UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO
FINANČNA MATEMATIKA - 2. STOPNJA

Špela Kern

OCENJEVANJE IZGUBE OB NEPLAČILU Z ODLOČITVENIM DREVESOM

Magistrsko delo

Mentor: prof. dr. Tomaž Košir

Ljubljana, 2016

Izjava o avtorstvu

Podpisana Špela Kern izjavljam:

- da sem magistrsko delo z naslovom *Ocenjevanje izgube ob neplačilu z odločitvenim drevesom* izdelala samostojno pod mentorstvom prof. dr. Tomaža Koširja in
- da Fakulteti za matematiko in fiziko Univerze v Ljubljani dovoljujem objavo elektronske oblike svojega dela na spletnih straneh.

Ljubljana, dne 10. 6. 2016 Podpis: ....

Zahvala

Zahvaljujem se svojemu mentorju prof. dr. Tomažu Koširju za pomoč in strokovno usmerjanje pri pisanju magistrskega dela.

Še posebej velika zahvala gre mojim najbližjim, atiju Francu, mami Danici, bratu Maticu ter partnerju Juretu, ki so mi vsa leta študija stali ob strani ter me podpirali in me vzpodbujali.

Magistrsko nalogo posvečam svojim najbližjim.

Kazalo

Program dela

V magistrskem delu opišite ocenjevanje izgube ob neplačilu kredita s stališča banke s pomočjo metode odločitvenih dreves. Predstavite to metodo tudi na zgledu.

Ljubljana, dne 7. 9. 2015 prof. dr. Tomaž Košir

Ocenjevanje izgube ob neplačilu z odločitvenim drevesom

Povzetek

Kreditno tveganje je eno izmed glavnih bančnih tveganj. Predstavlja tveganje nastanka izgube, ki je posledica dolžnikove nezmožnosti izpolnjevanja pogodbenih obveznosti v dogovorjenem roku. Prvi del magistrskega dela je namenjen ocenjevanju izgub v bankah (pričakovana, nepričakovana ter ekstremna izguba), nanje pa pomembno vplivajo faktorji tveganj verjetnost neplačila, izpostavljenost ob neplačilu in izguba ob neplačilu. Na ravni posameznega posla izguba ob neplačilu pomeni razmerje med izgubo na izpostavljenosti, ki je posledica neplačila nasprotne stranke, in vrednostjo izpostavljenosti ob neplačilu. Drugi del magistrskega dela je namenjen opisu gradnje odločitvenih dreves, ki se pogosto uporabljajo za segmentacijo portfelja. Na ravni posameznega segmenta portfelja se izguba ob neplačilu izračuna z modeli kot so tehtana povprečja deležev izgub ob neplačilu poslov, ki se preko odločitvenega drevesa uvrstijo v opazovani segment portfelja. Zadnji del magistrskega dela pojasnjuje pomen segmentacije portfelja pri ocenjevanju izgube ob neplačilu na primeru že zgrajenega odločitvenega drevesa.

Estimation of loss given default with a decision tree

Abstract

Credit loss is one of the most important bank risks. The risk of loss results from the debtors inability to fulfill contractual obligations within the agreed timeline. The first part of the master thesis is devoted to assessing losses in banks (the expected, unexpected and extreme loss), which are significantly affected by risk factors probability of default, exposure at default and loss given default. The loss given default on the loan level is equal to ratio of losses, which is the result of default, to exposure at default. The second part of the master thesis is devoted to the description of the building of decision trees, which are often used for portfolio segmentation. The loss given default on the portoflio segment level is calculated with models, such as weighted average of losses given default of loans which are through decision tree classified to the observed segment. The last part of the master thesis, from the example of already built decision tree, explains the importance of the portfolio segmentation in assessing the loss given default.

Math. Subj. Class. (2010): 05C05, 05C85, 91B06,

Ključne besede: banke, izguba ob neplačilu, odločitvena drevesa

Keywords: banks, loss given default, decision trees

Poglavje 1

Uvod

Banke so vsakodnevno izpostavljene številnim tveganjem, ki predstavljajo negotovosti o prihodnosti ter lahko negativno vplivajo na poslovanje in stabilnost banke.

Na splošno se bančna tveganja delijo na kreditno, tržno, obrestno, valutno, likvidnostno, operativno, kapitalsko, poslovno tveganje ter tveganje ugleda. V zadnjih letih so slovenske banke najbolj izpostavljene kreditnemu tveganju, ki je opredeljeno kot tveganje nastanka izgub v primeru neplačila ali poslabšanja kreditne kakovosti dolžnika. Največji vir izgub izvira iz tveganja nastanka izgube, ki je posledica dolžnikove nezmožnosti izpolnjevanja pogodbenih obveznosti v dogovorjenem roku. Izpostavljenost kreditnemu tveganju se glede na [1] postopoma zmanjšuje, kar je posledica ukrepov sanacije bank, aktivnosti na področju prestrukturiranja podjetij ter stabilnejše gospodarsko okrevanje.

Težišče magistrske naloge je na kreditnem tveganju, natančneje na ocenjevanju izgube ob neplačilu, ki je poleg verjetnosti neplačila in izpostavljenosti ob neplačilu eden od ključnih fakorjev tveganj v modelih za ocenjevanje in napovedovanje izgub iz kreditnega tveganja v bankah.

Prvo poglavje povzema magistrsko delo. Drugo poglavje je namenjeno predstavitvi glavnih bančnih tveganj s poudarkom na kreditnem tveganju. Opisan je življenjski cikel posojila, ki je osnova za preučevanje kreditnega tveganja. Sledi definicija dogodka neplačila in opis treh najpomembnejših faktorjev kreditnega tveganja, izpostavljenosti ob neplačilu, verjetnosti neplačila in izgube ob neplačilu.

V tretjem poglavju so opisane pričakovana, nepričakovana in ekstremna izguba. Predstavljena sta modela, ki z različnimi predpostavkami ocenjujeta pričakovano in nepričakovano izgubo. V nadaljevanju poglavja sledijo cilji in motivacija pristopa notranjih bonitetnih sistemov, ki v sklopu novega baselskega sporazuma banke vzpodbuja k naprednejšim metodam ocenjevanja kreditnega tveganja. Predstavljen je izračun pričakovane izgube na ravni posameznega posla ter na ravni portfelja. Sledi opis pričakovanih posledic uporabe pristopa notranjih bonitetnih sistemov ter pogojev za pridobitev dovoljenja s strani pristojnega organa za uporabo tega pristopa.

Četrto poglavje je namenjeno ocenjevanju izgube ob neplačilu, ki v literaturi pogosto nastopa v obliki svojega komplementa, stopnje poplačila. Izguba ob neplačilu se najpogosteje ocenjuje kot razmerje med oceno dejanske izgube in kreditno izpostavljenostjo v trenutku neplačila. Opisana je dinamika denarnih tokov, ki vpliva na oceno dejanske izgube. Sledi opis izračuna izgube ob neplačlu na ravni posameznega posla ter na ravni portfelja. Na koncu poglavja so predstavljeni podatki, katere je priporočljivo zbirati za namen modeliranja izgube ob neplačilu.

V petem poglavju so definirani osnvni pojmi iz teorije grafov. Predstavljena so binarna drevesa, ki so osnova za nadaljevanje magistrske naloge.

Šesto poglavje je namenjeno teoretičnemu ozadju gradnje odločitvenih dreves. Njihova značilnost je, da v vsaki notranji točki na izbiro ponujajo natanko dve možnosti, s čimer se zgradijo poti od korena do listov drevesa. Opisani so metoda rekurzivnega deljenja, kriteriji za ustavitev ter postopek obrezovanja dreves z navzkrižno validacijo.

V sedmem poglavju je predstavljena študija modeliranja izgube ob neplačilu s pomočjo gradnje odločitvenega drevesa. Predstavljeni so podatki, ki so bili uporabljeni v postopku modeliranja, opisana so odločitvena drevesa za različna casovna obdobja ter ocenjene so napovedne moči modelov.

Osmo poglavje magistrsko delo zaključuje.

Poglavje 2

Bančna tveganja

Tveganja predstavljajo negotovosti o prihodnosti, ki lahko negativno vplivajo na poslovanje in stabilnost banke. Banka se pri svojem delovanju in poslovanju sooča s tveganji različnih vrst, ki se med sabo razlikujejo po vsebini in razsežnosti.

S prevzemanjem tveganj se povečuje pričakovan zaslužek, vendar pa se s tem povečuje tudi verjetnost nastanka velikih izgub, ki so lahko usodne za delovanje banke. Te se odločajo za različen nivo prevzemanja tveganj, kar se odraža tudi v njihovem načinu poslovanja, strukturi bančnega portfelja in ponudbi produktov. Za banke je posledično ključna ustrezno organizirana funkcija tveganj, ki zagotavlja, da so vsa pomembna tveganja prepoznana ter ustrezno poročana vodstvu banke.

V tem poglavju sta predstavljena funkcija in pomen upravljanja tveganj, saj lahko pomanjkljivo in neustrezno upravljanje tveganj vodi v izgubo. Sledijo kratki opisi glavnih bančnih tveganj s poudarkom na kreditnem tveganju. Za ustrezno razumevanje kreditnega tveganja je predstavljen življenjski cikel posojila, v nadaljevanju pa so navedeni razlogi za merjenje kreditnega tveganja ter opisi osnovnih pojmov in faktorjev kreditnega tveganja.

2.1 Upravljanje tveganj

Glavna dejavnost bank je opravljanje finančnega posredništva, ki je v [2] opredeljeno kot dejavnost zbiranja finančnih sredstev od tistih, ki imajo presežke ter posojanja finančnih sredstev tistim, ki ta sredstva potrebujejo. Pri tem prihaja do prevzemanja tveganj, pomanjkljivo in neustrezno upravljanje tveganj pa lahko vodi v izgube, zato je pomembno, da banke tveganja prevzemajo previdno.

Zakon o bančništvu [3] v četrti točki 138. člena navaja, da mora funkcija upravljanja tveganj zagotoviti:

1. da so vsa pomembna tveganja ugotovljena, ocenjena oziroma izmerjena in da se o njih ustrezno poroča,
2. aktivno sodelovanje pri pripravi strategije upravljanja tveganj banke in pri vseh pomembnih odločitvah glede upravljanja tveganj,
3. oblikovanje celovitega pregleda nad tveganji, katerim je banka pri svojem poslovanju izpostavljena ali bi jim lahko bila;

Učinkovito upravljanje tveganj je eden ključnih elementov za varno in stabilno poslovanje banke, ki vodstvu omogoča porazdeljevanje virov v različna področja glede na kvanitativno oceno tveganj. Pri podaji ocene tveganosti se morajo upravljalci tveganj glede na [4] vprašati naslednje:

1. Koliko lahko izgubimo?
2. Ali lahko absorbiramo pomembno izgubo ne da bi banka postala insolventna?
3. Ali je donos posla dovolj visok, da sprejmemo tveganje?
4. Kako lahko zmanjšamo tveganje, brez da bi pomembno zmanjšali donos?

Sposobnost prevzemanja tveganj je opredeljena v šestem členu Sklepa o upravljanju s tveganji in izvajanju procesa ocenjevanja ustreznega notranjega kapitala za banke in hranilnice [5], ki navaja, da sposobnost prevzemanja tveganj pomeni, da mora banka zagotoviti, da je vsakokratno prevzemanje pomembnih tveganj v mejah njene sposobnosti prevzemanja tveganj.

2.1.1 Glavna bančna tveganja

Sledijo kratke definicije glavnih bančnih tveganj, ki so povzete iz [2] in [6].

