavanja, korisno je imati mogućnost redukcije modela, tako da u njemu ostanu samo promenljive koje nam zaista obezbeđuju važne informacije o vremenu preživljavanja. Nije uvek trivijalno odlučiti koju promenljivu treba ostaviti u modelu. Maksimalni model definišemo kao model koji sadrži sve nezavisne promenljive koje mogu biti pristrasne u modelu i pretpostavimo da ih ima r.

Prilikom definisanja maksimalnog modela, važno je da on sadrži sve nezavisne promenljive koje mogu uticati na vreme preživljavanja, ali se mora paziti da u model ne uđe previše nezavisnih promenljivih u poređenju sa brojem subjekata u uzorku jer nas ovakva situacija dovodi do nepreciznih rezultata. Takođe, što je veći broj nezavisnih promenljivih to je veći rizik da dođe do korelacije među promenljivama. Opšte pravilo je da treba da se uzme u obzir veličina uzorka, te što je manji uzorak manji je i broj nezavisnih promenljivih u maksimalnom modelu. U praksi se obično poštuje pravilo da je ukupan broj subjekata u uzorku $n \geq 5r$.

Nakon definisanja maksimalnog modela, sledeći korak je uporediti dva modela i odrediti koji je od njih bolji. U tu svrhu koristimo selekcione kriterijume, čiji je zadatak da porede maksimalan model sa redukovanim modelom koji proizilazi iz maksimalnog modela. Cilj je utvrditi da li redukovani model odgovara podacima jednako dobro kao maksimalan model i u tom slučaju ćemo se odlučiti da koristimo redukovani model umesto maksimalnog. Sada ćemo navesti par selekcionih kriterijuma.

1. **Cox-Snell reziduali (1968)** se definišu na sledeći način

$$r_{c_i} = \widehat{H}_0(t_{(i)})e^{\mathbf{X_i}\boldsymbol{\beta}^T} = \widehat{H}_i(t_{(i)}) = -\ln \widehat{S}_i(t_{(i)})$$

   gde je $\widehat{H}_0(t_{(i)})$ ocenjena bazna kumulativna funkcija hazarda;$\widehat{H}_i(t_{(i)})$ je ocenjena kumulativna funkcija hazarda za i-tog subjekta u vremenu preživljavanja $t_{(i)}$ i $\widehat{S}_i(t_{(i)})$ je ocenjena funkcija preživljavanja i-tog subjekta u vremenu preživljavanja $t_{(i)}$.

   Collett (1994) je pokazao da $-\ln S(t)$ ima eksponencijalnu raspodelu sa parametrom 1 bez obzira na samu formu funkcije preživljavanja $S(t)$. Ako je model adekvatno redukovan, ocenjena funkcija preživljavanja

$\widehat{S}(t)$ će biti približna $S(t)$ sa sličnim osobinama. Stoga $-\ln \widehat{S}(t_{(i)}) = r_{c_i}$ će činiti skup opservacija sa eksponencijalnom raspodelom sa parametrom 1.

Prilikom testiranja da li reziduali imaju eksponencijalnu raspodelu, Kaplan - Mejerova ocena se izračunava. Ocenjena regresijska linija za $\ln(-\ln \widehat{S}(r_{c_i}))$ i $\ln(r_{c_i})$ prolazi kroz koordinatni početak i ima jedinični nagib što ukazuje da je ocenjen model adekvatan.

2. **Martingalni reziduali (Therneau 1990)** predstavljaju jednu transformaciju Cox-Snell reziduala, tako da važi

$$r_{m_i} = \delta_i - r_{c_i}$$

Martingalni reziduali se mogu interpretirati kao razlika između posmatranog broja neuspeha u intervalu $(0, t_i)$ u oznaci δ_i i očekivan broj neuspeha prema razvijenom modelu. Grafički prikaz u skladu sa rastućim vremenom preživljavanja ne bi trebalo da prikazuje nikakav šablon za r_{m_i}.

3. **Reziduali u odstupanju(Therneau 1990)** se definišu

$$r_{D_i} = sgn(r_{m_i})\,[-2(r_{m_i} + \delta_i \log(\delta_i - r_{m_i})]$$

Reziduali odstupanja predstavljaju transformaciju martingalnih reziduala, pri čemu su grafici lakši za interpretaciju jer su reziduali simetrično raspoređeni oko 0.

Ako je $\mathbf{X_i} = [X_{i1}, X_{i2}, ..., X_{ip}]^T$ vektor karakteristika i-tog subjekta; $R_{t_{(i)}}$ je skup rizičnih klijenata u trenutku $t_{(i)}$, onda **Schoenfeld-ovi reziduali (1982)** u vremenu $t_{(i)}$ se definišu kao vektor $r_i = [r_{i1}, r_{i2}, ..., r_{ip}]$ gde je

$$r_{ik} = \mathbf{X_{ik}} - E(\mathbf{X_{ik}}|R_i).$$

Schoenfeld-ovi reziduali prikazuju razliku između uočenih vektora karakteristika $\mathbf{X_i}$ i njegovih očekivanih vrednosti, pod uslovom da subjekat pripada $R_{(t_{(i)})}$ skupu. Ako je vrednost reziduala veća, verovatnije je da će subjekat doživeti neuspeh u vremenu $t_{(i)}$.

Osnovna razlika ovog reziduala od ostalih je da sadrži vektor karakteristika za svakog subjekta i vrednost za svaku nezavisnu promenljivu.

Schooenfeld-ovi reziduali nam ukazuju između ostalog na postojanje vremenski zavisnih promenljivih i da li je pojedinim nezavisnim promenljivama neophodna transformacija.

Kada smo se upoznali sa selekcionim kriterijumima, navešćemo i selekcione metode koje koriste ove kriterijume da bi obredile da li se u modelu nalazi optimalan broj promenljivih.

$IBM^{@}$ $SPSS^{@}$ *Modeler* softver koji ćemo koristiti u daljem radu, nudi metod u kome ulaze sve dostupne promenljive; metod po etapama unapred i metod po etapama unazad. Metod po etapama unapred počinje praznim modelom i u svakom koraku dodaje promenljivu koja je najznačajnija ili kombinaciju promenljivih i tako sve dok se dođe do situacije da se više nema koristi od dodavanja novih promenljivih. Veoma bitno je da napomenemo da one promenljive koje ulaskom novih promenljivih gube na značajnosti, izlaze iz modela. U praksi se često dešava situacija da promenljiva ili kombinacija promenljivih koje su ušle u model u prethodnom koraku nisu značajne u sledećem i samim tim izlaze iz modela. Ovaj proces se nastavlja sve dok ne dođe do toga da nema više koristi dodavati nove promenljive.
Metod po etapama unazad počinje modelom koji sadrži sve promenljive i u svakom koraku izbacuje promenljivu ili kombinaciju promenljivih koje su najmanje značajne, sve dok ne dođe do toga da nema više koristi izbacivati promenljive.

5.5 Testiranje hipoteza

Postoji nekoliko metoda na osnovu kojih se može oceniti značaj modela i u ovom radu ćemo navesti tri najznačajnije. Međutim, potrebno je uvesti osnovne pretpostavke. U prethodnom poglavlju smo definisali da maksimalan model sadrži r promenljivih. Pretpostavimo da je $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T]^T$, pri čemu je $\boldsymbol{\beta}_1$ q-dimenzionalni podvektor koji sadrži promenljive od interesa a $\boldsymbol{\beta}_2$ je $(r-q)$-dimenzionalni podvektor koji sadrži preostale promenljive. Testiraćemo nultu hipotezu H_0 protiv alternativne H_1 pri čemu su hipoteze definisane na sledeći način

$$\begin{aligned} H_0 : \boldsymbol{\beta}_1 = 0 &\implies \text{redukovani model je značajan} \\ H_1 : \boldsymbol{\beta}_1 \neq 0 &\implies \text{maksimalan model je značajan} \end{aligned}$$

5.5.1 Test količnika verodostojnosti

Test statistika za test količnika verodostojnosti se dobija kao dvostruka razlika logaritma funkcije parcijalne verodostojnosti maksimalnog modela i redukovanog modela. Dvostruka razlika se uzima za postizanje odgovarajuće χ^2 raspodele za nultu hipotezu.

Test statistika za test količnika verodostojnosti je data na sledeći način

$$\widehat{L}_p = -2 \log \frac{\widehat{L}_p(\text{redukovani})}{\widehat{L}_p(\text{maksimalan})} = -2 \left[\log \widehat{L}_p(\text{redukovani}) - \log \widehat{L}_p(\text{maksimalan}) \right] : \chi^2_q$$

Pod pretpostavkom da važi nulta hipoteza, asimptotska raspodela $\widehat{L}_p$ odgovara $\chi^2_q(\alpha)$, pri čemu je α verovatnoća greške prvog reda, odnosno interval poverenja je $100(1-\alpha)\%$.

Traženu p vrednost jednostranog $\chi^2_q(\alpha)$ testa dobijamo izračunavajući

$$P(\chi^2_q \geq \widehat{L}_p).$$

Ukoliko je vrednost izračunate test statistke manja od odgovarajuće tablične vrednosti, prihvata se alternativna hipoteza.

5.5.2 Wald-ov test

Za testiranje hipoteza se može koristiti i Wald-ov test koji se najčešće koristi da pokaže da li efekat postoji ili ne, odnosno da li nezavisna promenljiva ima statistički značajan odnos sa zavisnom promenljivom. Wald-ov test takođe poredi razliku između maksimalnog i redukovanog modela. Razliku ove dve vrednosti aproksimiramo normalnom raspodelom, pa se kvadrat ove raspodele aproksimira χ^2 raspodelom.

Neka je sa $\boldsymbol{\beta} = [\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T]^T$ označena ocena vektora $\boldsymbol{\beta}$ za maksimalan model dobijena funkcijom parcijalne verodostojnosti. Ako matricu informacija $\mathbf{I}(\boldsymbol{\beta})$ podelimo na blokove na sledeći način

$$\mathbf{I}^{-1}(\boldsymbol{\beta}) = \begin{bmatrix} \mathbf{I}^{11} & \mathbf{I}^{12} \\ \mathbf{I}^{21} & \mathbf{I}^{22} \end{bmatrix}$$

pri čemu $\mathbf{I}^{11}$ označava blok $q \times q$ koji odgovara $\boldsymbol{\beta}_1$.

Wald-ovu test statistiku možemo zapisati

$$W = \widehat{\boldsymbol{\beta}}_1^T \left[\mathbf{I}^{11}(\widehat{\boldsymbol{\beta}}) \right]^{-1} \widehat{\boldsymbol{\beta}}_1$$

Pod pretpostavkom da H_0 važi, raspodela W asimptotski konvergira ka χ_q^2.