- **Kreditno tveganje** je tveganje nastanka izgube, ki je posledica dolžnikove nezmožnosti, da zaradi kateregakoli razloga ne izpolni svoje finančne ali pogodbene obveznosti v celoti.
- **Tržno tveganje** izvira iz negotovosti zaradi neugodnih sprememb dejavnikov tveganj, kot so obrestne mere, devizni tečaji, kreditni razmiki, cene finančnih instrumentov, cene blaga in drugi pomembni dejavniki, ki neugodno vplivajo na vrednost finančnega instrumenta.
- **Valutno tveganje** je tveganje spremembe poštene vrednosti ali prihodnjih denarnih tokov finančnih instrumentov zaradi sprememb tečajev tujih valut.
- **Obrestno tveganje** je tveganje nastanka izgube, ki nastane zaradi neugodnega gibanja obrestnih mer.
- **Likvidnostno tveganje** je tveganje nastanka izgube, ko banka ni sposobna poravnati vseh dospelih obveznosti oziroma, ko je banka zaradi nezmožnosti zagotavljanja zadostnih sredstev za poravnavo obveznosti ob dospelosti prisiljena zagotavljati potrebna sredstva s pomembno višjimi stroški od običajnih.
- **Operativno tveganje** je tveganje nastanka izgube zaradi neprimernega ali neuspešnega izvajanja notranjih procesov, ravnanj ljudi in delovanja sistemov ali zaradi zunanjih dogodkov. Vključuje tudi tveganje neustrezne informacijske tehnologije in procesiranja predvsem z vidika obvladljivosti, dostopov, integralnosti, nadzora in neprekinjenosti.
- **Strateško tveganje** je tveganje nastanka izgube zaradi nepravilnih poslovnih odločitev banke, neustreznega izvajanja sprejetih odločitev ter premajhne odzivnosti banke na spremembe poslovnega okolja.
- **Tveganje ugleda** je tveganje nastanka izgube zaradi negativne podobe, ki jo imajo o banki njeni komitenti, poslovni partnerji, lastniki in investitorji ali nadzorniki.

- **Kapitalsko tveganje** izhaja iz neustrezne višne kapitala glede na obseg in način poslovanja ali iz težav pri pridobivanju novega kapitala.

- **Tveganje dobičkonosnosti** se nanaša na neustrezno sestavo oziroma razpršenost prihodkov ali na nesposobnost banke, da zagotavlja zadostno in stalno raven dobičkonosnosti.

Sledi podrobnejši opis kreditnega tveganja, ki je težišče magistrskega dela.

2.2 Kreditno tveganje

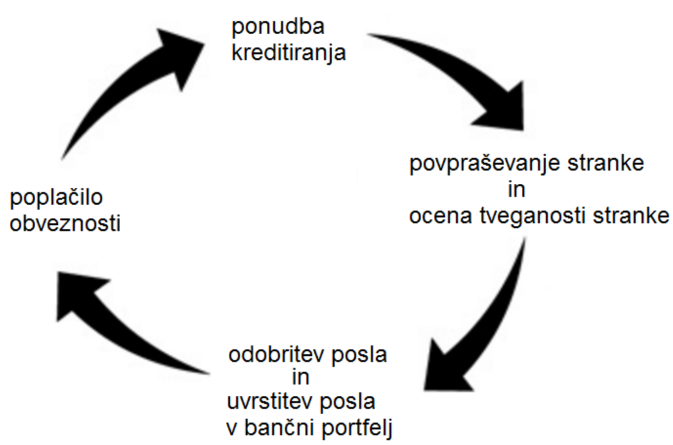
Eno izmed najpomembnejših bančnih tveganj je kreditno tveganje, ki predstavlja tveganje nastanka izgube zaradi neizpolnitve obveznosti, ki jih ima nasprotna stranka do banke. Gre za tveganje, da dolžnik ne bo pripravljen ali ne bo sposoben v celoti izpolniti svojih pogodbenih obveznosti v dogovorjenem roku.

Proces upravljanja kreditnega tveganja glede na [6] vključuje postopke ugotavljanja, merjenja, obvladovanja in spremljanja kreditnega tveganja, vključno s poročanjem o kreditnem tveganju. Temeljna cilja procesa upravljanja kreditnega tveganja sta doseganje in ohranjanje kakovosti in razpršenosti kreditnega portfelja, ki zagotavlja stabilno in varno poslovanje ob hkratnem doseganju ciljnih vrednosti donosnosti in kapitalske ustreznosti.

Sledi opis življenjskega cikla posojila, ki je osnova za preučevanje kreditnega tveganja. Sledi opis področij, pri katerih merjenje kreditnega tveganja pomembno pripomore k odločitvam. Nato sledijo še definicija neplačila ter opisi osnovnih faktorjev kreditnega tveganja.

2.2.1 Življenjski cikel posojila

Osnova za preučevanje kreditnega tveganja je poznavanje življenjskega cikla posameznega posojila, ki je prikazan na sliki 2.1.



Slika 2.1: Življenjski cikel posojila.

Cikel se začne s ponudbo kreditiranja, ki jo banka pripravi za potencialne stranke. V naslednji fazi potencialna stranka banki posreduje svoje povpraševanje, vključno z osnovnimi informacijami o svoji kreditni kvaliteti. Za presojanje kreditne kvalitete fizična oseba banki predloži nekaj osebnih informacij, kot na primer dokazila o osebnem dohodku, pravna oseba pa predloži informacije o bilančnem stanju podjetja.

Poleg informacij, ki jih stranke predložijo same, svojo oceno banke pogosto podprejo s podatki zunanjih ogencij. V Sloveniji se podatki o fizičnih osebah zbirajo v informacijskem sistemu Sisbon, ki je glede na [7] vzpostavljen z namenom izpolnitve obveznosti, predpisane v štirinajstem poglavju Zakona o bančništvu [3], upravljanja s kreditnim tveganjem bank, hranilnic in drugih dajalcev kreditov, zagotavljanja odgovornega kreditiranja in preprečevanja prezadolženosti posameznikov, izvajanje nadzora nad upravljanjem kreditnega tveganja v bankah ter statističnega poročanja. Za posamezno stranko so v sistemu zbrani osebni podatki, podatki o poslih, pozitivni dogodki, ki vplivajo na potek posla in negativni dogodki, ki niso v skladu s pogodbenimi obveznostmi.

Podobno so za pravne osebe v pomoč različne bonitetne agencije, ki z lastnimi modeli ocenjujejo stranke na podlagi bilančnih podatkov in profitabilnosti, obstajajo pa tudi organizacije kot je Agencija Republike Slovenije za javnopravne evidence in storitve (AJPES), ki je s strani države pooblaščena agencija za vodenje javnih registrov, kot so finančni podatki podjetij.

Na podlagi vseh pridobljenih informacij upravljalci tveganj ocenijo tveganost stranke in ji določijo bonitetno oceno, ki ima vlogo indikatorja kreditnega tveganja.

Banka se nato odloči, ali je pod dogovorjenimi pogoji, kot so višina posojila, obdobje in način odplačevanja ter cena posojila, posojilo pripravljena odobriti. Poleg naštetih pogojev je pomembno skleniti tudi dogovor o zavarovanju posojila. Če se s predlaganimi pogoji stinjata tako banka kot kreditojemalec, se posel dokončno sklene.

Po sklenitvi posojila se ta uvrsti v bančni portfelj sredstev. Sledi obdobje poplačila obveznosti, v katerem večina strank sčasoma vse obveznosti poplača v celoti, nekatere stranke pa se soočijo z nezmožnostjo izpolnjevanja kreditnih obveznosti. Takšne stranke se predajo oddelkom za izterjavo, katerih cilj je doseči čim večje poplačilo zapadlih terjatev.

2.2.2 Razlogi za merjenje kreditnega tveganja

Merjenje kreditnega tveganja glede na [4] koristi pri naslednjih področjih:

- **Podpora odobritvi posojila**

  V postopku odobritve posojila se je pomembno vprašati, ali se novo posojilo izplača ob danem tveganju in ceni posojila? Če vprašanje preoblikujemo, se sprašujemo, ali je pričakovana donosnost posla večja od minimalne zahtevane donosnosti kapitala? Da bi odgovorili na to vprašanje, moramo poznati pričakovan donos, ki upošteva tudi pričakovano izgubo in pričakovane stroške ter višino kapitala, ki ga bo novi posel obremenil.

  Podobno vprašanje je, kakšna cena posojila je ustrezna za dano tveganje, da je posojilo še dobičkonosno? Gre za ravno obraten pristop, pri katerem se najprej oceni obremenitev kapitala, s čimer se določi minimalna zahtevana donosnost kapitala, šele nato sledi izračun minimalne zahtevane donosnosti posojila.

- **Podpora optimizaciji portfelja**

V postopku optimiziranega upravljanja portfelja je naloga upravljalca tveganj minimizirati stopnjo tveganosti portfelja, zato mora dobro poznati koncentracijo in razpršenost tveganj. Za pridobitev teh informacij je potreben portfeljski model kreditnih tveganj, ki diagnostificira visoko koncentracijo močno koreliranih sredstev. Do močne korelacije prihaja zaradi skupnih lastnosti segmenta portfelja, kot sta ista industrija ali regija, ali ker na del sredstev vpliva isti ekonomski faktor, na primer cena goriva.

Portfeljski model mora prikazovati trenutna tveganja koncentracije in omogočati testiranje strategij za razpršitev portfelja.

- **Podpora upravljanju kapitala**

  Banka glede na tveganost portfelja določa ustrezne provizije za pričakovano izgubo v naslednjem letu in določa rezerve za primer, da bi bile izgube nenavadno visoke. Zagotoviti mora tudi zadostno višino kapitala, da banka ohrani ciljno bonitetno oceno. Če kapitala glede na tveganje v banki ni zadosti, mora banka pridobiti dodaten kapital, zmanjšati tveganja ali pa se ji poslabša bonitetna ocena.

  Za določitev ustreznih provizij in rezerv je potrebno ustrezno oceniti in napovedati višino izgube, ki jo banka lahko utrpi v nenavadno slabem letu.

2.2.3 Osnovni pojmi in faktorji kreditnega tveganja

Sledijo definicija neplačila in opisi treh ključnih faktorjev kreditnega tveganja, ki so potrebni za razumevanje nadaljevanja magistrskega dela.

- **Dogodek neplačila** (ang. default)

  Definicija dogodka neplačila je odvisna od namena uporabe pojma **neplačilo**, zato se definicije v literaturi razlikujejo. Navedimo nekaj primerov:

  - Dogodek neplačila označuje stanje stranke, v katerem trenutna nesposobnost rednega plačevanja obveznosti postane trajna nesposobnost rednega plačevanja. [8]
  - Da je prišlo do neplačila s strani dolžnika, se šteje, ko se zgodi eden ali oba od sledečih dogodkov [9]:
    a) Banka smatra, da obstaja majhna verjetnost, da bo dolžnik poravnal svoje obveznosti do bančne skupine v celoti, brez da bi se banka za poplačilo poslužila postopkov kot je npr. sprostitev ali koriščenje varščine.
    b) Dolžnik zaostaja s plačilom katerekoli pomembne obveznosti do bančne skupine več kot 90 dni. Prekoračenja se bodo smatrala kot zapadla, ko bo stranka prekoračila dovoljeno mejo (limit) ali ko bo obveščena o znižanju limita na raven, nižjo od vsote strankinih tekočih obveznosti.
  - V skladu z 178. členom Uredbe CRR [10] neplačilo nastopi, ko se zgodi eden ali oba od naslednjih dogodkov:

a) Institucija meni, da obstaja majhna verjetnost, da bo dolžnik poravnal svoje kreditne obveznosti do institucije, njene nadrejene osebe ali katere koli njej podrejene družbe v celoti, ne da bi institucija za poplačilo uporabila ukrepe, kot je unovčenje zavarovanja.
b) Dolžnik več kot 90 dni zamuja s plačilom katere koli pomembne kreditne obveznosti do institucije, njene nadrejene osebe ali katere koli njej podrejene družbe.

– Neplačilo pomeni, da stranka ni več sposobna poravnavati svojih obveznosti. Do takega stanja lahko pride zaradi vrste vzrokov, kot so zamuda pri plačilni obveznosti, prestrukuriranje posojilnih obveznosti zaradi poslabšanja bonitetne ocene dolžnika ali stečaj dolžnika;[8]

Večina literature obravnava dogodek neplačila kot enkratni slučajni dogodek, saj predpostavlja, da bodo banke svoje problematične naložbe prodale hitro po nastanku dogodka neplačila, prodajne cene pa definirajo končno poplačilo in posledično potencialno izgubo. Vendar pa mnoge banke svojih problematičnih naložb ne prodajo, zato je iz perspektive banke neplačilo dolgotrajnejši proces z večimi fazami.

V primeru, da ima problematična stranka potencial za prihodnost, se banka lahko strinja z prestrukturiranjem strankinih pogodbenih obveznosti [11], do česar pride zaradi dvoma, da stranka brez spremembe strukture posojila ne bo sposobna redno poravnavati obveznosti. Posledično restrukturiranje posojila upravičeno vodi v dogodek neplačila [8].

Potrebno je poudariti, da vsaka zamuda plačilnih obveznosti še ne pomeni dogodka neplačila. Večina zaostankov plačil se poravna znotraj obdobja treh mesecev in posledično ne pride do dogodka neplačila, saj je zamuda krajša od 90 dni [8].

Definicija dogodka neplačila je ključna za modeliranje faktorjev tveganj, katerih opisi sledijo v nadaljevanju, zato je potrebno uporabljati definicijo, ki je jasna in točna.

Izpostavljenost banke kreditnemu tveganju je glede na [6] odvisna od treh ključnih komponent kreditnega tveganja: višine izpostavljenosti, verjetnosti neplačila dolžnika, ki jo odraža njegova bonitetna ocena, in od stopnje poplačila dolga v primeru neplačila, ki je odvisna od zavarovanja in izterjave.

- **Izpostavljenost ob neplačilu (EAD)** (ang. exposure at default) izraža znesek izpostavljenosti banke v trenutku neplačila.

  Izziv pri modeliranju izpostavljenosti ob neplačilu predstavlja upoštevanje pričakovanih prihodnjih odlivov iz naslova neizkoriščenega dela odobrenega zneska posla, zaradi česar se izpostavljenost ob neplačilu povečuje.

- **Verjetnost neplačila (PD)** (ang. probability of default) izraža verjetnost neplačila nasprotne stranke v obdobju enega leta.

  Po 115. členu Zakona o bančništvu [3], je verjetnost neplačila enaka verjetnosti, da bo v naslednjem enoletnem obdobju nastopil položaj neplačila nasprotne pogodbene stranke, do česar pride, če po oceni banke obstaja majhna verjetnost, da bo nasprotna pogodbena stranka v celoti poravnala svoje kreditne obveznosti do banke, njene nadrejene družbe ali katerekoli njene podrejene družbe, ne da bi bilo za poplačilo treba uresničiti kreditno zavarovanje ali druge ukrepe prisilne izterjave te obveznosti ali če nasprotna pogodbena stranka več kot 90 dni zamuja s plačilom katerekoli pomembne kreditne obveznosti do banke, njene nadrejene družbe ali katerekoli njene podrejene družbe.

- **Izguba ob neplačilu (LGD)** (ang. loss given default) izraža delež izpostavljenosti, ki jo banka utrpi kot izgubo, če pri dolžniku pride do neplačila.

  Izguba ob neplačilu je po 55. točki 4. člena Uredbe CRR [10] definirana kot razmerje med izgubo na izpostavljenosti, ki je posledica neplačila nasprotne stranke, in vrednostjo izpostavljenosti ob neplačilu.

  Izguba ob neplačilu je po 115. členu Zakona o bančništvu [3] enaka razmerju med zneskom izgube pri posamezni izpostavljenosti banke zaradi neplačila nasprotne stranke in vrednostjo izpostavljenosti po stanju ob neplačilu; Pričakovana izguba je razmerje med zneskom, za katerega se pričakuje, da bo izgubljen pri posamezni izpostavljenosti banke zaradi morebitnega neplačila nasprotne stranke ali zaradi uresničitve tveganja zmanjšanja vrednosti odkupljene denarne terjatve v naslednjem enoletnem obdobju, in vrednostjo izpostavljenosti po stanju ob neplačilu.

Poglavje 3

Izgube

Ena izmed ključnih funkcij kapitala je zagotavljanje zaščite pred tveganji, natančneje, pokrivanje nepričakovanih izgub. Kapital ima pomemben vpliv na donosnost poslovanja banke, zato mora biti optimalno izkoriščen. Banke morajo zagotoviti, da ustrezno ocenjujejo tveganja ter potencialne izgube, katerim so izpostavljene, da si zagotovijo razpolaganje z urezno višino kapitala. Novi kapitalski sporazum prav iz tega razloga banke vspodbuja k uporabi naprednejših ter na tveganja občutljivejših metod za določanje potrebnega kapitala za pokrivanje kreditnega tveganja.

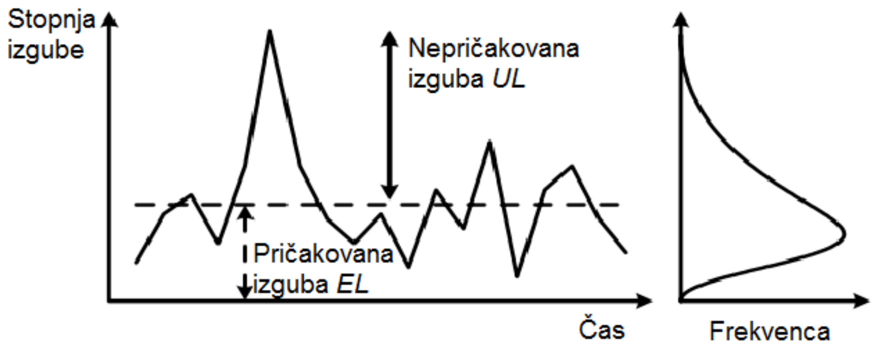
V nadaljevanju so opisane pričakovana, nepričakovana in ekstremna izguba. Sledi model za ocenjevanje pričakovane in nepričakovane izgube, v katerem kot najpomembnejši parametri nastopajo verjetnost neplačila (PD), izguba ob neplačilu (LGD) in izpostavljenost ob neplačilu (EAD), ki smo jih opisali že v prejšnjem poglavju. Prav ti trije parametri so za ocenjevanje kreditnega tveganja ključni tudi v pristopu notranjih bonitetnih sistemov, ki je opisan v nadaljevanju poglavja.

3.1 Pričakovana, nepričakovana in ekstremna izguba

Banka je preko tveganj izpostavljena izgubam, ki jih je težko predvideti. V veliko pomoč je časovna vrsta izgub iz preteklih let, ki za prihodnje obdobje omogoča napoved povprečne stopnje izgube, ki ji pravimo **pričakovana izguba (EL)** (ang. expected loss). Enaka je povprečni dejanski izgubi preteklih let prilagojeni z višino kreditne izpostavljenosti znotraj opazovanega obdobja. Banke pričakovano izgubo sprejemajo kot strošek poslovanja in jo v večini pokrivajo z oblikovanjem rezervacij.

Lahko se zgodi, da je dejanska izguba znotraj opazovanega obdobja, navadno enega leta, večja od ocenjene pričakovane izgube, zato takšni izgubi pravimo **nepričakovana izguba (UL)** (ang. unexpected loss). Do nepričakovanih izgub prihaja redko, višina izgube, ki jo banka lahko utrpi, pa je zelo visoka. Časovna vrsta razlik med pričakovanimi in nepričakovanimi izgubami na sliki 3.1 tvori porazdelitev verjetnosti izgub iz kreditnega tveganja v banki. Frekvenca pričakovanih izgub je višja od nepričakovanih izgub, saj se izgube v okviru pričakovanih izgub pojavljajo pogosteje kot nepričakovane izgube.

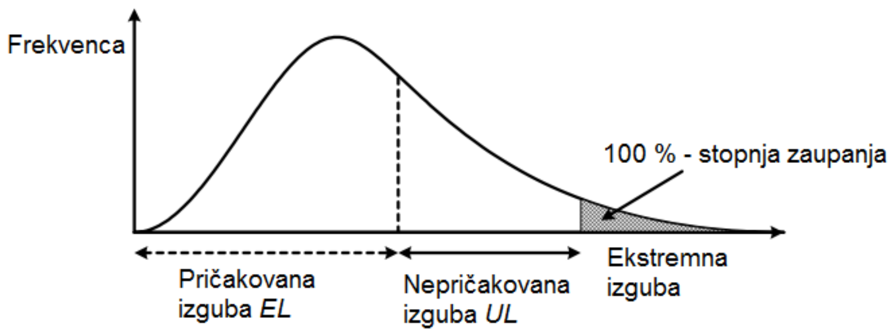
Glede na [15] je ena od glavnih funkcij kapitala absorbiranje nepričakovane izgube. Banke se zavedajo, da se nepričakovanam izgubam ni moč izogniti, težko pa je predvideti, kdaj bo do njh prišlo in v kakšnem obsegu. Obrestne mere, ki vključujejo pribitke za tveganje neplačila, lahko absorbirajo nekatere komponente nepričakovanih izgub, vendar pa obrestne mere, ki bi v celoti pokrivale nepričakovane izgube, na trgu niso konkurenčne. Kapital je torej potreben za kritje nepričakovanih izgub in ima zato pomembno funkcijo absorbiranja izgub.



Slika 3.1: Frekvenca pričakovanih in nepričakovanih izgub. Slika je povzeta iz [15].

Ključni vprašanji sta, kolikšen delež kapitala naj banka ob določeni stopnji zaupanja znotraj opazovanega časovnega obdobja zadržuje za namen pokrivanja izgub ter kolikšen del naj ga vloži v dobičkonosne naložbe? Zadrževanje prevelike količine kapitala ni dobičkonosno, premajhna količina zadržanega kapitala pa povečuje verjetnost, da izgub danega leta banka ne bo sposobna pokriti z dobičkom in razpoložljivim kapitalom, kar lahko privede do insolventnosti banke. Iz tega razloga morajo banke pazljivo uravnotežiti razmerje med ceno tveganja ter koristmi držanja kapitala.

Na sliki 3.2 je značilna porazdelitev verjetnosti izgub iz naslova kreditnega tveganja. Izgube, ki so po velikosti enake ali delno manjše od pričakovane izgube, se pojavljajo pogosteje kot pa večje izgube. Izgubi, ki presega vsoto pričakovane in nepričakovane izgube, pravimo **ekstremna izguba** in je malo verjetna. Na sliki 3.2 je senčena s črno barvo. Porazdelitev ima težek desni rep, kar pomeni, da se verjetnost ekstremnih izgub zmanjšuje zelo počasi.



Slika 3.2: Porazdelitev izgub. Slika je povzeta iz [15].

Obstaja ogromno pristopov za iskanje optimalnega deleža kapitala, ki naj ga banka zadržuje za namen pokrivanja izgub. Eden od njih je Pristop notranjih bonitetnih sistemov, ki je podrobneje predstavljen v razdelku 3.3. Sledi opis modela za ocenjevanje pričakovane in nepričakovane izgube, opisan v [4].

3.2 Ocenjevanje pričakovane in nepričakovane izgube

Porazdelitev izgube iz kreditnega tveganja je določena z matematičnim upanjem in standardnim odklonom izgub znotraj enega leta. Matematično upanje izgube je pogosto opisano s pričakovano izgubo EL, standardni odklon izgube pa z nepričakovano izgubo UL.