5.5.3 Skor test

Definisali smo ranije vektor efikasnih rezultata, u oznaci $U(\boldsymbol{\beta})$. Označimo sa $U_1(\boldsymbol{\beta})$ podvektor prvih q elemenata vektora $U(\boldsymbol{\beta})$. Test statistika skor testa je

$$SC = U_1(\widehat{\boldsymbol{\beta}}_{\text{redukovan}})^T \mathbf{I}^{11}(\widehat{\boldsymbol{\beta}}_{\text{redukovan}}) U_1(\widehat{\boldsymbol{\beta}}_{\text{redukovan}})$$

Za velike uzorke, raspodela SC asimptotski konvergira ka χ_q^2 pod pretpostavkom H_0.

6

Razvoj Koksovog PH skoring modela

U numeričkom delu ćemo prikazati razvoj kreditnog skoring modela preko Koksove metode upotrebom softvera $IBM^@$ $SPSS^@$ Modeler. Prvo ćemo razviti model kreditnog skoringa na osnovu kojeg se mogu odobravati plasmani potencijalnim klijentima izračunavajući rizičnost klijenta prilikom otplate tog plasmana narednih $6, 12, 18, 24, 30$ i 36 meseci koristeći Koksov metod. Potom ćemo tako razvijen model uporediti sa standardnim modelom koji se koristi u bankarstvu razvijen preko logističke regresije.

Posmatran uzorak se sastoji od 7358 klijenata kojima je odobren kredit u vremenskom periodu od 01.07.2010. do 01.07.2011. obezbeđen od strane jedne finansijske institucije. Rok otplate posmatranih kredita se kreće od 12 do 300 meseci. Međutim, obezbeđen uzorak prati otplatu klijenata do 01.07.2013., što implicira da je najduži period posmatranja 36 meseci. Indikator promenljiva uzima vrednost 1 ako je klijent dostigao 90 dana kašnjenja što ukazuje da je klijent propustio sa otplatom 3 uzastopna mesečna anuiteta. U daljoj analizi indikator promenljivu nazivamo Default Flag. Stopa defaulta klijenta u uzorku je 8.96%. Spisak promenljivih dostupnik za ovo istraživanje je dato u Tabeli 6.1.

Promenljiva	**Vrsta promenljive**	**Kategorije**
Datum odobravanja kredita	Datum	-
Godine starosti	Neprekidna	-
Pol	Binarna	1 = muško 0 = žensko

Godine stanovanja na datoj adresi	Neprekidna	-
Mesto stanovanja	Kategorijalna	1 = vlastita kuća 2 = vlastiti stan 3 = unajmljen stan
Region	Kategorijalna	1 = Beograd 2 = Niš 3 = Novi Sad
Bračni status	Kategorijalna	1 = oženjen/udata 2 = neoženjen/neudata 3 = razveden/razvedena 4 = udovac/udovica
Broj izdržavanih lica u domaćinstvu	Neprekidna	-
Broj zaposlenih lica u domaćinstvu	Neprekidna	-
Ukupan broj članova u domaćinstvu	Neprekidna	-
Ugovor za mobilni telefon	Binarna	1 = postpaid ugovor 0 = pripaid
Obrazovanje	Kategorijalna	1 = visoka stručna sprema 2 = viša škola 3 = srednja škola 4 = osnovna škola
Tip zaposlenja	Binarna	1 = zaposlen 0 = penzioner
Tip zaposlenja detaljno	Kategorijalna	1 = Menadžeri, lekari, advokati, profesori 2 = službenici, nastavnici 3 = ostalo 4 = penzioner
Radno iskustvo	Neprekidna	-
Prihod pri apliciranju	Neprekidna	-
Iznos kredita	Neprekidna	-
Rok otplate	Neprekidna	-
Mesečni anuitet	Neprekidna	-
Istorija iz Kreditnog Biroa	Kategorijalna	1 = nije problematičan 2 = pred-problematičan 3 = problematičan
Default Flag	Binarna	1 = default 0 = cenzurisanje
Vreme preživljavanja (u mesecima)	Neprekidna	-
Vreme default-a klijenta ili cenzurisanja	Datum	-

Tabela 6.1. Lista dostupnih promenljivih

Radi detaljnijeg upoznavanja sa uzorkom u Tabeli 6.2. je prikazan odnos default i cenzurisanih klijenata u zavisnosti od vremena odobravanja. Na ovaj način je potvrđeno ekonomsko stanovište, da je najveća stopa defaulta u prvoj i drugoj godini nakon odobravanja kredita.

Godina odobravanja	Godina zatvaranja	Default plasmani	Cenzurisani plasmani	Ukupno plasmana	Default Stopa	Stopa preživljavanja
2010	2010	10	39	49	20.41%	79.59%
	2011	212	749	961	22.06%	77.94%
	2012	180	911	1091	16.50%	83.50%
	2013+	47	2961	3008	1.56%	98.44%
	Total	**449**	**4660**	**5109**	**8.79**%	**91.21**%
2011	2011	66	284	350	18.86%	81.14%
	2012	113	669	782	14.45%	85.55%
	2013+	31	1086	1117	2.78%	97.22%
	Total	**210**	**2039**	**2249**	**9.34**%	**90.66**%

Tabela 6.2. Default klijenti u zavisnosti od godine zatvaranja plasmana

Ako posmatramo sumarno, default stopa klijenata čiji su plasmani odobreni u 2010 je 8.79% a u 2011 godini je 9.34%, što zaista verno odražava ekonomsku situaciju na tržištu, koja je bila na udaru krize ovih godina. Iz tog razloga je default stopa viša nego ranijih godina.

Tokom razvoja modela, bilo je neophodno izvršiti određene transformacije i uzeti u obzir i nelinearne efekte nezavisnih promenljivih.

1. korak Nakon inicijalnog upoznavanja sa uzorkom i karakteristikama subjekata, pozabavili smo se outlierima i ekstremnim vrednostima. Standardno skaliranje outliera i ekstremnih vrednosti je na srednjoj vrednosti $+/-$ $3*$ standardna devijacija. Ukoliko postoje podaci koji nedostaju, kako zbog tehničke greške ili u momentu odobravanja određen podatak nije bio dostupan, neophodno je dopuniti uzorak sa adekvatnom vrednošću. Izbor istraživača predstavlja da li će koristiti srednju vrednost, medianu, mod ili neku drugu vrednost koja predstavlja logički izbor i sa ekonomskog i sa matematičkog stanovišta. Kombinovanjem dostupnih promenljivih istraživači obično stvaraju nove promenljive u cilju utvrđivanja najznačajnijih promenljivih.

2. korak Neophodno je sprovesti univarijantu analizu svake nezavisne promenljive u slučaju da pronađemo potencijalnu nekonzistentnost sa ostalim podacima. Posmatrajmo na primer godine klijenta. Može se desiti da je operativnom greškom referenta unesen podatak da klijent ima 136 godina, a u stvari klijent ima 36 godina. Ovo je jednostavan primer koji ocrtava potrebu sprovođenja univarijante analize.

3. korak Ako posmatramo kategorijalne promenljive, potrebno je proveriti koliko se procenatnualno subjekata iz uzorka nalazi u svakoj kategoriji. Nepisano pravilo je da svaka kategoriji sadrži barem 5% od ukupnog uzorka, ali ne više od 90% da bi razvijen model bio adekvatan. Pored ovog pravila, potrebno je izbegavati velik broj kategorija, pa se u praksi obino uzima do 4 kategorije po promenljivoj. Ukoliko ekonomska logika podržava a stope defaulta klijenata su približno iste, problem velikog broja kategorija rešavamo spajanjem određenih kategorija. U praksi, to se postiže upotrebom algoritma *C&R*, *Quest* ili *CHAID*. Isto važi i za neprekidne promenljive. Ako određene nezavisne promenljive neprekidnog tipa pokazuju značajnu nelinearnu vezu sa zavisnom promenljivom potrebno ih je kategorizovati. Novonastale kategorizovane neprekidne promenljive i rekategorizovane kategorijalne promenljive moraju podržavati ekonomsku praksu i ocrtavati realno stanje u suprotnom se odbacuju iz modela.

4. korak Potrebno je ponovo osmotriti sve dostupne promenljive u cilju pronalaženja promenljivih koje su u korelaciji i proveriti očekivanja istraživača. Kada mapiramo takve promenljive, proverava se značajnost svake pojedinačno i najznačajnija ulazi u model. Za izračunavanje značajnosti koristimo *Information Value* metod i *Feaute Selection* koji imaju drugačiji algoritam u pozadini ali daju iste rezultate.

5. korak Nakon pripreme uzorka i inicijalnim upoznavanjem, sprovodimo Koksov metod u cilju dobijanja neophodnih promenljivih koje definišu finalni model za odobravanje kredita.

Praksa nalaže da se uzorak podeli na deo za razvoj modela koji čini 70% od ukupnog broja subjekata i 30% deo za validaciju modela, pod uslovom da je stopa defaulta približno jednaka u obe particije.

Poštujući navedene korake, razvijamo model za kreditni skoring. Koristimo *CHAID* algoritam za kategorizaciju neprekidnih i rekategorizaciju kategorijalnih promenljivih.Za selekciju potencijalno značajnih i irelevantnih

promenljivih koristimo metod *Feature selection*. Irelevantne promenljive odbacujemo pre upotrebe Koksove metode. Kako smo na početku podelili uzorak, Koksov metod se primenjuje na uzorku za razvoj modela, a potom se proverava na delu uzorka za validaciju. Korišćen je metod po etapama unapred. Kriterijumi konvergencije koji su korišćeni:

- maksimalan broj iteracija je 20;
- konvergencija parametara koja predstavlja najveću razliku između apsolutne vrednosti aproksimacije parametra u dve uzastopne iteracije je 0.0001;
- konvergencija logaritma funkcije verodostojnosti koja predstavlja apsolutnu razliku logaritma funkcije verodostojnosti između dve uzastopne iteracije podeljena sa logaritmom funkcije verodostojnosti iz prethodne iteracije je 0.00001.

Pri svakom koraku izgradnje modela po etapama za utvrđivanje značajnosti određene etape koristi se test količnika verodostojnosti. Granice za ulazak i izlazak iz modela su 0.05 i 0.1, redom. Interval poverenja je 95%.

6.1 Opis promenljivih u finalnom modelu

U Tabeli 6.3. su prikazane promenljive koje definišu finalni model. Analiziraćemo pojedinačno svaku.

Promenljive u finalnom modelu i njihove osnovne karakteristike								
	B	SE	Wald	df	Sig	Exp (B)	95.0% CI for Exp(B)	
							Lower	Upper
Radno iskustvo	-.041	.004	83.024	1	.000	.960	.952	.968
V_Pol_Mesto_Boravka_KAT(1)	.471	.100	22.333	1	.000	1.602	1.317	1.947
Godine stanovanja na datoj adresi _KAT(1)	.564	.138	16.759	1	.000	1.757	1.342	2.301
Godine stanovanja na datoj adresi _KAT(2)	.293	.108	7.359	1	.007	1.341	1.085	1.657
Broj zaposlenih lica u domaćinstvu	-.439	.105	17.605	1	.000	.645	.525	.791
Prihod pri apliciranju _KAT(1)	.776	.137	31.901	1	.000	2.173	1.660	2.844
Prihod pri apliciranju _KAT(2)	.356	.136	6.819	1	.009	1.428	1.093	1.866
Rok otplate _KAT(1)	-.834	.135	38.140	1	.000	.434	.333	.566
Rok otplate _KAT(2)	-.299	.128	5.459	1	.019	.741	.577	.953
Mesečni anuitet _KAT(1)	-.756	.208	13.147	1	.000	.470	.312	.707
Mesečni anuitet _KAT(2)	-.254	.111	5.251	1	.022	.776	.625	.964
Istorija iz Kreditnog Biroa_KAT(1)	-.700	.097	51.883	1	.000	.496	.410	.601
Ukupan broj članova u domaćinstvu _KAT(1)	-.572	.120	22.737	1	.000	0.564	0.446	0.714

Tabela 6.3. Promenljive u finalnom modelu i njihove osnovne karakteristike

Radno iskustvo. Radno iskustvo predstavlja neprekidnu promenljivu pri čemu je iskazano radno iskustvo klijenta u godinama. Distribucija je prikazana histogramom, gde su crvenom bojom označeni default klijenti a plavom ukupan broj klijenata.

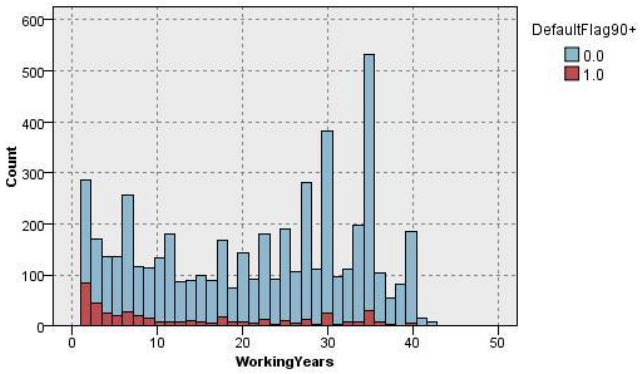


Tabela 6.4. Histogram radnog iskustva klijenta

Osnovne karakteristike promenljive su:

Min	Max	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skewness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.
1	49	48	21.370	0.168	11.999	143.987	-0.202	0.034	-1.288	0.069

Pre upotrebe Koksovog modela izračunata je potencijalna značajnost promenlijve preko *Feature Selection* opcije i ona iznosi 1, što predstavlja dobar indikator da će se promenljiva naći u finalnom modelu. Upotrebom Koksovog modela, pretpostavka je potvrđena. U finalnom modelu, uočavamo da je predznak koeficijenta β negativan, to znači da sa porastom radnog iskustva rizičnost klijenta smanjuje. Ako poredimo sa najrizičnijim subjektom, koji ima najmanje radno iskustvo, zaključujemo da ako se radno iskustvo povećava po jedinici godine tako rizik opada 0.960 puta, odnosno za svaku godinu rizik je manji za $100\% - (0.960 * 100\%) = 4\%$. Ako hoćemo da proverimo za 5 godina, onda se rizik smanjio za $100\% - (0.960^5 * 100\%) = 18.47\%$. Donja i gornja granica 95%-tnog intervala poverenja su 0.952 i 0.968.Kako je vrednost *Wald*-ove test statistike 83.024, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000.

V Pol Mesto Boravka. U Tabeli 6.1. je data lista dostupnih promenljivih. Kombinacijom promenljivih *Pol* i *Mesto Boravka* dobija se promenlijva *V Pol Mesto Boravka* koja ima 6 kategorija. Distribucija je prikazana u nastavku.

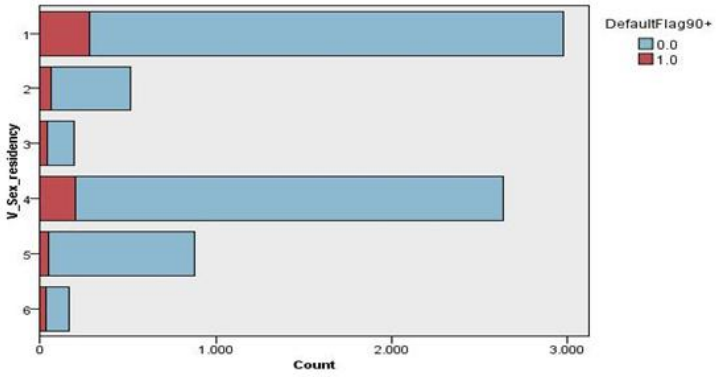


Tabela 6.5. Distribucija V Pol Mesto Boravka

Kategorija 1 označava muškarce u vlastitoj kući (40.49%),
Kategorija 2 označava muškarce u vlastitom stanu (6.99%),
Kategorija 3 označava muškarce u iznajmljenom stanu (2.65%),
Kategorija 4 označava ženu u vlastitoj kući (35.78%),
Kategorija 5 označava ženu u vlastitom stanu (11.91%),
Kategorija 6 označava ženu u iznajmljenom stanu (2.19%).

U Tabeli 6.6. je prikazan udeo cenzurisan i default klijenata po kategoriji.

	Broj subjekata u kategoriji	Izraženo u procentima	Default slučajeva u kategoriji	Izraženo u procentima	Cenzurisanih slučajeva u kategoriji	Izraženo u procentima
Kategorija 1	2979	40.49 %	282	9.47%	2697	90.53%
Kategorija 2	514	6.99%	59	11.48%	455	88.52%
Kategorija 3	195	2.65%	41	21.02%	154	78.98%
Kategorija 4	2633	35.78%	200	7.60%	2433	92.40%
Kategorija 5	876	11.91%	45	5.14%	831	94.86%
Kategorija 6	161	2.19%	32	19.88%	129	80.12%

Tabela 6.6. Raspodela po kategorijama

Primetimo da su stope defaulta klijenata znatno niže kod kategorije 4 i kategorije 5, što i podupire stanovište da su klijenti žene koji žive u vlastitoj kući ili stanu znatno bezbednije od ostalih klijenata. Zapažamo i da postoji problem udela u kategorijama 3 i 6 jer je manji od 5%.

Korišćenjem *CHAID* algoritma spajamo određene kategorije i dobijamo novu promenljivu *V Pol Mesto Boravka KAT*.

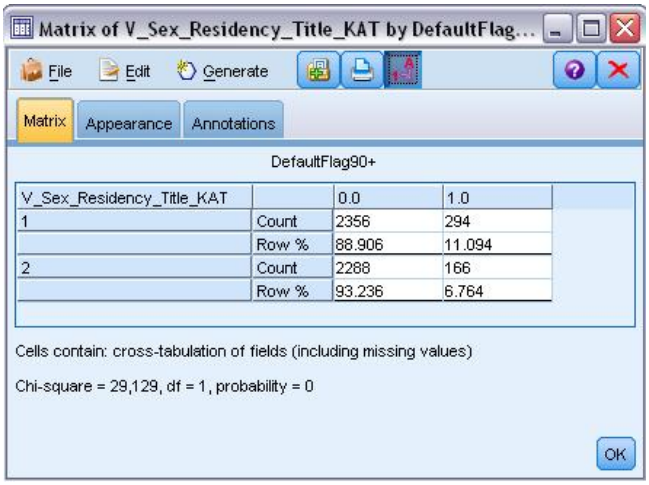


Tabela 6.7. V Pol Mesto Boravka KAT

Nakon spajanja kategorija dobija se *V Pol Mesto Boravka KAT(1)* što predstavlja kategoriju klijenata muškaraca koji žive u vlastitoj kući ili vlastitom stanu ili unajmljenom stanu ili žene koje žive u unajmljenom stanu. Kategorija *V Pol Mesto Boravka KAT(2)* predstavlja žene koje žive u vlastitom stanu ili vlastitoj kući. Kao što se može videti iz Tabele 6.7. udeo klijenata u kategorijama je približno jednak, a stopa defaulta je znatno veća u prvoj kategoriji i čini 11.094%, što opravdava prethodno stanovište da su žene u vlastitoj kući ili stanu znatno bezbedniji klijenti banke sa stopom defaulta od 6.764%.

Pre upotrebe Koksovog modela izračunata je potencijalna značajnost promenljive preko *Feature Selection* opcije i ona iznosi 1 , što ponovo predstavlja dobar indikator da će se promenljiva *V Pol Mesto Boravka KAT* naći u finalnom modelu. Upotrebom Koksovog modela, pretpostavka je potvrđena i promenljiva *V Pol Mesto Boravka KAT(1)* se zaista našla u modelu. Uočavamo da je predznak koeficijenta β pozitivan, znači da je kategorija koja ne obuhvata žene koje žive u vlastitom stanu ili kući znatno rizičnija od kategorije koja ih obuhvata, tačno 1.602 puta. Donja i gornja granica 95%-tnog intervala poverenja su 1.317 i 1.947. Kako je vrednost *Wald*-ove test statistike 22.333, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000.

Godine stanovanja na datoj adresi. Godine stanovanja klijenta na datoj adresi predstavljaju neprekidnu promenljivu pri čemu su iskazane godine boravka klijenta na istoj adresi. Očekujemo da je klijent bezbedniji ukoliko

duže živi na istoj adresi. Distribucija je prikazana histogramom, pri čemu su crvenom bojom označeni default klijenti a plavom ukupan broj klijenata.

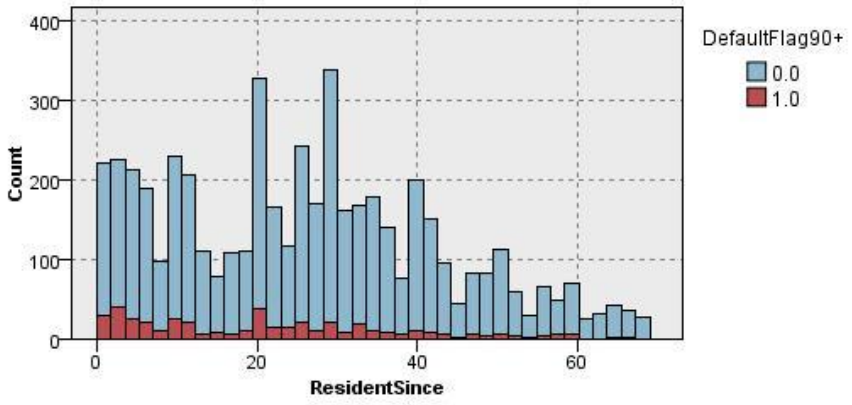


Tabela 6.8. Histogram godine stanovanja na istoj adresi

Osnovne karakteristike promenljive su:

Min	Max	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skewness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.
0	69	36	26.280	0.236	16.882	285.000	0.375	0.034	-0.592	0.069

Kako je promenljiva pokazala značajnu nelinearnu vezu sa zavisnom promenljivom kategorizovali smo je korišćenjem *CHAID* algoritma.