Naj bo dejanska izguba $Izguba$ enaka produktu izpostavljenosti ob neplačilu EAD, izgube ob neplačilu LGD in indikatorja neplačila I:

$$Izguba = EAD \cdot LGD \cdot I \tag{3.1}$$

pri čemer za I velja:

$$I = \begin{cases} 1 & \text{neplačilo se zgodi,} \\ 0 & \text{neplačilo se ne zgodi.} \end{cases} \tag{3.2}$$

Posledično za spremenljivko $Izguba$ velja:

$$Izguba = \begin{cases} EAD \cdot LGD & \text{če } I = 1, \\ 0 & \text{če } I = 0. \end{cases} \tag{3.3}$$

3.2.1 LGD in EAD sta konstantni

Naj bosta izguba ob neplačilu LGD in izpostavljenost ob neplačilu EAD konstantni. Potem je v izrazu (3.3) neznano le še, ali se neplačilo zgodi ali ne. Dva možna izida ob koncu leta sta:

- z verjetnostjo p se neplačilo zgodi
- z verjetnostjo $1 - p$ se neplačilo ne zgodi

Pričakovana izguba EL je potem naslednja:

$$\begin{aligned} EL &= p[1 \cdot EAD \cdot LGD] + (1-p)[0 \cdot EAD \cdot LGD] \\ &= p \cdot EAD \cdot LGD \end{aligned} \tag{3.4}$$

Tudi pri izračunu nepričakovane izgube UL naj bosta izguba ob neplačilu LGD in izpostavljenost ob neplačilu EAD konstantni. Indikator neplačila I je diskretna spremenljivka, zato je varianca nepričakovane izgube enaka izračunu (3.5), v katerem z $\overline{L}$ označujemo izraz za EL iz (3.4).

$$\begin{aligned} UL^2 &= p[1 \cdot EAD \cdot LGD - \overline{L}]^2 + (1-p)[0 \cdot EAD \cdot LGD - \overline{L}]^2 \\ &= p \cdot [EAD^2 \cdot LGD^2 - 2EAD \cdot LGD \cdot \overline{L} + \overline{L}^2] + (1-p) \cdot [\overline{L}]^2 \\ &= p \cdot [EAD^2 \cdot LGD^2 - 2EAD \cdot LGD \cdot \overline{L} + \overline{L}^2] + [\overline{L}]^2 - p \cdot [\overline{L}]^2 \\ &= p \cdot [EAD^2 \cdot LGD^2 - 2EAD \cdot LGD \cdot \overline{L}] + [\overline{L}]^2 \\ &= p \cdot [EAD^2 \cdot LGD^2 - 2EAD \cdot LGD \cdot p \cdot EAD \cdot LGD] + [p^2 \cdot EAD^2 \cdot LGD^2] \\ &= p \cdot EAD^2 \cdot LGD^2 - p^2 \cdot EAD^2 \cdot LGD^2 \\ &= [p - p^2][EAD \cdot LGD]^2 \end{aligned} \tag{3.5}$$

S preoblikovanjem enakosti (3.5) dobimo izraz za nepričakovano izgubo UL:

$$UL = \sqrt{p - p^2} \cdot EAD \cdot LGD \tag{3.6}$$

3.2.2 LGD in EAD nista konstantni

Predpostavimo, da spremenljivki izguba ob neplačilu LGD in izpostavljenost ob neplačilu EAD nista konstantni. Potem je v opazovanem primeru pričakovana izguba enaka matematičnemu upanju v vseh fazah, pomnožena z verjetnostjo stanja:

$$\begin{aligned} EL &= \sum_{I=0,1} p(I) \int_{LGD} \int_{EAD} I \cdot LGD \cdot EAD \cdot pr(LGD, EAD|I) \cdot dLGD \cdot dEAD \\ &= p \cdot \int_{LGD} \int_{EAD} LGD \cdot EAD \cdot pr(LGD, EAD|I=1) \cdot dLGD \cdot dEAD + \\ &\quad + (1-p) \cdot 0 \\ &= p \cdot \int_{LGD} \int_{EAD} LGD \cdot EAD \cdot pr(LGD, EAD|I=1) \cdot dLGD \cdot dEAD \end{aligned} \tag{3.7}$$

Kjer:

- s $p(I)$ označujemo verjetnost, da ima indikator I vrednost 0 ali 1,
- s $pr(LGD, EAD|I)$ označujemo pogojno skupno porazdelitev slučajnih spremenljivk izgube ob neplačilu LGD in izpostavljenosti ob neplačilu EAD pri pogoju indikatorja I.

V resnici nas zanima le pogoj, ko je indikator $I = 1$, kar pomeni, da nas zanimata izguba ob neplačilu LGD in izpostavljenost ob neplačilu EAD v primeru nastanka neplačila.

Opazimo, da je zveza znotraj dvojnega integrala v izrazu (3.7) podobna definiciji kovariance med LGD in EAD:

$$\begin{aligned} &\int_{LGD} \int_{EAD} LGD \cdot EAD \cdot pr(LGD, EAD) \cdot dLGD \cdot dEAD \\ = \; &\sigma^2_{EAD,LGD} + \overline{EAD} \cdot \overline{LGD} \end{aligned} \tag{3.8}$$

Z upoštevanjem zveze (3.8) lahko poenostavimo enakost (3.7):

$$EL = p \cdot (\overline{EAD} \cdot \overline{LGD} + \sigma^2_{EAD,LGD}) \tag{3.9}$$

Ta rezultat je enak kot rezultat za EL v enakosti (3.4), le da je upoštevana še kovarianca med izpostavljenostjo ob neplačilu EAD in izgubo ob neplačilu LGD. Če odvisnost med spremenljivkama EAD in LGD ni jasna, vseeno pa je njuna korelacija enaka 0, potem je pričakovana izguba enaka:

$$EL = p \cdot \overline{EAD} \cdot \overline{LGD} \tag{3.10}$$

Sledi še izračun nepričakovane izgube UL.

V prvem primeru, v katerem odvisnost med spremenljivkama EAD in LGD ni jasna, je varianca izgube UL^2 enaka kvadratu standardnega odklona izgube v vsakem stanju, pomnožena z verjetnostjo stanja. Pričakovano izgubo ponovno označimo z $\overline{L}$.

$$
\begin{aligned}
UL^2 &= \sum_{I=0,1} p(I) \int_{LGD} \int_{EAD} (I \cdot LGD \cdot EAD - \overline{L})^2 \cdot \\
&\cdot pr(LGD, EAD|I) \cdot dLGD \cdot dEAD
\end{aligned} \tag{3.11}
$$

Pričakovana izguba $\overline{L}$ je enaka konstanti, zato jo lahko nesemo pred dvojni integral in vsoto:

$$
\begin{aligned}
UL^2 &= \Big(\sum_{I=0,1} p(I) \int_{LGD} \int_{EAD} (I \cdot LGD \cdot EAD)^2 \cdot \\
&\cdot pr(LGD, EAD|I) \cdot dLGD \cdot dEAD \Big) - \\
&- \Big(2 \cdot \overline{L} \sum_{I=0,1} p(I) \int_{LGD} \int_{EAD} (I \cdot LGD \cdot EAD) \cdot \\
&\cdot pr(LGD, EAD|I) \cdot dLGD \cdot dEAD \Big) + \\
&+ \Big(\overline{L}^2 \sum_{I=0,1} p(I) \int_{LGD} \int_{EAD} pr(LGD, EAD|I) \cdot dLGD \cdot dEAD \Big) \\
&= \Big(\sum_{I=0,1} p(I) \int_{LGD} \int_{EAD} (I \cdot LGD \cdot EAD)^2 \cdot \\
&\cdot pr(LGD, EAD|I) \cdot dLGD \cdot dEAD \Big) - (\overline{L}^2)
\end{aligned} \tag{3.12}
$$

Sedaj lahko upoštevamo še verjetnost indikatorja I. Ko je $I = 0$, je izraz znotraj integralov enak 0.

$$
\begin{aligned}
UL^2 &= \Big((1-p) \cdot 0 + p \cdot \int_{LGD} \int_{EAD} (LGD \cdot EAD)^2 \cdot \\
&\cdot pr(LGD, EAD|I=1) \cdot dLGD \cdot dEAD \Big) - (\overline{L}^2) \\
&= \Big(p \cdot \int_{LGD} \int_{EAD} (LGD \cdot EAD)^2 \cdot \\
&\cdot pr(LGD, EAD|I=1) \cdot dLGD \cdot dEAD \Big) - (\overline{L}^2)
\end{aligned} \tag{3.13}
$$

Dvojni integral v enakosti (3.13) je dokaj težko poenostaviti, razen, če predpostavimo, da sta spremenljivki LGD in EAD neodvisni, kar bi pomenilo, da je višina izgube neodvisna od izpostavljenosti ob neplačilu. V tem primeru lahko verjetnostno porazdelitev razdelimo v produkt dveh porazdelitev, vsako pa lahko integriramo posebej:

$$
\begin{aligned}
UL^2 &= \Big(p \cdot \int_{LGD} \int_{EAD} (LGD \cdot EAD)^2 \cdot pr(LGD) \cdot pr(EAD) \cdot \\
&\cdot dLGD \cdot dEAD \Big) - (\overline{L}^2)
\end{aligned}
$$

$$
\begin{aligned}
&= \Big(p \cdot \int_{LGD} LGD^2 \cdot pr(LGD) \cdot dLGD \cdot \\
&\quad \cdot \int_{EAD} EAD^2 \cdot pr(EAD) \cdot dEAD\Big) - (\overline{L}^2) \qquad (3.14)
\end{aligned}
$$

Opazimo, da so integrali skoraj enaki formuli za varianco, zato zvezo (3.14) preoblikujemo:

$$
\begin{aligned}
UL^2 &= p \cdot \left(\sigma_{LGD}^2 + \overline{LGD}^2\right)\left(\sigma_{EAD}^2 + \overline{EAD}^2\right) - (\overline{L}^2) \\
&= p \cdot \left(\sigma_{LGD}^2 + \overline{LGD}^2\right)\left(\sigma_{EAD}^2 + \overline{EAD}^2\right) - \\
&\quad - (p^2 \cdot \overline{LGD}^2 \cdot \overline{EAD}^2) \\
&= (p - p^2) \cdot \overline{LGD}^2 \cdot \overline{EAD}^2 + \\
&\quad + p \cdot \left(\sigma_{LGD}^2 \cdot \overline{LGD}^2 + \sigma_{EAD}^2 \cdot \overline{EAD}^2 + \sigma_{LGD}^2 \cdot \sigma_{EAD}^2\right) \qquad (3.15)
\end{aligned}
$$

Rezultat (3.15) je podoben rezultatu iz (3.6), le da je upoštevana še kovarianca med izpostavljenostjo ob neplačilu EAD in izgubo ob neplačilu LGD.

3.2.3 Primer izračuna

Za lažjo predstavo si oglejmo izračun pričakovane in nepričakovane izgube še na primeru, ko banka kreditira podjetje z bonitetno oceno BB, za katero velja, da je verjetnost neplačila p enaka 22 bazičnih točk.