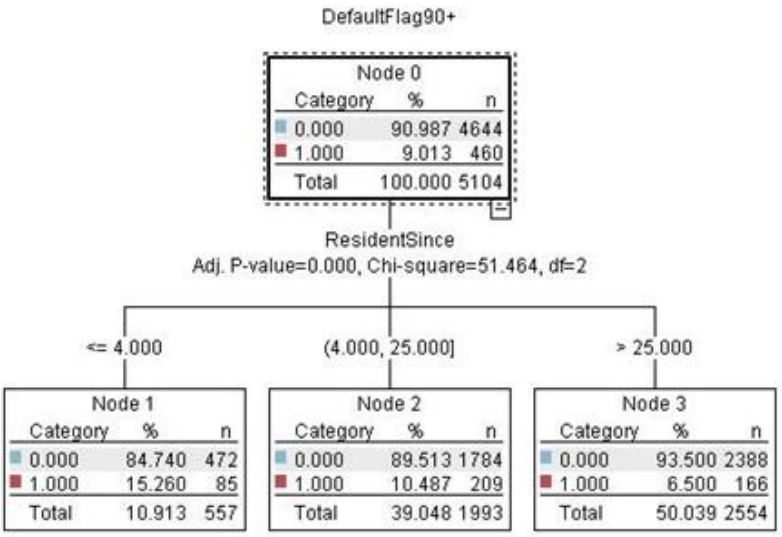


Tabela 6.9. *CHAID* godine stanovanja na istoj adresi

U Tabeli 6.9. je prikazana kategorizovana neprekidna promenljiva. Klijent pripada kategoriji

- *Godine stanovanja na datoj adresi KAT(1)* ako živi na datoj adresi između $[0, 4]$ godine što ujedno predstavlja i najrizičniju kategoriju sa stopom defaulta 15.260%;
- *Godine stanovanja na datoj adresi KAT(2)* ako živi na datoj adresi između $(4, 25]$ godina što ujedno predstavlja i srednje rizičnu kategoriju sa stopom defaulta 10.487%;
- *Godine stanovanja na datoj adresi KAT(3)* ako živi na datoj adresi više od 25 godina što ujedno predstavlja i najmanje rizičnu kategoriju sa stopom defaulta 6.500%.

Pre upotrebe Koksovog modela izračunata je potencijalna značajnost promenljive preko *Feature Selection* opcije i ona iznosi 1 i u neprekidnom i u kategorijalnom obliku. Upotrebom Koksovog modela promenljiva kategorizovana

preko *CHAID* algoritma *Godine stanovanja na datoj adresi KAT* se našla u modelu. Za kategoriju *Godine stanovanja na datoj adresi KAT(1)* uočavamo da je vrednost koeficijenta β najveća a njegov predznak pozitivan što ukazuje da je ova kategorija najrizičnija i to 1.757 puta u odnosu na najbezbedniju kategoriju *Godine stanovanja na datoj adresi KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 1.342 i 2.301. Kako je vrednost *Wald*-ove test statistike 16.759, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000. U finalnom modelu, svoje mesto je zauzela i kategorija *Godine stanovanja na datoj adresi KAT(2)*. Uočavamo da je vrednost koeficijenta β manja u odnosu na kategoriju *Godine stanovanja na datoj adresi KAT(1)* a njegov predznak je opet pozitivan, što znači da je i ova kategorija rizičnija i to 1.341 puta u odnosu na najbezbedniju kategoriju *Godine stanovanja na datoj adresi KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 1.085 i 1.657. Kako je vrednost *Wald*-ove test statistike 7.359, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i $p-vrednost$ je 0.007.

Broj zaposlenih lica u domaćinstvu. Broj zaposlenih lica u domaćinstvu predstavlja neprekidnu promenljivu. Očekujemo da se rizičnost smanjuje pri porastu broja zaposlenih članova u domaćinstvu. Distribucija je prikazana histogramom, pri čemu su crvenom bojom označeni default klijenti a plavom ukupan broj klijenata.

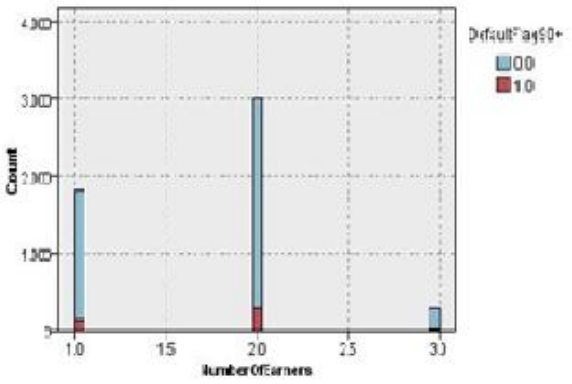


Tabela 6.10. Distribucija broja zaposlenih u domaćinstvu

Osnovne karakteristike promenljive su:

Min	Max	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skewness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.
1	3	2	1.701	0.008	0.567	0.321	0.087	0.034	-0.583	0.069

Feature Selection metod potvrđuje značajnost ove promenljive i ona iznosi 0.972. U finalnom modelu, uočavamo da je predznak koeficijenta negativan, što znači da se porastom broja zaposlenih lica u domaćinstvu rizičnost klijenta smanjuje. Ako se broj zaposlenih u domaćinstvu povećava za jedan, rizik opada 0.645 puta. Donja i gornja granica 95%-tnog intervala poverenja su 0.525 i 0.791. Kako je vrednost *Wald*-ove test statistike 17.605, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000.

Prihod pri apliciranju. Prihod klijenta pri apliciranju za plasman predstavlja neprekidnu promenljivu pri čemu su iskazana prosečna tromesečna primanja klijenta potvrđena od strane poslodavca. Očekujemo da je klijent bezbedniji ukoliko je prihod pri apliciranju veći. Distribucija je prikazana histogramom, gde su crvenom bojom označeni default klijenti a plavom ukupan broj klijenata.

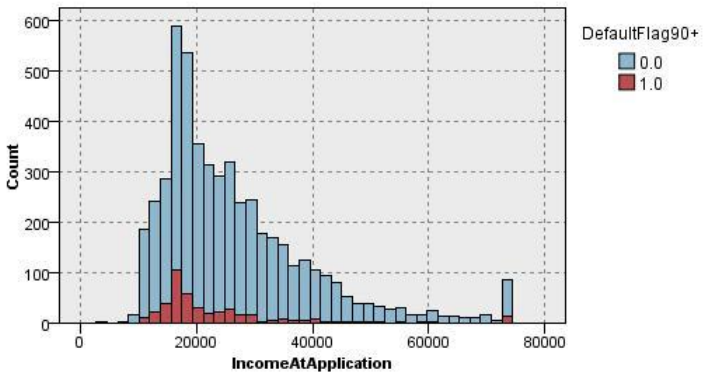


Tabela 6.11. Histogram prihoda pri apliciranju

Osnovne karakteristike promenljive su:

Min	Max	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skewness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.
2800.00	74446.18	71646.18	26705.69	184.44	13176.97	173632732	1.528	0.034	2.508	0.069

Kako je promenljiva pokazala značajnu nelinearnu vezu sa zavisnom promenljivom kategorizovali smo je korišćenjem *CHAID* algoritma.

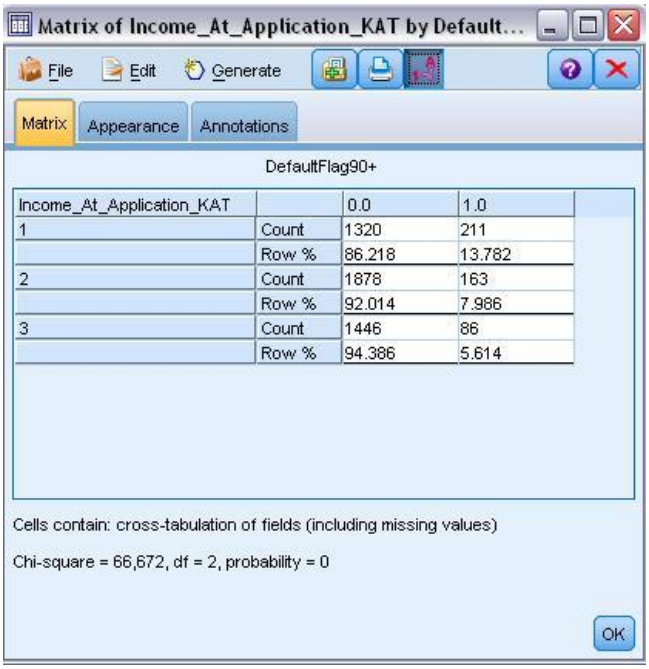


Tabela 6.12. Matrica prihoda pri apliciranju KAT

U Tabeli 6.12. su prikazane stope defaulta i broj slučajeva u kategorizovanoj neprekidnoj promenljivoj. Klijent pripada kategoriji

- *Prihod pri apliciranju KAT(1)* ako je prihod pri apliciranju manji ili jednak od 18, 119.82 RSD što ujedno predstavlja i najrizičniju kategoriju sa stopom defaulta 13.782%;
- *Prihod pri apliciranju KAT(2)* ako je prihod pri apliciranju u opsegu (18, 119.82 RSD, 30, 000.00 RSD] što ujedno predstavlja i srednje rizičnu kategoriju sa stopom defaulta 7.986%;

- *Prihod pri apliciranju KAT(3)* ako je prihod pri apliciranju veći od 30, 000.00 RSD što ujedno predstavlja i najmanje rizičnu kategoriju sa stopom defaulta 5.614%;

Upotrebom *Feature Selection* metode značajnost promenljive je 1 i u neprekidnom i u kategorijalnom obliku. Upotrebom Koksovog modela promenljiva kategorizovana preko *CHAID* algoritma *Prihod pri apliciranju KAT* se našla u modelu. Za kategoriju *Prihod pri apliciranju KAT(1)*, uočavamo da je vrednost koeficijenta β najveća a njegov predznak pozitivan što ukazuje da je ova kategorija najrizičnija i to čak 2.173 puta u odnosu na najbezbedniju kategoriju *Prihod pri apliciranju KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 1.660 i 2.844. Kako je vrednost *Wald*-ove test statistike 31.901, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000. Za kategoriju *Prihod pri apliciranju KAT(2)* uočavamo da je vrednost koeficijenta β manja u odnosu na kategoriju *Prihod pri apliciranju KAT(1)* a njegov predznak je opet pozitivan, što znači da je i ova kategorija rizična i to 1.428 puta u odnosu na najbezbedniju kategoriju *Prihod pri apliciranju KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 1.093 i 1.866. Kako je vrednost *Wald*-ove test statistike 6.819, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.009.

Rok otplate. Rok otplate plasmana predstavlja neprekidnu promenljivu pri čemu je ova promenljiva iskazana u mesecima. Očekujemo da je klijent bezbedniji ukoliko je rok otplate kraći, jer je lakše oceniti ponašanje klijenta i potencijalnu rizičnost za kraći vremenski period. Distribucija je prikazana histogramom, pri čemu su crvenom bojom označeni default klijenti a plavom ukupan broj klijenata.