1. Po modelu iz 3.2.1 sta izguba ob neplačilu (LGD) in izpostavljenost ob neplačilu (EAD) konstantni, zato predpostavimo, da je izguba ob neplačilu LGD enaka 30 %, izpostavljenost ob neplačilu pa 100.000 EUR. Potem je pričakovana izguba z uporabo enakosti (3.4) enaka:

$$
\begin{aligned}
EL &= p \cdot EAD \cdot LGD \\
&= 0,0022 \cdot 100.000 \cdot 0,3 \\
&= 66\, EUR \qquad (3.16)
\end{aligned}
$$

   Nepričakovana izguba pa je z uporabo izraza (3.6) enaka:

$$
\begin{aligned}
UL &= \sqrt{p - p^2} \cdot EAD \cdot LGD \\
&= \sqrt{0,0022 - 0,0022^2} \cdot 100.000 \cdot 0,3 \\
&= 1.410\, EUR \qquad (3.17)
\end{aligned}
$$

2. Po modelu 3.2.2 sta izguba ob neplačilu (LGD) in izpostavljenost ob neplačilu (EAD) zvezni spremenljivki. Če v enačbi 3.9 upoštevamo, da je kovarianca med LGD in EAD enaka 0, potem je EL kar enak enačbi 3.10, kjer sta $\overline{EAD}$ in $\overline{LGD}$ kar enaka matematičnem upanju, naj bo to 30 % in 100.000 EUR.

$$\begin{aligned} EL &= p \cdot (\overline{EAD} \cdot \overline{LGD} + \sigma^2_{EAD,LGD}) \\ &= p \cdot \overline{EAD} \cdot \overline{LGD} \\ &= 0,0022 \cdot 100.000 \cdot 0,3 \\ &= 66\, EUR \qquad (3.18) \end{aligned}$$

Če predpostavimo, da je varianca izpostavljenosti ob neplačilu EAD za posojila blizu 0, je nepričakovana izguba enaka:

$$\begin{aligned} UL^2 &= (p - p^2)\overline{LGD}^2\,\overline{EAD}^2 + \\ &\quad + p \cdot \left(\sigma^2_{LGD}\overline{LGD}^2 + \sigma^2_{EAD}\overline{EAD}^2 + \sigma^2_{LGD}\sigma^2_{EAD}\right) \\ &= (p - p^2)\overline{LGD}^2\,\overline{EAD}^2 + p \cdot \sigma^2_{LGD}\overline{LGD}^2 \qquad (3.19) \end{aligned}$$

Z upoštevanjem ocene iz [4], da standardni odklon izgube ob neplačilu LGD v povprečju znaša 25 %, je izračun nepričakovane izgube sledeč:

$$\begin{aligned} UL &= \sqrt{((0,0022 - 0,0022^2)(0,3^2 \cdot 100.000^2) + 0,0022 \cdot (0.25^2 \cdot 100.000^2))} \\ &= 1.830\, EUR \qquad (3.20) \end{aligned}$$

Rezultat nepričakovane izgube v (3.20) je ravno 30 % povečanje glede na rezultat v (3.17), kjer smo predpostavili, da je izguba ob neplačilu LGD fiksna.

3.3 Pristop notranjih bonitetnih sistemov

Novi baselski sporazum strmi k izboljšanju ocenjevanja tveganj ter banke vzpodbuja k uporabi na tveganja občutljivejših metod za določanje potrebnega kapitala za pokrivanje kreditnega tveganja. Z uporabo Pristopa notranjih bonitetnih sistemov (v nadaljevanju IRB pristop) bodo banke z manj tveganim portfeljem in učinkovitimi sistemi za obvladovanje kreditnega tveganja občutile kapitalske olajšave, kar je ena izmed glavnih spodbud za uporabo in nadaljnji razvoj tega pristopa.

V nadaljevanju so predstavljeni cilji in motivacija za uporabo IRB pristopa, izračun pričakovane izgube v skladu z IRB pristopom na ravni posameznega posla ter na ravni portfelja ter pričakovane posledice uporabe IRB pristopa.

3.3.1 Cilji in motivacija za uporabo IRB pristopa

Novi baselski sporazum glede na [12] omogoča, da banke smejo uporabljati ali **standardiziran pristop**, pri katerem so vse ključne komponente kreditnega tveganja določene s strani nadzornika, ali eno od dveh različ **Pristopa notranjih bonitetnih sistemov (IRB pristop)**, pri katerem banke smejo uporabljati svoje lastne interne modele za ocenjevanje ključnih dejavnikov tveganj v okviru IRB pristopa, to so verjetnost neplačila (PD), izpostavljenost ob neplačilu (EAD), izguba ob neplačilu (LGD) ter zapadlost (M).

Glavna razlika med IRB pristopoma je v načinu pridobitve ocen ključnih dejavnikov tveganj. Banke, ki uporabljajo **osnovni IRB pristop**, smejo uporabljati lastni model za ocenjevanje verjetnosti neplačila, ostale faktorje tveganj pa določi nadzornik - deleži izgube ob neplačilu (LGD) so vnaprej določeni, prav tako izpostavljenost ob neplačilu (EAD), če ta ni znana. Banke, ki uporabljajo **napredni IRB pristop**, poleg lastnega modela za ocenjevanje verjetnosti neplačila (PD), smejo uporabljati še lastne modele za ocenjevanje izgube ob neplačilu (LGD) ter izpostavljenosti ob neplačilu (EAD). Vsi interni modeli za ocenjevanje ključnih dejavnikov tveganj morajo biti predhodno potrjeni s strani nadzornikov.

Osnovni cilj IRB pristopa je zajeti dejansko ekonomsko tveganje posameznih naložb ali poslov banke. Temelji na lastnem sistemu internih bonitetnih ocen oz. sistemu razvrščanja komitentov, ki omogoča diferenciacijo kreditnega tveganja in s tem posledično natančnejši izračun kapitalskih zahtev za kreditno tveganje. (Čargo, Štajner [13], 2004)

Ena od spodbud za uporabo občutljivejših metod je pričakovanje, da bo tak pristop pri bankah z manj tveganim portfeljem in učinkovitimi sistemi za obvladovanje kreditnega tveganja prispeval h kapitalskim olajšavam. (Čargo, Štajner, 2004)

Glavna motivacija za uporabo IRB pristopa je izpostavljena v članku [14]. Ocene določene z internimi modeli so prilagojene lastnostim bančnega portfelja in so posledično za banko ustreznejše kot ocene določene s strani nadzornikov. Prav možnost uporabe lastnih ocen banke motivira k izdelavi internih modelov za ocenjevanje ključnih parametrov in posledično k izboljšanju izkaza poslovnega izida, s tem pa tudi k uporabi naprednega IRB pristopa za izračun kapitalske zahteve za kreditno tveganje. (Mićković in Kern, 2015)

3.3.2 Pričakovana izguba na ravni posameznega posla

Pričakovano izgubo lahko definiramo s pristopom »od spodaj navzgor«, ki je opisan v članku [15]. Enaka je produktu naslednjih parametrov:

- deleža dolžnikov, ki bi v roku enega leta lahko postali neplačniki,
- neporavnane izpostavljenostjo ob neplačilu,
- delež izgube ob neplačilu (t.j. delež kreditne izpostavljenosti, ki ne bo poplačana z unovčitvijo zavarovanja kreditnega posla);

Teh parametrov v tekočem letu ni možno natančno meriti, saj gre za slučajne spremenljivke, njihove vrednosti pa se ocenjujejo na podlagi časovne vrste podatkov. Vsi trije parametri se ujemajo s ključnimi dejavniki tveganj po IRB pristopu, to so verjetnost neplačila PD, izpostavljenost ob neplačilu EAD in izguba ob neplačilu LGD, zato se znesek pričakovane izgube EL izračuna s produktom:

$$EL = PD \cdot LGD \cdot EAD \tag{3.21}$$

Na podlagi Sklepa o izračunu kapitalske zahteve za kreditno tveganje po pristopu IRB [5] se delež pričakovane izgube za posamezni kreditni posel izračuna kot produkt verjetnosti neplačila in izgube ob neplačilu (izražene v odstotku) po formuli:

$$EL = PD \cdot LGD \tag{3.22}$$

Znesek pričakovane izgube za posamezni kreditni posel se izračuna kot produkt pričakovane izgube in izpostavljenosti ob neplačilu.

3.3.3 Pričakovana izguba na ravni portfelja

Pričakovana izguba EL na ravni portfelja se izračuna kot vsota pričakovanih izgub posameznih poslov po formuli:

$$EL_{portfelj} = \sum_i EL_i \qquad (3.23)$$

Znesek pričakovane izgube na ravni portfelja pa se izračuna kot vsota zneskov pričakovanih izgub vseh kreditnih poslov portfelja.

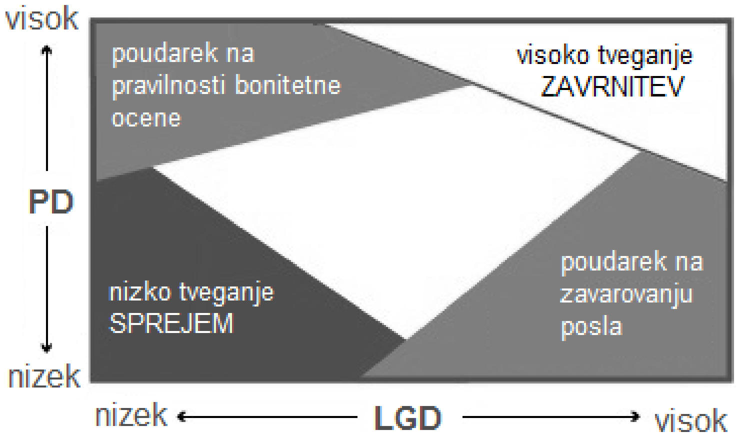
Na pričakovano izgubo banke gledajo kot na strošek poslovanja in jo pokrivajo preko oblikovanja cen za kreditno izpostavljenost in prek rezervacij. Zavedajo se, da lahko pride tudi do nepričakovanih izgub, vendar je težko oceniti, kdaj bo do njih prišlo in v kakšnem obsegu. Obrestne mere, ki bi v celoti pokrivale nepričakovane izgube, bi bile previsoke, zato ima pri pokrivanju izgub, ki presegajo pričakovano izgubo, pomembno vlogo kapital banke.

3.3.4 Učinki uporabe IRB pristopa

Gradivo Banke Slovenije [16] obravnava pričakovane učinke Novega baselskega sporazuma, s katerim so regulatorji želeli zagotoviti sisteme učinkovitejšega upravljanja tveganj. Že pred uveljavitvijo novih kapitalskih pravil je bilo na podlagi kvantitativne študije pričakovano, da bodo kapitalske zahteve izračunane s preprostejšo različico IRB pristopa višje od kapitalskih zahtev izračunanih s standardiziranim pristopom. S tem razlogom so se številne banke odločile za lastne analize pričakovanih učinkov Novega baselskega sporazuma. Rezultati analiz so pokazali, da so s prehodom na IRB pristop pričakovani prihranki iz naslova presežnega kapitala. Največji izziv za prehod na nove pristope je zagotovitev ustrezne podatkovne osnove, saj sporazum določa zahtevane dolžine časovnih vrst za ocenjevanje verjetnosti neplačila, izgube ob neplačilu in stopnje poplačila, ki so ključni v izračunih po IRB pristopu.

Gradivo Banke Slovenije [16] opisuje pričakovane posledice, ki jih prinaša novi kapitalski sporazum:

- nivo regulatornega kapitala naj bi bil bliže dejanskemu ekonomskemu tveganju bank,
- kapital v posameznih bankah in med njimi naj bi se prerazporedil glede na tveganost njihovega poslovanja,
- agregatna raven kapitala v bančnih sistemih naj ne bi padla pod obstoječo raven,
- procesi obvladovanja tveganj v bankah naj bi se priznavali v večji meri kot doslej,
- spodbude za kapitalsko arbitražo naj bi se zmanjšale;



Slika 3.3: Razlikovanje med dobrimi in slabimi posojili.

Verjetnost neplačila (PD) in stopnja izgube ob neplačilu (LGD) sta pomembna podatka, ki ju banka lahko uporablja tudi kot usmeritev pri oblikovanju politike kreditiranja. Oba kreditna faktorja pripomoreta k učinkovitem rezlikovanju med dobrimi in slabimi posojili ter opozarjata na potencialne nevarnosti, kar prikazuje slika 3.3.

- Nizko kreditno tveganje predstavljajo stranke, ki imajo nizko verjetnost neplačila (PD) ter nizko pričakovano izgubo (LGD). Gre za tveganje, ki ga je banka pripravljena sprejeti.
- Če je verjetnost neplačila nizka, izguba ob neplačilu pa je visoka, se je potrebno prepričati o pravilni bonitetni oceni dolžnika.
- Če je odstotek izgube ob neplačilu nizek, verjetnost neplačila pa je visoka, se je potrebno osredotočiti na vrednost zavarovanj ter možnost njihovega unovčenja.
- Če sta vrednosti obeh faktorjev visoki, gre za visoko kreditno tveganje, ki ga banka ni pripravljena sprejeti.

3.3.5 Dovoljenje za uporabo IRB pristopa

Banke morajo za uporabo IRB pristopa pridobiti dovoljenje s strani pristojnega organa, ki dovoljenje izda, če presodi, da so izpolnjeni standardi iz 144. člena Uredbe CRR [10]:

- bonitetni sistemi institucije omogočajo smiselno oceno značilnosti dolžnika in izpostavljenosti, smiselno razlikovanje tveganj ter natančne in dosledne kvantitativne ocene tveganj,
- notranje bonitetne ocene ter ocene neplačila in izgube, ki se uporabljajo pri izračunu kapitalskih zahtev, ter z njimi povezani sistemi in procesi so ključni pri upravljanju tveganj in postopku odločanja ter pri odobritvi kreditov, razporeditvi notranjega kapitala in funkcijah korporativnega upravljanja,

- banka ima enoto za nadzor kreditnega tveganja, pristojno za bonitetne sisteme, ki je ustrezno neodvisna in ni izpostavljena neprimernim vplivom,
- institucija zbira in hrani vse pomembne podatke za zagotavljanje učinkovite podpore svojemu procesu merjenja in upravljanja kreditnega tveganja,
- banka dokumentira svoje bonitetne sisteme in načelo za njihovo sestavo ter svoje bonitetne sisteme ovrednoti,
- banka je pristop notranjih modelov za izpostavljenosti iz naslova lastniških instumentov ovrednotila v ustreznem časovnem obdobju pred izdajo dovoljenja za uporabo pristopa, pri čemer je ocenila, ali so ti pristopi primerni,
- banka je v skladu s pristopom IRB izračunala kapitalske zahteve na podlagi svojih ocen parametrov tveganja in jih je sposobna poročati,
- banka je razvrstila in še naprej razvršča vsako izpostavljenost iz obsega uporabe bonitetnega sistema v bonitetni razred ali skupino tega bonitetnega sistema;

Poleg naštetih standardov morajo biti za uporabo modela LGD izpolnjeni še pogoji iz 174. člena Uredbe CRR [10]. Če banka uporablja statistične modele, morajo biti izpolnjene naslednje zahteve:

- model ima dobro napovedno moč in kapitalske zahteve zaradi njegove uporabe niso izkrivljene. Vhodne spremenljivke predstavljajo razumno in učinkovito osnovo za napovedovanja. Model je v čim večji meri nepristranski;
- banka vzpostavi proces za vrednotenje vhodnih podatkov v model, ki vključuje oceno natančnosti, popolnosti in primernosti podatkov;
- podatki, uporabljeni za izgradnjo modela, so reprezentativni za populacijo dejanskih dolžnikov ali izpostavljenosti institucije;
- banka vzpostavi redni ciklus preverjanja veljavnosti modela, ki vključuje spremljanje delovanja in stabilnosti modela, pregled specifikacij modela in testiranje pridobljenih rezultatov modela glede na dejanske izide;
- banka dopolnjuje statistični model s človekovo presojo in človekovim nadzorom in na ta način pregleduje razvrstitve na podlagi modela ter zagotavlja pravilno uporabo modelov. Cilj postopkov pregledovanja je ugotavljanje in omejevanje napak, povezanih s šibkostmi modela. Človekova presoja upošteva vse pomembne informacije, ki jih model ne obravnava. Institucija dokumentira, kako naj bodo človekova presoja in rezultati modela medsebojno povezani;

V skladu s 145. členom Uredbe CRR [10] mora banka, ki zaprosi za dovoljenje za uporabo IRB pristopa, pred zaprositvijo za dovoljenje, najmanj tri leta uporabljati bonitetne sisteme in lastne ocene LGD ter konverzijskih faktorjev, ki so vsaj okvirno skladni z zgornjimi pogoji iz 174. člena Uredbe CRR [10], za notranje namene merjenja in upravljanja tveganja, preden lahko postane upravičena do uporabe pristopa IRB.

Vse podrobnosti o delovanju svojih bonitetnih sistemov mora banka skrbno dokumentirati. Če banka uporablja statistične modele, mora vse uporabljene metodologije dokumentirati ter jih predložiti pristojnemu organu, ki odloča o ustreznosti modelov. Dokumentacija mora po 175. členu Uredbe CRR [10] zajemati natančen opis teorije, predpostavk ter matematične in empirične osnove za dodelitev ocen bonitetnim razredom, posameznim dolžnikom, izpostavljenostim ali skupinam. Določati mora natačen statističen proces za ovrednotenje modela, vključno s testom delovanja statističnega modela zunaj časa in zunaj vzorca, v katerem je bil generiran razvojni vzorec, ter navaja okoliščine, v katerih model ne deluje učinkovito.

V skladu s 144. členom Uredbe CRR [10] mora banka pridobiti dovoljenje za uporabo IRB pristopa, tudi, kadar je banka model kupila od zunanjega prodajalca. V skladu s 175. členom Uredbe CRR [10] je razumljivo, da so lahko metodologija, predpostavke in druge informacije o modelu poslovna skrivnost prodajalca, vendar mora banka pristojnemu organu vseeno na zadovoljiv način dokazati, da so zgornje zahteve izpolnjene.

Poglavje 4

Ocenjevanje izgube ob neplačilu (LGD)

Izguba ob neplačilu se najpogosteje ocenjuje kot razmerje med oceno dejanske izgube in kreditno izpostavljenostjo v trenutku nastanka neplačila. Dejanska izguba je odvisna od denarnih tokov v obdobju po nastanku neplačila, pravimo mu obdobje poplačevanja. V tem obdobju prihaja tako do poplačil obveznosti kot do dodatnih stroškov iz naslova izterjave. Vsi denarni tokovi vplivajo na višino izgube tako na ravni posamezne terjatve kot na ravni celotnega portfelja.

Sledi opis dinamike denarnih tokov, vključno s časovnico stečaja, ki pomembno vpliva na dinamiko denarnega toka. V nadaljevanju je opisana stopnja poplačila, ki je enaka komplimentu izgube ob neplačilu. Sledita opisa postopkov izračuna izgube ob neplačilu na ravni posamezne terjatve in na ravni portfelja. Nato sledi še opis potrebnih podatkov za modeliranje izgube ob neplačilu.

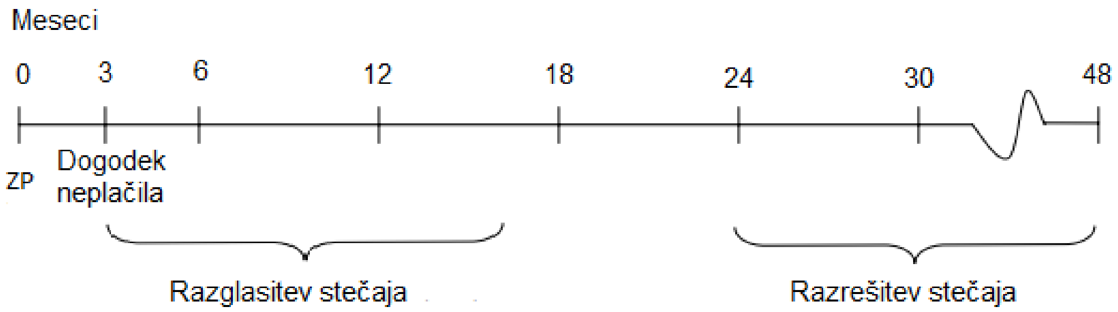
4.1 Dinamika denarnih tokov

Na višino izgube ob neplačilu pomembno vplivajo plačila in stroški, do katerih pride po nastanku dogodka neplačila, zato je pomembno poznati dinamiko denarnih tokov iz obdobja poplačila. Eden izmed najpogostejših razlogov neplačila je stečaj podjetja, zato za boljše razumevanje dinamike denarnih tokov predstavimo časovnico stečaja podjetja.

4.1.1 Časovnica stečaja

Časovnica stečaja se začne z zadnjim plačilom obveznosti, ki je na sliki 4.1 označeno z okrajšavo ZP. Po definiciji dogodka neplačila do neplačila pride, ko dolžnik zaostaja s plačilom obveznosti več kot 90 dni, kar je približno 3 mesece. V obdobju enega leta po nastanku dogodka neplačila se razglasi stečaj podjetja, datum razglasitve pa je odvisen tudi od pogajanj med podjetjem in njegovimi upniki. Nato sledijo postopki likvidacije, v roku od dveh do štirih let pa se stečajni postopek razreši.

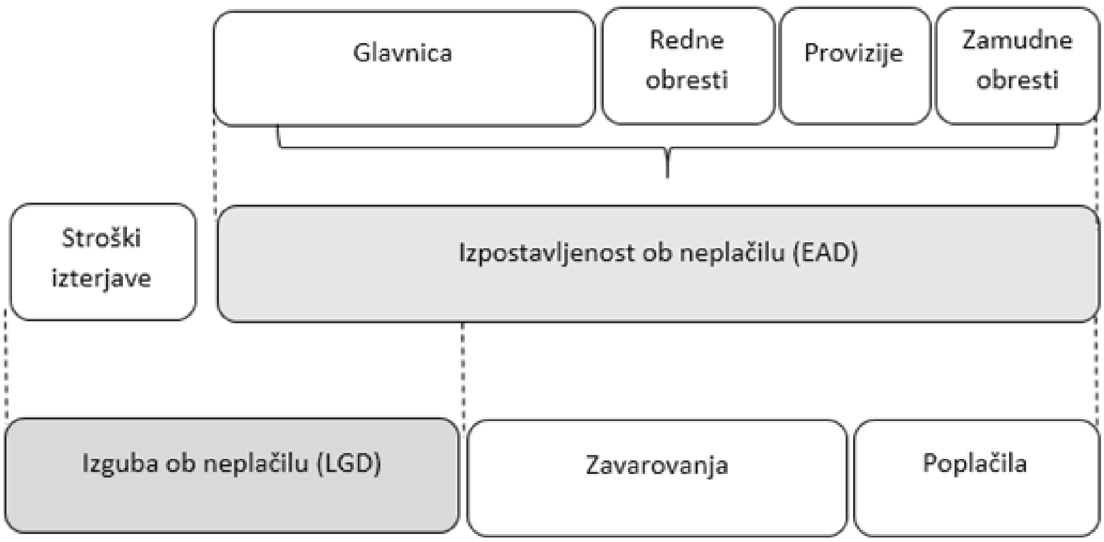
Na izgubo ob neplačilu močno vpliva trajanje razrešitve stečajnega postopka, saj se z daljšanjem tega obdobja zmanjšuje vrednost poplačil. Po raziskavah, omenjenih v [12], povprečno obdobje stečajnega postopka traja dve leti, kar je tudi upoštevano na sliki 4.1.



Slika 4.1: Časovnica stečaja. Slika je povzeta iz [12].

4.1.2 Denarni tokovi ob neplačilu

Na višino izpostavljenosti ob neplačilu (EAD) vplivajo plačila pred nastankom dogodka neplačila. Izpostavljenost ob neplačilu (EAD) je enaka vsoti preostale glavnice, rednih obresti, provizij ter zamudnih obresti, kar prikazuje slika 4.2. Daljše kot je obdobje redne poravnave obveznosti pred nastankom dogodka neplačila, nižja je izpostavljenost ob neplačilu v primerjavi z odobrenim zneskom posojila.



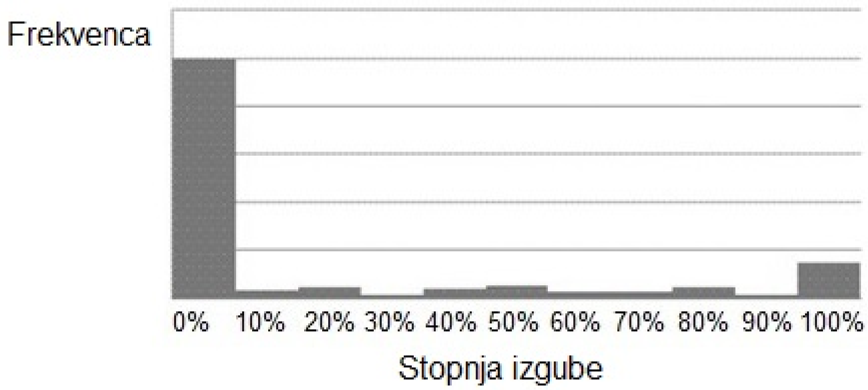
Slika 4.2: Dinamika denarnih tokov v povezavi z izpostavljenostjo ob neplačilu (EAD) in izgub ob neplačilu (LGD). Slika je povzeta iz [14].