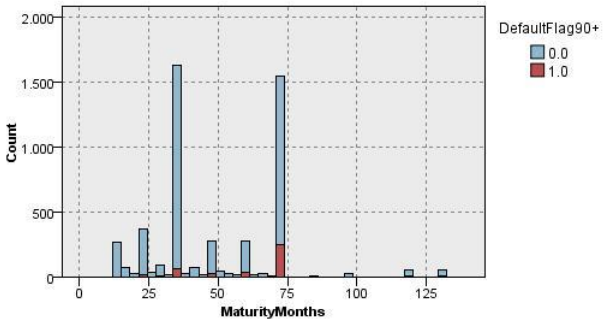


Tabela 6.13. Distribucija roka otplate

Osnovne karakteristike promenljive su:

Min	Max	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skew ness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.
12	132.529	120.53	49.64	0.324	23.119	534.487	0.779	0.034	0.966	0.069

Kako je promenljiva pokazala značajnu nelinearnu vezu sa zavisnom promenljivom kategorizovali smo je korišćenjem *CHAID* algoritma.

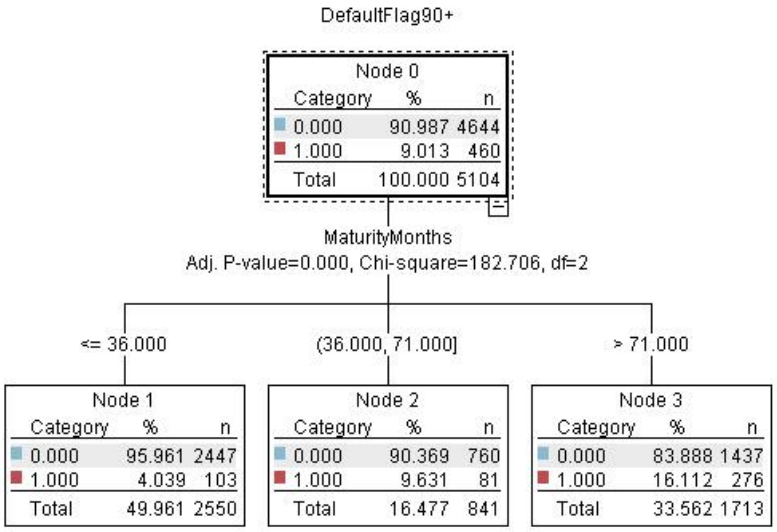


Tabela 6.14. *CHAID* roka otplate

U Tabeli 6.14. su prikazane stope defaulta i broj slučajeva u kategorizovanoj neprekidnoj promenljivoj. Klijent pripada kategoriji

- *Rok otplate KAT(1)* ako je rok otplate manji ili jednak od 36 meseci što ujedno predstavlja i najmanje rizičnu kategoriju sa stopom defaulta 4.039%;
- *Rok otplate KAT(2)* ako je rok otplate u intervalu od $(36, 71]$ meseci što ujedno predstavlja i srednje rizičnu kategoriju sa stopom defaulta 9.631%;

- *Rok otplate KAT(3)* ako je rok otplate veći od 71 mesec što ujedno predstavlja i najrizičniju kategoriju sa stopom defaulta 16.112%.

Kako je značajnost promenlijve 1 i u neprekidnom i u kategorijalnom obliku, kategorijalni oblik promenljive *Rok otplate KAT* se našao u modelu. Za kategoriju *Rok otplate KAT(1)*, uočavamo da je vrednost koeficijenta β najmanja i njegov predznak je negativan što ukazuje da je ova kategorija najmanje rizična i to čak 0.434 puta u odnosu na najrizičniju kategoriju *Rok otplate KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 0.333 i 0.566. Kako je vrednost *Wald*-ove test statistike 38.140, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000. Za kategoriju *Rok otplate KAT(2)* uočavamo da je vrednost koeficijenta β veća u odnosu na kategoriju *Rok otplate KAT(1)* a njegov predznak je opet negativan, što znači da je i ova kategorija manje rizična i to 0.741 puta u odnosu na najrizičniju kategoriju *Rok otplate KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 0.577 i 0.953. Kako je vrednost *Wald*-ove test statistike 5.459, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.019.

Mesečni anuitet. Mesečni anuitet po plasmanu predstavlja neprekidnu promenljivu koja prikazuje mesečne obaveze klijenta. Očekujemo da je klijent bezbedniji ukoliko su njegove mesečne obaveze manje. U praksi, ova promenljiva se transformiše u procenat zaduženosti u odnosu na primanja klijenta, ali pošto je ovaj model razvijen čisto u eksperimentalne svrhe, neće se vršiti nikakva transformacija. Distribucija je prikazana histogramom, pri čemu su crvenom bojom označeni default klijenti a plavom ukupan broj klijenata.

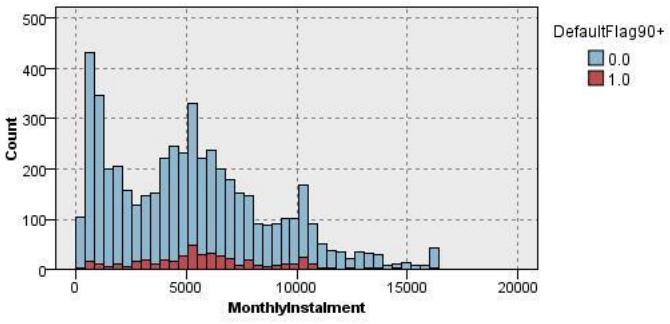
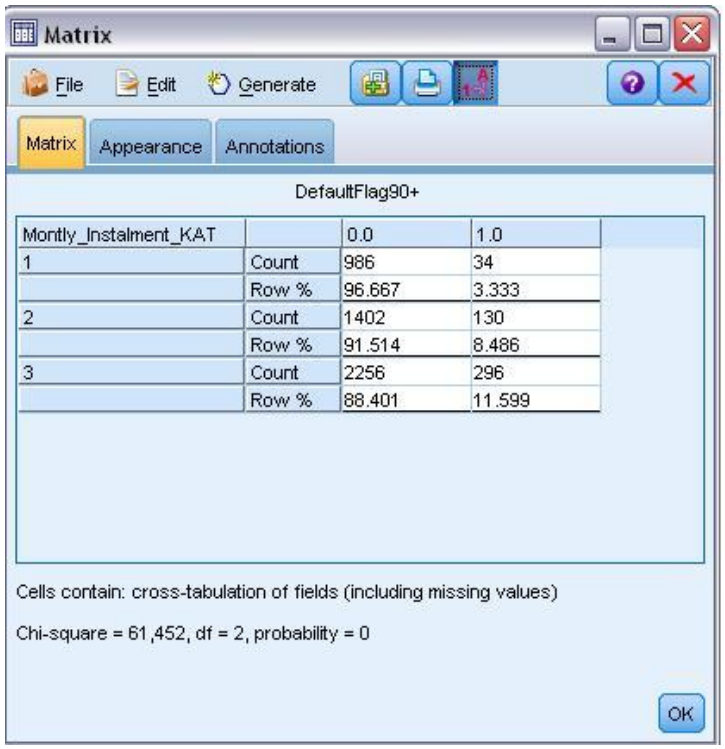


Tabela 6.15. Distribucija mesečnog anuiteta

Osnovne karakteristike promenljive su:

Min	Max	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skew ness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.
24.37	16420	16395	5276	50.19	3585	12856847	0.646	0.034	-0.032	0.069

Kako je promenljiva pokazala značajnu nelinearnu vezu sa zavisnom promenljivom kategorizovali smo je korišćenjem *CHAID* algoritma.



Montly_Instalment_KAT		0.0	1.0
1	Count	986	34
	Row %	96.667	3.333
2	Count	1402	130
	Row %	91.514	8.486
3	Count	2256	296
	Row %	88.401	11.599

Tabela 6.16. Matrica *Mesečnih anuiteta KAT*

U Tabeli 6.16. su prikazane stope defaulta i broj slučajeva u kategorizovanoj neprekidnoj promenljivoj. Klijent pripada kategoriji

- *Mesečnih anuiteta KAT(1)* ako ima mesečne obaveze manje ili jednake od 1526.43 RSD što ujedno predstavlja i najmanje rizičnu kategoriju sa stopom defaulta 3.33%;

- *Mesečnih anuiteta KAT(2)* ako ima mesečne obaveze u intervalu od $(1526.43, 5050.75]$ što ujedno predstavlja i srednje rizičnu kategoriju sa stopom defaulta 8.486%;
- *Mesečnih anuiteta KAT(3)* ako ima mesečne obaveze veće od 5050.75 RSD što ujedno predstavlja i najrizičniju kategoriju sa stopom defaulta 11.599%.

Kako je značajnost promenlijve 1 i u neprekidnom i u kategorijalnom obliku, u finalnom modelu se našao kategorijalni oblik promenljive *Mesečni anuitet KAT* dobijen preko *CHAID* algoritma. Za kategoriju *Mesečni anuitet KAT(1)* uočavamo da je vrednost koeficijenta β najmanja i njegov predznak je negativan što ukazuje da je ova kategorija najmanje rizična i to čak 0.470 puta u odnosu na najrizičniju kategoriju *Mesečni anuitet KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 0.312 i 0.707. Kako je vrednost *Wald*-ove test statistike 13.147, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000. Za kategoriju *Mesečni anuitet KAT(2)* uočavamo da je vrednost koeficijenta β veća u odnosu na kategoriju *Mesečni anuitet KAT(1)* a njegov predznak je opet negativan, što znači da je i ova kategorija manje rizična i to 0.776 puta u odnosu na najrizičniju kategoriju *Mesečni anuitet KAT(3)*. Donja i gornja granica 95%-tnog intervala poverenja su 0.625 i 0.964. Kako je vrednost *Wald*-ove test statistike 5.251, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.022.

Istorija iz Kreditnog Biroa. Iz Tabele 6.1. uočavamo da promenljiva *Istorija iz Kreditnog Biroa* ima tri kategorije. Distribucija je prikazana u nastavku.

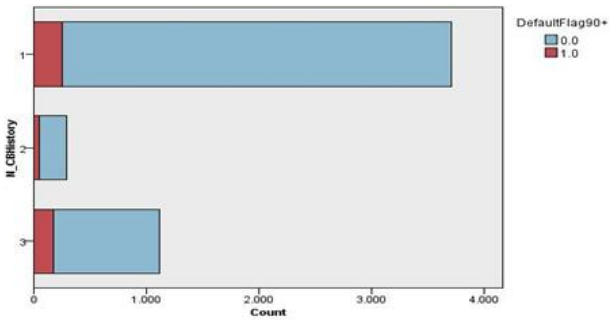


Tabela 6.17. Distribucija Istorija iz Kreditnog Biroa

Kategorija 1 označava klijente koji nisu problematični (72.59%),
Kategorija 2 označava pred-problematične klijente (5.62%),
Kategorija 3 označava potencijalno problematične klijente (21.79%).

U Tabeli 6.18. je prikazan udeo cenzurisan i default klijenata po kategoriji.

	Broj subjekata u kategoriji	Izraženo u procentima	Default slučajeva u kategoriji	Izraženo u procentima	Cenzurisanih slučajeva u kategoriji	Izraženo u procentima
Kategorija 1	3705	72.59%	244	6.59%	3461	93.41%
Kategorija 2	287	5.62%	41	14.29%	246	85.71%
Kategorija 3	1112	21.79%	175	15.74%	937	84.26%

Tabela 6.18. Raspodela po kategorijama

Primetimo da je stopa defaulta u kategoriji 1 znatno niža od preostalih kategorija jer su u pitanju neproblematični klijenti.