Po nastopu dogodka neplačila se prične obdobje poplačevanja, kamor spada tudi postopek izterjave, ki za banko predstavlja dodaten strošek. Dolg iz naslova izpostavljenosti ob neplačilu in stroškov izterjave je lahko poplačan s plačili stranke ali z unovčitvijo zavarovanj. V zadnjem času se ravno iz tega razloga kvaliteti zavarovanj daje vse večji pomen.

4.1.3 Porazdelitev izgub ob neplačilu

V obdobju poplačevanja sta najpogostejša primera, ko stranka dolg do banke poplača v celoti, in zato banka ne utrpi izgube, ter ko stranka dolg poplača le delno. V tem primeru banka pogosto utrpi izgubo, njena višina pa je odvisna od kvalitete zavarovanj posla.

Opisana najpogostejša primera pojasnjujeta tipično porazdelitev izgube ob neplačilu, ki ima navadno dve izrazitejši izboklini, pravimo ji **bimodalna porazdelitev**. Primer takšne porazdelitve izgube ob neplačilu je prikazan na sliki 4.3.



Slika 4.3: Tipična porazdelitev izgube ob neplačilu (LGD).

Najvišja izboklina porazdelitve iz slike 4.3 je leva izboklina, ki predstavlja izgubo ob neplačilu v višini 0 %, kar pomeni, da banka sploh ne utrpi izgube. Takšnih primerov je največ zaradi dveh razlogov:

1. Definicija neplačila je dokaj stroga, ker se kot neplačilo smatra vsaka zamuda, ki je daljša od 90 dni. Stranke se pogosto soočajo s plačilno nesposobnostjo, ki ni trajna, pa vendarle zamujajo več kot 90 dni. Večinoma stranke takšen dolg poravnajo dokaj hitro ter do unovčitve zavarovanj sploh ne pride.

2. Ko se stranka sooča s trajno plačilno nesposobnostjo in so posli dobro zavarovani, se dolg pogosto poplača v coloti z unovčitvijo zavarovanj.

Druga najvišja izboklina porazdelitve je desna izboklina, ki predstavlja izgubo ob neplačilu v višini 100 %, kar za banko pomeni popolno izgubo. Do popolne izgube pride, kadar banka ne prejme nikakršnih poplačil, posel pa je nezavarovan ter zato unovčitev zavarovanj ni možna. Takšnih primerov je manj, saj je dolg najpogosteje vsaj delno poplačan, kar na sliki 4.3 predstavlja del med levo in desno izboklino. Od tod izhaja zelo značilna bimodalna porazdelitev izgube ob neplačilu.

4.2 Stopnja poplačila

Izguba ob neplačilu (LGD) je v literaturi pogosto izražena v obliki njenega komplementa, to je **stopnja poplačila** RR (ang. recovery rate), za katero velja:

$$LGD = 1 - RR \tag{4.1}$$

Na podlagi akademskih in praktičnih raziskav so bile v članku [12] ugotovljene naslednje pogoste karakteristike stopnje poplačila v povezavi z izgubami:

- Večinoma je stopnja poplačila relativno visoka (okrog 70 - 80 %) ali relativno nizka (okrog 20 - 30 %). Posledično je razmišljanje o povprečni stopnji poplačila lahko zavajujoče.
- Eden od najpomembnejših dejavnikov, ki vplivajo na stopnjo poplačila, je zavarovanje posla.
- Stopnje poplačila so v obdobju recesije do ene tretjine nižje kot sicer.
- Zdi se, da na stopnjo poplačila vpliva tudi dejavnost, s katero se ukvarja stranka.
- Zdi se, da višina izpostavljenosti ne vpliva na višino izgube.

Zgoraj opisane karakteristike stopnje poplačila še dodatno pojasnjujejo bimodalno porazdelitev izgube o neplačilu iz slike 4.3.

4.3 Izračun izgube ob neplačilu na ravni posamezne terjatve

Z izgubo ob neplačilu (LGD) označujemo delež izpostavljenosti, ki jo bo banka utrpela kot izgubo, če stranka postane neplačnik. Enaka je razmerju med izpostavljenostjo ob neplačilu, povečani za administrativne stroške povezane z neplačilom in zmanjšani za sedanjo vrednost morebitnih poplačil, ter izpostavljenostjo ob neplačilu.

$$LGD = \max\left(0, \frac{EAD + \sum NPV_{stroški} - \sum NPV_{prihodki}}{EAD}\right) \tag{4.2}$$

pri čemer je:

- LGD ... izguba ob neplačilu za posamezni kreditni posel
- EAD ... izpostavljenost ob neplačilu
- $NPV_{stroški}$ ... neto sedanja vrednost stroškov po nastanku neplačila
- $NPV_{prihodki}$ ... neto sedanja vrednost prihodkov po nastanku neplačila

V postopku izterjave prihaja do stroškov, ki vplivajo na povečanje izpostavljenosti stranke do banke, zato so stroški izterjave v izračunu izgube ob neplačilu upoštevani kot negativni denarni tok. Na drugi strani se zaradi poplačil in unovčitve zavarovanj izpostavljenost zmanjšuje, zato so ti prihodki v izračunu izgube ob neplačilu upoštevani kot pozitivni denarni tok.

Lahko se zgodi, da prihodki presežejo izpostavljenost ob neplačilu povečano za stroške izterjave, vendar izguba ne more biti negativna, zato je izguba ob neplačilu vedno navzdol omejena z 0 %.

Primer 4.3.1. Naj bo dolžnik srednje veliko podjetje, ki ima z banko sklenjen posel v višini 10 mio EUR. Ob dogodku neplačila je njegova izpostavljenost ob neplačilu, vključno z neporavnanimi zamudnimi obrestmi in provizijami EAD znašala še 7 mio EUR. Po enem letu od dogodka neplačila je dolžnik poravnal 2 mio EUR, po dveh letih pa je banka z unovčitvijo zavarovanj prejela še 4 mio EUR. Postopek izterjave je trajal 2 leti,v tem času pa je nastalo 100.000 EUR stroškov.

Izračun izgube ob neplačilu z upoštevanjem 1,5 % diskontne stopnje denarnih tokov je naslednji:

$$
\begin{aligned}
LGD &= \max\left(0, \frac{1}{EAD}\left[EAD - \frac{P_1}{(1+r)^{t_1}} - \frac{P_2}{(1+r)^{t_2}}\right]\right) \\
&= \max\left(0, \frac{1}{7.000.000}\left[7.000.000 - \frac{2.000.000}{(1+0,015)} - \right.\right. \\
&\quad \left.\left. - \frac{4.000.000}{(1+0,015)^2} + \frac{100.000}{(1+0,015)^2}\right]\right) \\
&= \max(0, 18\%) \\
&= 18\% \qquad (4.3)
\end{aligned}
$$

Banka je utrpela izgubo ob neplačilu v višini 18 %.

4.4 Izračun izgube ob neplačilu na ravni portfelja

Banka za ocenjevanje tveganosti portfelja uporablja povprečne vrednosti izgube ob neplačilu za posamezne segmente portfelja. Povprečja so izračunana kot tehtana povprečja izgub ob neplačilu na ravni terjatev opazovanega segmenta portfelja.

Za način tehtanja v izračunu povprečne vrednosti LGD na ravni portfelja ali posameznega segmenta se banke odločijo glede na namen nadaljnje uporabe izračunanih vrednosti LGD. Glede na [17] ločimo med neplačilnim in časovnim tehtanjem ter tehtanje s številom neplačil in z višino izpostavljenosti neplačil.

4.4.1 Neplačilno tehtanje

Povprečje izračunano z neplačilnim tehtanjem za utež uporablja število neplačil ali delež izpostavljenosti neplačila v primerjavi s celotno izpostavljenostjo neplačil v opazovanem letu.

- Tehtanje s številom neplačil

$$LGD = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n_j} LGD_{i,j}}{\sum_{j=1}^{m} n_j} \qquad (4.4)$$

- Tehtanje z izpostavljenostjo neplačil

$$LGD = \frac{\sum_{j=1}^{m}\sum_{i=1}^{n_j} EAD_{i,j}LGD_{i,j}}{\sum_{j=1}^{m}\sum_{i=1}^{n_j} EAD_{i,j}} \qquad (4.5)$$

pri čemer je:

- i obravnavano neplačilo
- LGD ... izguba ob neplačilu
- j ... leto neplačila
- n_j ... število neplačil znotraj leta j
- m ... število opazovanih let

Tehtanje s številom neplačil je priporočljivo za segment podjetij, saj so visoke izpostavljenosti ob neplačilu pogostejše ter bi pri tehtanju z izpostavljenostjo neplačil visoka izpostavljenost pretehtala ostale, nižje izpostavljenosti ob neplačilu. Ravno nasprotno so v segmentu prebivalstva ogromne razlike med najnižjo in najvišjo izpostavljenostjo ob neplačilu manj pogoste, zato se uporablja tehtanje z izpostavljenostjo neplačil.

4.4.2 Časovno tehtanje

Povprečje izračunano s časovnim tehtanjem za utež uporablja število opazovanih let. Podobno kot pri neplačilnem tehtanju, tudi pri časovnem ločimo med tehtanjem s številom neplačil in tehtanjem z izpostavljenostjo neplačil.

- Tehtanje s številom neplačil

$$LGD = \frac{\sum_{j=1}^{m}(\frac{\sum_{i=1}^{n_j} LGD_{i,j}}{n_j})}{m} \tag{4.6}$$

- Tehtanje z izpostavljenostjo neplačil

$$LGD = \frac{\sum_{j=1}^{m}(\frac{\sum_{i=1}^{n_j} EAD_{i,j} LGD_{i,j}}{\sum_{i=1}^{n_y} EAD_{i,j}})}{m} \tag{4.7}$$

pri čemer je:

- i obravnavano neplačilo
- LGD ... izguba ob neplačilu
- j ... leto neplačila
- n_j ... število neplačil znotraj leta j
- m ... število opazovanih let

Časovno tehtanje je manj zaželeno, saj se leta z visokimi vrednostmi LGD »zgladijo« z leti z nižjimi vrednostmi LGD, kar lahko privede do podcenjevanja vrednosti deleža izgube ob neplačilu. Iz tega razloga se pogosteje kot časovno tehatnje uporablja neplačilno tehtanje.

4.5 Potrebni podatki za modeliranje izgube ob neplačilu (LGD)

Za izračun izgube ob neplačilu je potreben širok nabor podatkov z dolgo zgodovinsko časovno vrsto. Najpomembnejši dejavniki za zagotavljanje kvalitete zgrajenega modela so kvalitetni vhodni podatki, zato je procesu zbiranja podatkov potrebno nameniti veliko pozornosti.

Nabor potrebnih podatkov lahko razdelimo v tri sklope: podatki ob sklenitvi posla, o dogodku neplačila in o dinamiki denarnih tokov.