Korišćenjem *CHAID* algoritma spajamo određene kategorije i dobijamo novu promenljivu *Istorija iz Kreditnog Biroa KAT*.

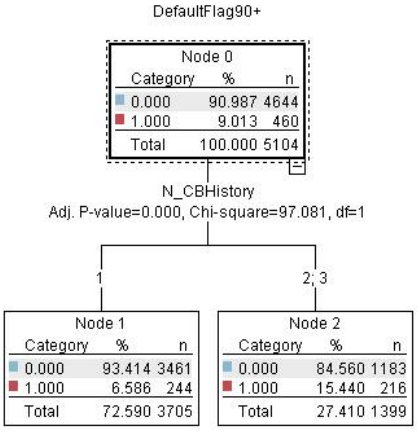


Tabela 6.19. *CHAID* Istorija kreditnog biroa

U Tabeli 6.19. su prikazane stope defaulta i broj slučajeva u rekategorizovanoj kategorijalnoj promenljivoj. Klijent pripada kategoriji

- *Istorija Kreditnog Biroa KAT(1)* ako nisu problematični što ujedno predstavlja i manje rizičnu kategoriju sa stopom defaulta 6.586%;
- *Istorija Kreditnog Biroa KAT(2)* ako su pred-problematični ili potencijalno problematični što ujedno predstavlja i rizičniju kategoriju sa stopom defaulta 15.44%;

Značajnost rekategorizovane promenljive *Istorija Kreditnog Biroa KAT* je 1 i ona se našla u finalnom modelu. Uočavamo da je predznak koeficijenta β negativan, znači da je kategorija neproblematičnih klijenata manje rizična od *Istorija Kreditnog Biroa KAT(2)*, tačno 0.498 puta. Donja i gornja granica 95%-tnog intervala poverenja su 0.410 i 0.601. Kako je vrednost *Wald*-ove test statistike 51.883, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000.

Ukupan broj članova u domaćinstvu. Distribucija ove promenljive je prikazana u nastavku. Očekujemo ukoliko je broj članova u domaćinstvu veći, klijent ima više izdržavanih članova u domaćinstvu, stoga je rizičniji.

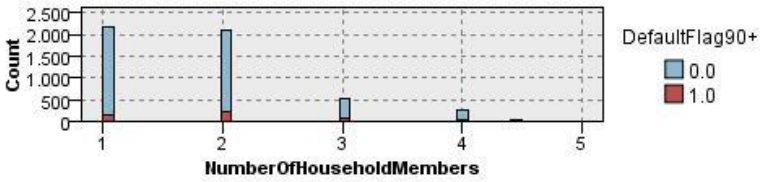


Tabela 6.20. Histogram promenljive Ukupan broj članova u domaćinstvu

Osnovne karakteristike promenljive su:

Min	Max	Range	Mean	Mean Std. Err.	Std. Dev	Variance	Skew ness	Skewness Std. Err.	Kurtosis	Kurtosis Std. Err.
1	4.496	3.496	1.807	0.012	0.872	0.761	1.040	0.034	0.605	0.069

Korišćenjem *CHAID* algoritma kategorizujemo neprekidnu promenljivu i dobijamo novu promenljivu *Ukupan broj članova u domaćinstvu KAT*.

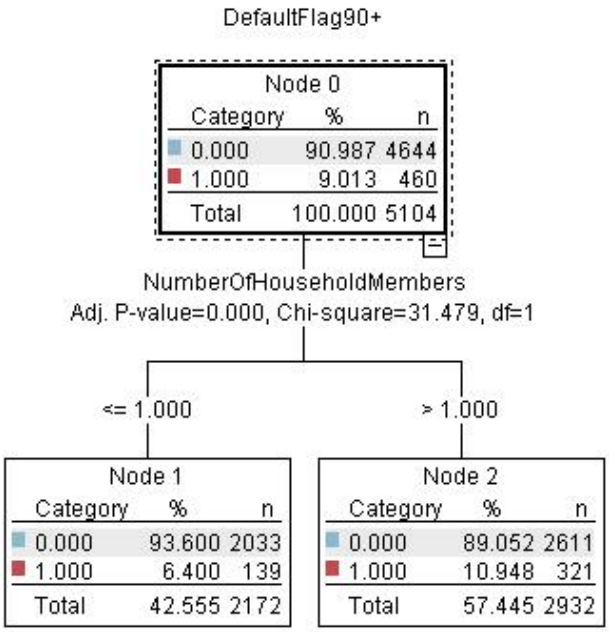


Tabela 6.21. *CHAID* Ukupan broj članova u domaćinstvu

Značajnost promenlijve je 1 i u neprekidnom i u kategorijalnom obliku. Upotrebom Koksovog modela, promenljiva kategorijalnog tipa *Ukupan broj članova u domaćinstvu KAT* se našla u modelu. Uočavamo da je predznak koeficijenta β negativan, znači da je kategorija *Ukupan broj članova u domaćinstvu KAT(1)* manje rizična od *Ukupan broj članova u domaćinstvu KAT(2)*, tačno 0.564 puta. Donja i gornja granica 95%-tnog intervala poverenja su 0.446 i 0.714. Kako je vrednost *Wald*-ove test statistike 22.737, sa jednim stepenom slobode, promenljiva je statistički značajna za predviđanje rizičnosti klijenta i *p-vrednost* je 0.000. Ova promenljiva je komplemenatrna sa promenljivom *Broj zaposlenih lica u domaćinstvu*.

6.2 Rezultati

Pri razvoju modela koristili smo Koksov metod po etapama unapred. U Tabeli 6.22. je prikazan omnibus test koeficijenata u modelu, koliko precizno model predviđa rizičnost klijenta i njegovo vreme preživljavanja. χ^2 vrednost se dobija kao razlika test količnika verodostojnosti prethodnog i narednog koraka. Može se pratiti promena kako po blokovima tako po pojedinačnim koracima u toku razvoja modela.

Omnibus test koeficijenata u modelu										
Korak	-2Log Likelihood	Ukupno			Promena u odnosu na prethodan korak			Promena u odonsu na prethodan blok		
		χ^2	df	Sig.	χ^2	df	Sig.	χ^2	df	Sig.
1(b)	7386.466	203.209	2	.000	200.128	2	.000	200.128	2	.000
2(c)	7249.348	335.401	3	.000	137.118	1	.000	337.246	3	.000
3(d)	7199.470	392.934	4	.000	49.878	1	.000	387.124	4	.000
4(e)	7178.650	417.492	6	.000	20.821	2	.000	407.945	6	.000
5(f)	7158.418	441.727	7	.000	20.232	1	.000	428.177	7	.000
6(g)	7140.987	462.167	9	.000	17.431	2	.000	445.608	9	.000
7(h)	7126.396	469.419	11	.000	14.591	2	.001	460.199	11	.000
8(i)	7116.156	479.055	12	.000	10.239	1	.001	470.438	12	.000
9(j)	7097.930	496.319	13	.000	18.226	1	.000	488.665	13	.000
a.Početni blok Korak 1 metod= Forward Stepwise (Likelihood Ratio)										
b.Promenljiva/e ušle u Koraku 1: Rok otplate _KAT										
c.Promenljiva/e ušle u Koraku 2: Radno iskustvo										
d.Promenljiva/e ušle u Koraku 3: Istorija iz Kreditnog Biroa_KAT										
e.Promenljiva/e ušle u Koraku 4: Prihod pri apliciranju _KAT										
f.Promenljiva/e ušle u Koraku 5: V_Pol_Mesto_Boravka_KAT										
g.Promenljiva/e ušle u Koraku 6: Godine stanovanja na datoj adresi _KAT										
h.Promenljiva/e ušle u Koraku 7: Mesečni anuitet _KAT										
i.Promenljiva/e ušle u Koraku 8: Ukupan broj članova u domaćinstvu _KAT										
j.Promenljiva/e ušle u Koraku 9: Broj zaposlenih lica u domaćinstvu										

Tabela 6.22. Omnibus test koeficijenata u modelu

Na Grafiku 6.1. je prikazana funkija preživljavanja i funkcija rizika u zavisnosti od vremena preživljavanja. Kao što smo i očekivali, sa porastom vremena preživljavanja, funkcija preživljavanja opada a funkcija rizika raste. Vidimo da krive nisu glatke već su stepenastog oblika jer u uzorku nemamo dovoljan broj subjekata.

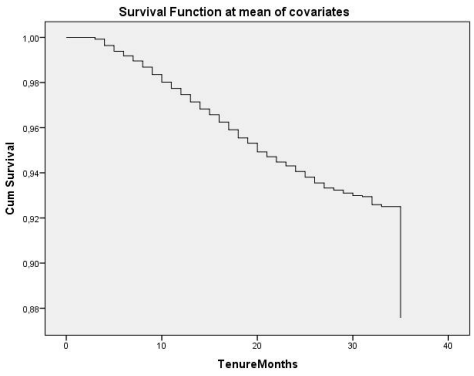


Figure 6.1: Funkcija preživljavanja

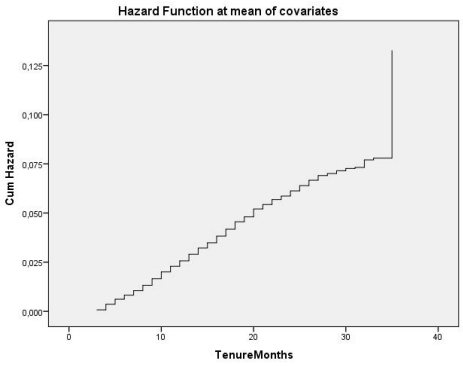


Figure 6.2: Funkcija hazarda

U Tabeli 6.23. je prikazana ocena funkcije preživljavanja i kumulativne i bazne hazard funkcije u zavisnosti od vremena preživljavanja. Pozivajući se na Poglavlje 5.3. i imajući u vidu da postoje ista vremena preživljavanja kod različitnih subjekata, za dobijanje ovih ocena se koristi iterativni postupak opisan ranije.

Survival Table

Time	Baseline Cum Hazard	At mean of covariates		
		Survival	SE	Cum Hazard
3	0.008	0.999	0.000	0.001
4	0.035	0.996	0.001	0.004
5	0.060	0.994	0.001	0.006
6	0.080	0.992	0.001	0.008
7	0.102	0.990	0.001	0.011
8	0.129	0.987	0.001	0.013
9	0.161	0.984	0.002	0.017
10	0.195	0.980	0.002	0.020
11	0.222	0.977	0.002	0.023
12	0.250	0.975	0.002	0.026
13	0.282	0.971	0.002	0.029
14	0.313	0.968	0.003	0.032
15	0.339	0.966	0.003	0.035
16	0.372	0.962	0.003	0.038
17	0.405	0.959	0.003	0.042
18	0.442	0.955	0.003	0.046
19	0.466	0.953	0.003	0.048

Survival Table				
Time	Baseline Cum Hazard	At mean of covariates		
		Survival	SE	Cum Hazard
20	0.505	0.949	0.004	0.052
21	0.527	0.947	0.004	0.054
22	0.552	0.945	0.004	0.057
23	0.569	0.943	0.004	0.059
24	0.595	0.941	0.004	0.061
25	0.621	0.938	0.004	0.064
26	0.647	0.936	0.004	0.067
27	0.670	0.933	0.004	0.069
28	0.680	0.932	0.005	0.070
29	0.694	0.931	0.005	0.071
30	0.705	0.930	0.005	0.073
31	0.710	0.929	0.005	0.073
32	0.747	0.926	0.005	0.077
33	0.756	0.925	0.005	0.078
35	1.288	0.876	0.033	0.133

Tabela 6.23. Tabela preživljavanja

Neophodno je proveriti da li se javlja multikolinearnost u finalnom modelu. Kao pokazatelji će nam poslužiti *Tolerance* koja mora biti veća od 0.4 i *VIF* što predstavlja parcijalnu korelaciju koja mora imati vrednost manju ili jednako od 2.5. U Tabeli 6.24. su prikazani izračunati koeficijenti i kao što vidimo u finalnom modelu ne postoji multikolinearnost.

	Collinearity Statistics	
	Tolerance	VIF
Radno iskustvo	.902	1.108
V_Pol_Mesto_Boravka_KAT	.968	1.033
Godine stanovanja na datoj adresi _KAT	.921	1.085
Broj zaposlenih lica u domaćinstvu	.751	1.331
Prihod pri apliciranju _KAT	.892	1.121
Rok otplate _KAT	.734	1.362
Mesečni anuitet _KAT	.757	1.322
Istorija iz Kreditnog Biroa_KAT	.929	1.076
Ukupan broj članova u domaćinstvu _KAT	.790	1.266

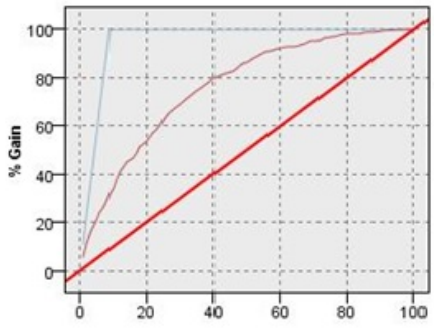
Tabela 6.24. Provera multikolinearnosti

Podsetimo se da *Kolmogorov-Smirnov (KS)* statistika prikazuje najveće rastojanje između kumulativnog procenta dobrih slučajeva i loših slučajeva. U praksi je prihvatljivo $KS > 30\%$. Koeficijenat *GINI* prikazuje koliko dobro razvijen model razlikuje dobre i loše slučajeve kao i efikasnost razvijenog modela. U praksi je prihvatljivo $GINI > 50\%$. Posmatramo razvijen model pomoću Koksove metode. Rezultati su sledeći:

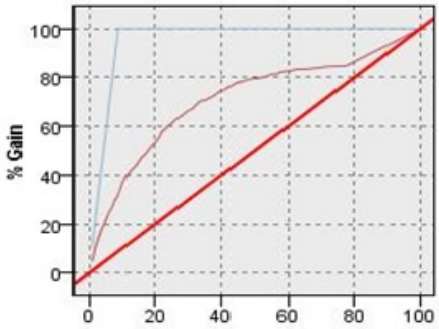
	GINI	KS
12 meseci	58.94%	44.96%
24 meseca	50.42%	38.91%

Vidimo da su i u prvoj i u drugoj godini rezultati i više nego zadovoljavajući, stoga smatramo da je razvijen model validan.

Na Grafiku 6.3. je prikazana ROC kriva razvijenog modela - kriva obeležena svetlo crvenom bojom, a kriva obeležena svetlo plavom bojom predstavlja idealan model. Što je veća površina ispod ROC krive, razvijen model bolje predviđa neuspeh subjekta.



GINI 12 meseci



GINI 24 meseca

Želimo da proverimo i kako se kreće stopa defaulta kod najboljih i najlošijih 10% slučajeva. Stopa defaulta predstavlja i procenat subjekata koji su doživeli neuspeh u odnosu na ukupan broj subjekata.

12 meseci	stopa defaulta	udeo u odnosu na ukupan broj defaulta
najgorih 10%	***33.63%***	***34.23%***
najboljih 10%	***0.70%***	***0.72%***

24 meseci	stopa defaulta	udeo u odnosu na ukupan broj defaulta
najgorih 10%	***57.55%***	***24.21%***
najboljih 10%	***3.77%***	***1.59%***

6.3 Poređenje sa modelom razvijenim pomoću logističke regresije

Već smo napomenuli ranije da želimo da razvijemo model pomoću logističke regresije, da bismo rezultate *GINI* i *KS* uporedili sa modelom razvijenim preko Koksove metode. Promenljive u logističkom modelu ostaju iste; koeficijenti su neznatno promenjeni ali najznačajnija razlika među ovim modelima predstavlja činjenica da logistički model podrazumeva fiksiran vremenski period od 12m na osnovu kog se računa rizičnost klijenta ali isključivo u tom periodu. U tabeli u nastavku su date promenljive i odgovarajući β koeficijenti.

Promenljive u finalnom modelu i njihove osnovne karakteristike								
	B	**SE**	**Wald**	**df**	**Sig**	**Exp (B)**	**95.0% CI for Exp(B)**	
							Lower	**Upper**
Presek	-.505	.298	2.871	1	.090			
Radno iskustvo	-.045	.005	87.744	1	.000	.956	.947	.965
V_Pol_Mesto_Boravka_KAT(1)	.596	.110	29.104	1	.000	1.815	1.461	2.253
Godine stanovanja na datoj adresi _KAT(1)	.652	.156	17.469	1	.000	1.918	1.413	2.604
Godine stanovanja na datoj adresi _KAT(2)	.318	.118	7.212	1	.007	1.375	1.090	1.734
Broj zaposlenih lica u domaćinstvu	-.461	.114	16.351	1	.000	.631	.504	.789
Prihod pri apliciranju _KAT(1)	.984	.151	42.640	1	.000	2.676	1.991	3.595
Prihod pri apliciranju _KAT(2)	.443	.147	9.065	1	.003	1.558	1.167	2.079
Rok otplate _KAT(1)	-1.013	.140	52.119	1	.000	.363	.276	.478
Rok otplate _KAT(2)	-.389	.141	7.674	1	.006	.677	.514	.892
Mesečni anuitet _KAT(1)	-.505	.214	5.582	1	.018	.604	.397	.918
Mesečni anuitet _KAT(2)	-.219	.123	3.136	1	.077	.804	.631	1.024
Istorija iz Kreditnog Biroa_KAT(1)	-.745	.108	47.454	1	.000	.475	.384	.587
Ukupan broj članova u domaćinstvu _KAT(1)	-.689	.130	27.953	1	.000	.502	.389	.648

Analiza modela je suštinski ista, osim analize β koeficijenta, koja se znatno razlikuje. Za kategorijalne promenljive, $Exp(\beta)$ se tumači kao odnos šansi da će klijent doživeti neuspeh ako se nalazi u definisanoj kategoriji u odnosu na baznu kategoriju. Ako imamo neprekidnu promenljivu, $Exp(\beta)$ opisuje odnos šansi neuspeha klijenta po jediničnoj promeni posmatrane neprekidne promenljive. I u ovom modelu, matematički je dokazano da promenljive podržavaju ekonomsko očekivanje.

Model razvijen preko logističke regresije daje rezultat za *GINI* 57.74% a *KS* 44.16% i ako uporedimo sa modelom razvijenim preko Koksove metode za 12 meseci, vidimo da su rezultati logističkog modela lošiji u oba slučaja.

Zaključujemo da model koji se dobija korišćenjem Koksove metode pruža kako preciznije i efikasnije rezultate u prvoj godini otplate, u poređenju sa logističkom metodom, tako i ostavlja mogućnost izračunavanja rizičnosti klijenta u svakom mesecu roka otplate kredita, što do sada nije bilo moguće.

Zaključak

Sumirajući dobijene rezultate u ovom radu, vidimo da je upotreba Koksovog PH modela u skoring modelima za odobravanje kredita opravdana i svakako predstavlja adekvatniji izbor od logističke regresije. Razvijen Koksov PH skoring model za odobravanje kredita pokazuje znatno bolje rezultate kako u prvoj godini otplate kredita, što smo pokazali poređenjem rezultata GINI i KS, tako i pruža mogućnost izračunavanja rizičnosti klijenta tokom celog perioda otplate, što ranije nije bilo moguće. Iako nije obrađeno, isto važi i za kreditne skoring modele o opštem ponašanju klijenta, stoga je neophodno zameniti standardnu metodologiju i upotrebu logističke regresije sa savremenom metodologijom i uporebom analize preživljavanja i Koksovog modela.

U ovom radu je razvijen model zasnovan isključivo na podacima dostupnim od strane klijenta u momentu apliciranja za dati kreditni proizvod, stoga nemamo dostupne bilo kakve istorijske podatke vezane za klijenta, osim onih karakteristika iz Kreditnog Biroa. Dalji razvoj i usavršavanje se može vršiti u pravcu uključivanja makro-ekonomskih vremenski zavisnih promenljivih koje su dostupne u svakom momentu. Naša pretpostavka je da će se u tom slučaju poboljšati preciznost i efikasnost modela. Predlog promenljivih koje se mogu u daljem razvoju uključiti su: stopa inflacije, promena kursa, promena kamatnih stopa, promena vrednosti potrošačke korpe, promena prosečne plate u Srbiji, itd. Interesantno bi bilo posmatrati kako se rizičnost klijenta preciznije izračunava u zavisnosti od gore navedenih faktora.

Rad završavamo sa konstatacijom da analiza preživljavanja i Koksov model predstavljaju savremen metod ocene rizičnosti klijenta, koji nadmašuje logističku regresiju u svakom pogledu. Shodno tome neophodno je da finansijske institucije modifikuju standardnu metodologiju sa savremenim načinom izračunavanja rizičnosti klijenta.

Literatura

[1] Maria Stepanova, Lyn Thomas, *Survival analysis methods for personal loan data*, Operations Research 2002 INFORMS Vol. 50, No. 2 (2002), pp. 277 − 289.

[2] Maria Stepanova, Lyn Thomas, *PHAB scores: proportional hazards analysis behavioral scores*, Journal of the Operational Research Society (2001)52, pp. 1007 − 1016

[3] Buda Bajić, Milena Kresoja, *Analiza preživljavanja i Koksov PH model*, Seminarski rad iz Statističkog modeliranja, Univerzitet u Novom Sadu, Novi Sad (2011)

[4] J.D Kalbfleisch, R.L. Prentice, *The Statistical Analysis of Failure Time Data*, John Wiley & Sons (2002)

[5] Jaroslav Pazdera, Michal Rychnovsky, Petr Zahradnik, *Survival analysis in credit scoring*, Seminar on Modelling in Economics, Charles University in Prague, Faculty of Mathematics and Physics, Prague (2009)

[6] Anupam Saha, Naeem Siddiqi *Survival Analysis Workflow assessing the impact of Macro-Economic Shocks on Credit Portfolios and predicting the Time of Default*,SAS Institute Inc

[7] N. Bucay, D. Rosen, *Applying Portfolio Credit Risk Models to Retail Portfolios*, Journal of Risk Finance 2 (2001), pp. 35 − 61.

[8] M. Malik, L. C. Thomas, *Modeling credit risk of portfolios of consumer loans*, Working Paper CORMSIS 07-12, to appear in Journal of the Operational Research Societ (2007), Southampton, University of Southampton.

[9] Jian-Jian Ren, Mai Zho, *Full likelihood inferences in the cox model: An empirical likelihood approach*, Working Paper, University of Central Florida and University of Kentucky (2009)

[10] K.R. Bailey, *Asymptotic equivalence between the Cox estimator and the general ML estimators of regression and survival parameters in the Cox model*,Ann. Statist. 12(1984), pp. 730 − 736.

[11] A. A Tsiatis, *A large sample study of Coxs regression model*,Ann. Statist. 9(1981), pp. 93 − 108.

[12] Alireza Ghasemi, Soumaya Yacout, M. Salah Ouali, *Parameter Estimation for Condition Based Maintenance with Proportional Hazard Model*, International Conferenceon Industrial Engineering and Systems Management (2009), Department of Mathematics and Industrial Engineering, Montral, Qubec, Canada

[13] Yuichi Hirose, *Efficiency of Profile/Partial Likelihood in the Cox Model*, Working Paper, School of Mathematics, Statistics and Operations Research, Victoria University of Wellington,New Zealand

[14] Shian-Chang Huang, *A new corporate credit scoring system using semi-supervised discriminant analysis*, Full Length Research Paper, African Journal of Business Management Vol. 5(22) (2011), pp. 9355 − 9362

[15] Steven Finlay, *Multiple classifier architectures and their application to credit risk assessment*, Working Paper, Lancaster University Management School (2008)

[16] Ross A. McDonald, A. Matuszyk, Lyn C. Thomas, *Application of survival analysis to cash flow modeling for mortgage products*, Working Paper, Quantitative Financial Risk Management Center, School of Management, University of Southampton, Southampton, UK

[17] Qamruz Zaman, Karl P. Pfeiffer, *Survival Analysis in Medical Research*, Working Paper, Department of Statistics, University of Peshawar, Pakistan

[18] Ivana Stokić, *Primena genetskog programiranja u strojnom učenju*, Diplomski rad (2011), Sveučilište u Zagrebu, Fakultet elektrotehnike i računarstva, Zagreb

[19] Srinvas Gumparthi, Dr. V.Manickavasagam, *Risk classification based on discriminant analysis for SMEs*, International Journal of Trade, Economics and Finance, Vol.1, No.3 (2010)

[20] Lyn C. Thomas, *A Survey of Credit and Behavioral Scoring;Forecasting financial risk of lending to consumers*,Department of Business Studies,University of Edinburgh,Edinburgh, UK

[21] Bart Baesens, Tony Van Gestel, Maria Stepanova, Dirk Van den Poel, *Neural Network Survival Analysis for Personal Loan Data*,Working Paper, Universiteit Gent, Faculteit economie en Bedrijfskunde (2004)

[22] Jih-Jeng Huang, Gwo-Hshiung Tzeng, Chorng-Shyong Ong, *Two-stage genetic programming (2SGP)for the credit scoring model*,Applied Mathematics and Computation 174 (2006) pp. $1039 - 1053$, Elsevier Inc

[23] J-K Im, DW Apley, C Qi, X Shan, *A time-dependent proportional hazards survival model for credit risk analysis*, Journal of the Operational Research Society (2012) 63, pp. $306 - 321$

[24] Ricardo Cao, Juan M. Vilar and Andres Devia, *Consumer credit risk via survival analysis*, Departamento de Matematicas, Facultad de Informatica, Universidade da Coruna, Spain (2009), pp. $3 - 30$

[25] *IBM SPSS Complex Samples 20*, IBM Corporation 1994(2012)

[26] *IBM SPSS Modeler 15 Algorithms Guide*, IBM Corporation 1994(2011)

[27] *IBM SPSS Statistics 20 Algorithms*, IBM Corporation 1994(2011)

[28] Gorica Gvozdić, *Primenjena logistička regresija*, Master rad, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Novi Sad (2011)

[29] Jelena Burgijašev, *Različiti pristupi kreditnom skoring sistemu*,Master rad, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Novi Sad (2009)

[30] Željka Tepavčević, *Analiza preživljavanja- Koksov model*, Diplomski rad, Univerzitet u Novom Sadu, Prirodno-matematički fakultet, Novi Sad (2006)

[31] John Watkins, Andrey Vasnev, Richard Gerlach, *Survival Analysis for Credit Scoring: Incidence and Latency*, Working Paper, Faculty of Economics and Business, The University of Sydney (2009)

[32] Alnoor Bhimani, Mohamed Azzim Gulamhussen, Samuel Da Rocha Lopes, *The role of financial, macroeconomic and non-financial information in bank loan default timing prediction*, Working Paper, London School of Economics and Political Science, London (2009)

[33] L. Thomas, S. Thomas, L. Tang, O. A. Gwilym, *The Impact of Demographic and Economic Variables on Financial Policy Purchase Timing Decisions*, Working Paper, School of Management, University of Southampton, UK

[34] John Fox, *Cox Proportional-Hazards Regression for Survival Data*, Working Paper, Appendix to An R and S-PLUS Companion to Applied Regression (2002)

[35] L.C. Thomas, D.B. Edelman, J.N. Crook, *Credit scoring and its applications*, Philadelphia, PA: Society for Industrial and Applied Mathematics(2002)

[36] Mr Tomislav M. Todorović, *Upravljanje kreditnim rizikom u banci*, Ekonomski horizonti (2009), 11, (2) str. 81 – 99

Biografija

Sandra Rackov je rođena 29. januara 1987. godine u Kikindi. Osnovnu školu "Vuk Karadžić" i gimnaziju "Dušan Vasiljev" je završila u Kikindi. Studije na Prirodno-matematičkom fakultetu u Novom Sadu, smer diplomirani matematičar matematika finansija, upisala je 2006. godine, a završila u oktobru 2010. godine sa prosečnom ocenom 9,00. Iste godine je upisala master studije primenjene matematike, modul matematika finansija, na istom fakultetu. Zaključno sa oktobarskim ispitnim rokom 2011. godine, položila je sve predviđene ispite sa prosečnom ocenom 9,29.

Novembra 2011. godine je započela jednogodišnje stručno usavršavanje i osposobljavanje u OTP Banci Srbija a.d. Novi Sad, gde se i zaposlila novembra 2012. godine kao savetnik za kvantitativne analize i modeliranje, u Sektoru za upravljanje rizicima.

UNIVERZITET U NOVOM SADU
PRIRODNO - MATEMATIČKI FAKULTET
KLJUČNA DOKUMENTACIJA INFORMACIJA

Redni broj:
RBR

Identifikacioni broj:
IBR

Tip dokumentacije: Monografska dokumentacija
TD

Tip zapisa: Tekstualni štampani materijal
TZ

Vrsta rada: Master rad
VR

Autor: Sandra Rackov
AU

Mentor: Prof. dr Zagorka Lozanov-Crvenković
ME

Naslov rada: Primena Koksovog PH modela u analizi kreditnog rizika
NR

Jezik publikacije: Srpski (latinica)
JP

Jezk izvoda: s / en
JI

Zemlja publikovanja: Republika Srbija
ZP

Uže geografsko područje: Vojvodina

UGP

Godina: 2013
GO

Izdavač: Autorski reprint
IZ

Mesto i adresa: Novi Sad, Trg D. Obradovića 4
MA

Fizički opis rada: (6/89/3/22/0/22/0)(broj poglavlja/broj strana/broj literarnih citata/broj tabela/broj slika/broj grafika/broj priloga)
FO:

Naučna oblast: Matematika
NO

Naučna disciplina: Primenjena matematika
ND

Ključne reči: Koksov PH model, Kreditni skoring modeli, Analiza preživljavanja, Analiza kreditnog rizika

PO, UDK

Čuva se: U biblioteci Departmana za matematiku i informatiku, Prirodno-matematički fakultet, Uviverzitet u Novom Sadu
ČU

Važna napomena:
VN

Izvod: Ovaj rad se bavi analizom preživljavanja i njegovom primenom u analizi kreditnog rizika. Razvijen je Koksov PH skoring model za odobravanje kredita pri čemu je prikazana rizičnost klijenta u zavisnosti od roka otplate.
IZ

Datum prihvatanja teme od strane NN veća: 11.06.2013.
DP

Datum odbrane:
DO

Članovi komisije:
ČK

Predsednik: dr Danijela Rajter-Ćirić, redovni profesor Prirodno-matematičkog fakulteta u Novom Sadu
Član: dr Zagorka Lozanov-Crvenković, redovni profesor Prirodno-matematičkog fakulteta u Novom Sadu
Mentor: dr Ivana Štajner-Papuga, vanredni profesor Prirodno-matematičkog fakulteta u Novom Sadu

UNIVERSITY OF NOVI SAD
FACULTY OF SCIENCE
KEY WORDS DOCUMENTATION

Accession number:
ANO

Identification number:
INO

Document type: Monograph type
DT

Type of record: Printed text
TR

Contents Code: Master's thesis
CC

Author: Sandra Rackov
AU

Mentor: Phd Zagorka Lozanov-Crvenković
MN

Title: Application of Cox PH model in credit risk analysis
TI

Language of text: Serbian (Latin)
LT

Language of abstract: s / en
LA

Country of publication: Republic of Serbia
CP

Locality of publication: Vojvodina

LP

Publication year: 2013
PY

Publisher: Author's reprint
PU

Publication place: Novi Sad, Trg D. Obradovića 4
PP

Physical description: (6/89/3/22/0/22/0)(chapters/ pages/ quotations/tables/ pictures/ graphics/ enclosures)
PD

Scientific field: Mathematics
SF

Scientific discipline: Applied mathematics
SD

Subject/Key words: Cox PH model, Credit scoring model, Survival analysis, Credit risk analysis
SKW

Holding data: The Library of the Department of Mathematics and Informatics, Faculty of Science and Mathematics, University of Novi Sad
HD

Note:
N

Abstract: This master thesis is about survival analysis and its application in credit risk analysis. Cox PH scoring model has been developed to show us dependence of clients risk status and maturity months of a loan.
AB

Accepted by the Scientific Board on: 11.06.2013.
ASB

Defended:

DE

Thesis defend board:
DB

President: Phd Danijela Rajter-Ćirić, full professor at Faculty of Science in Novi Sad
Member: Phd Zagorka Lozanov-Crvenković, full professor at Faculty of Science in Novi Sad
Mentor: Phd Ivana Štajner-Papuga, associate professor at Faculty of Science in Novi Sad