1. **Podatki o sklenitvi posla**

   Ob sklenitvi posla je potrebno zabeležiti čim več podatkov o stranki ter njenih poslih, kot so:

   - ID stranke,
   - tip stranke (npr. prebivalstvo, samostojni podjetnik ali podjetje),
   - velikost podjetja (npr. veliko, srednje, majhno ali mikro),
   - država,
   - ID posla,
   - datum sklenitve in zapadlosti posla,
   - odobreni znesek,
   - bonitetna ocena ob sklenitvi posla,
   - vrsta in ročnost posla,
   - ID zavarovalne pogodbe,
   - vrsta in višina zavarovanja;

2. **Podatki o dogodku neplačila**

   Ob nastanku neplačila je potrebno zbrati podatke o okoliščinah nastanka neplačila in o stranju stranke in njenih poslih, kot so:

   - datum neplačila,
   - tip neplačila (npr. vzrok za nastanek dogodka neplačila je lahko padec v bonitetno skupino D oz. E ali pomembna zamuda nad 90 dni ali restrukturiranje),
   - bonitetna ocena ob neplačilu,
   - aktivno bilančno stanje na dan neplačila,
   - zunajbilančno stanje na dan neplačila,
   - izpostavljenost na dan neplačila (EAD),
   - vrednost zavarovanja na dan neplačila,
   - datum izhoda iz statusa neplačila;

3. **Podatki o prihodkih in izdatkih**

   Po nastanku neplačila je potrebno zbirati podatke o poplačilih, unovčevanju zavarovanj ter stroških, ki nastanejo v postopku izterjave, kot so:

- datum in znesek plačila,
- datum unovčenja zavarovanja,
- ID unovčene zavarovalne pogodbe,
- znesek prihodka iz naslova unovčenja zavarovanja,
- datum in višina stroška;

Podatki morajo biti celoviti ter kvalitetni, kategorije podatkov pa morajo biti natančno definirane. Podatki so lahko opisane ali kvantitativne oblike, zato je smiselno uvesti tudi slamnate spremenljivke. Tiste spremenljivke, ki imajo le dve možni vrednosti (ali kategoriji), imenujemo **binarne spremenljivke**.

Tehnološka podprtost zbiranja podatkov je ključnega pomena, saj gre za obsežen nabor podatkov in bi ročno spremljanje podatkov zahtevalo veliko časa in organiziranosti. Prav tako je ročno spremljanje in ugotavljanja nastanka dogodka neplačila na celotnem portfelju posojil zaradi potrebnega štetja dni praktično nemogoče.

Poglavje 5

Binarna drevesa

V prejšnjih poglavjih smo pojasnili, kako se izguba ob neplačilu ocenjuje na ravni posameznega posla in na ravni portfelja. Za pridobitev optimalnih rezultatov je potrebna segmentacija portfelja. V ta namen se pogosto uporabljajo binarna drevesa, ki nastopajo v vlogi odločitvenih dreves.

V tem poglavju si bomo ogledali nekaj osnovnih definicij iz terorije grafov ter teoretično ozadje binarnih dreves.

5.1 Osnovne definicije in izreki iz teorije grafov

Navedli bomo nekaj osnovnih definicij in izrekov iz teorije grafov, ki so potrebni za razumevanje odločitvenih dreves. Povzeti so iz [18] in [19].

Definicija 5.1.1. **Graf** je urejeni par $G = (V, E)$, kjer je V množica točk in E množica neurejenih parov točk, ki jih imenujemo povezave.

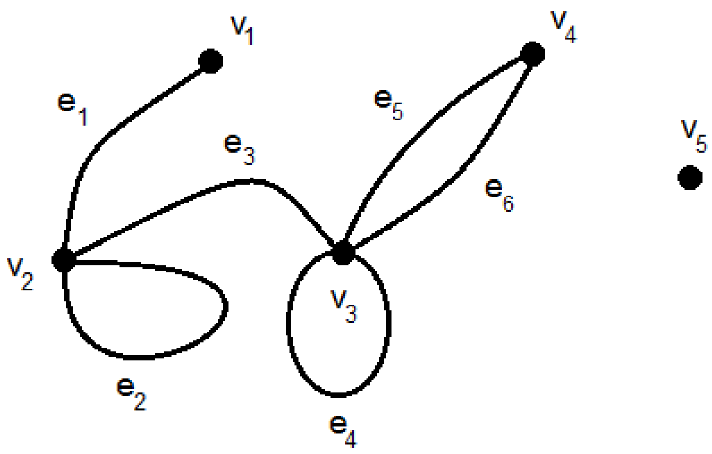
Definicija 5.1.2. Za graf $G = (V, E)$ velja:

- Točki u in v sta **končni točki** povezave (u, v).
- Povezavi, ki imajo skupne končne točke so **vzporedne povezave**.
- Povezavi (v, v) pravimo **zanka**.
- Graf je **enostaven**, če nima vzporednih povezav in zank.
- **Povezavi** sta **sosednji**, če imata skupno končno točko.
- **Točki** u in v sta **sosednji**, če sta povezani s povezavo (u, v).
- **Stopnja točke** $deg(v)$ točke v je enaka številu povezav s končno točko v. Najmanjšo stopnjo točke označimo z $\delta(G)$ največjo stopnjo točke pa z $\Delta(G)$.

Zgoraj definirane pojme si oglejmo še na primeru:

Primer 5.1.3. Graf G na sliki 5.1 je sestavljen iz množice točk $V = \{v_1, v_2, \ldots, v_5\}$ in povezav $E = \{(v_1, v_2), (v_2, v_2), (v_2, v_3), (v_3, v_3), (v_3, v_4), (v_3, v_4)\}$, kar pišemo krajše $E = \{e_1, e_2, e_3, e_4, e_5, e_6\}$. Zanj velja naslednje:

- v_3 in v_4 sta končni točki povezave e_5,



Slika 5.1: Graf G.

- e_5 in e_6 sta vzporedni povezavi,
- e_2 in e_4 sta zanki,
- G ni enostaven,
- e_1 in e_2 sta sosednji povezavi,
- v_1 in v_2 sta sosednji točki,
- stopnja točke v_4 je 2,
- $\delta(G) = 0$ in $\Delta(G) = 5$;

Zapletenejše grafe najlažje preučujemo tako, da opazovani graf razdelimo na več manjših delov, katerim pravimo podgrafi.

Definicija 5.1.4. Graf $G_1 = (V_1, E_1)$ imenujemo **podgraf** grafa $G = (V, E)$, če je $V_1 \subseteq V$ in $E_1 \subseteq E$, kar označimo $G_1 \subseteq G$.

Definicija 5.1.5. Naj bo $G = (V, E)$ graf in $S \subseteq V(G)$. Z $G - S$ označimo podgraf grafa G, ki smo mu **odstranili** točke iz množice S. Če grafu G odstranimo le točko v, potem pišemo $G - v$.

Eno izmed zelo pogostih vprašanj v teoriji grafov je, kako po povezavah grafa priti iz ene točke v drugo. Glede na omejitev kolikokrat prehodimo isto povezavo, da pridemo iz ene točke grafa v drugo, ločimo sprehod in pot.

Definicija 5.1.6. Sprehod S v grafu G je zaporedje točk in povezav

$$v_{i0}, e_{j1}, v_{i1}, e_{j2}, \ldots, e_{jk}, v_{ik}$$

Sprehod se začne in konča v končnih točkah v_{i0} in v_{ik}. Dolžina sprehoda S je enaka k. Če so vse točke sprehoda S različne, potem sprehod imenujemo pot. Dolžino poti med točkama u in v označimo z $d(u, v)$.

Če v grafu med dvema poljubnima točkama obstaja sprehod, potem med tema dvema točkama obstaja tudi pot. Najdemo jo tako, da potujemo po točkah in povezavah sprehoda na način, da izpustimo vse cikle in tako nobene točke ali povezave ne obiščemo več kot enkrat. Velja pa, da tako potovanje po točkah in povezavah lahko naredimo le v povezanem grafu.

Definicija 5.1.7. Graf G je **povezan**, če obstaja pot med poljubnim parom točk iz grafa G, sicer je **nepovezan**. Nepovezan graf razpade na povezane podgrafe, katerim pravimo **komponente** grafa.

5.2 Drevesa

Drevesa spadajo v enega izmed najenostavnejših razredov grafov. Neglede na njihovo enostavnost imajo bogato strukturo in so uporabni na ogromno področjih, kot na primer pri podatkvnih strukturah.

Definicija 5.2.1. Drevo je povezan graf brez ciklov.

Izrek 5.2.2. Za graf T z n točkami in m povezavami so ekvivalentne naslednje trditve:

1. T je drevo.
2. Med dvema poljubnima točkama grafa T obstaja natanko ena pot in graf T je brez zank.
3. Graf T je povezan in velja $m = n - 1$.
4. Graf T nima cikla in $m = n - 1$.
5. Graf T nima cikla, če pa bi mu dodali poljubno povezavo, potem bi imel graf G natanko en cikel.

Dokaz izreka se nahaja v [18].

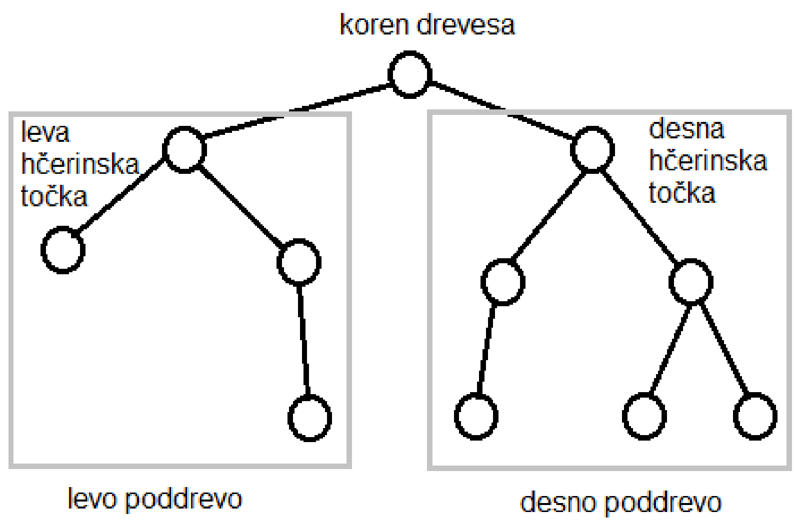
Definicija 5.2.3. Naj bo $T = (V, E)$ drevo. **Poddrevo** T' je podgraf grafa T.

5.2.1 Binarno drevo

Posebna oblika dreves so **binarna** ali dvojiška drevesa. Zanje velja, da so bodisi prazna (brez vozlišč) bodisi imajo končno množico vozlišč, ki jo sestavljajo tri disjunktne podmnožice: koren drevesa, levo poddrevo in desno poddrevo. Struktura binarnega drevesa je prikazana na sliki 5.2.

Definicija 5.2.4. Naj bo T drevo in $r \in V(T)$ koren drevesa T. Takšno drevo imenujemo **drevo s korenom**.

Definicija 5.2.5. Naj bo T drevo s korenom r in naj bo u sosednja točka točke v na edini poti do korena r. Potem je u **starševska točka** točke v in točka v je **hčerisnka točka** točke u.



Slika 5.2: Disjunktne podmnožice: koren drevesa, levo in desno poddrevo.

Definicija 5.2.6. Naj bo T drevo in naj bo $l \in V(T)$ točka stopnje 1. Takšni točki pravimo **list** (drevesa).

Iz zgornjih definicij sledi, da točki brez starševske točke pravimo koren drevesa, točki brez hčerinskih točk pa list drevesa.

Ena najpomembnejših lastnosti binarnih ali dvojiških dreves je, da v vsaki notranji točki drevesa na izbiro ponujajo natanko dve možnosti. Ravno zato se binarna drevesa uporabljajo kot odločitvena drevesa, saj potovanje po binarnem drevesu pripelje do končnih rezultatev odločitev, ki se nahajajo v listih binarnega drevesa.

Poglavje 6

Gradnja odločitvenega drevesa

V prejšnjem razdelku smo se spoznali z osnovnimi definicijami iz teorije grafov in z binarnimi drevesi, ki se pogosto uporabljajo kot odločitvena drevesa.

Ob gradnji odločitvenega drevesa se srečujemo z dilemo kriterija za nadaljnje delitve podatkovnih točk, pri čemer se sprašujemo, katera spremenljivka je ustrezna za kriterij delitve, v kakšnem zaporedju uporabiti kriterije za delitev ter kdaj se ustaviti z nadaljnjimi delitvami. Da bi odgovorili na ta vprašanja, je potrebno poznati način ocenjevanja odvisne spremenljivke v listih odločitvenega drevesa, saj je obnašanje modela za ocenjevanje odvisne spremenljivke eden izmed pomembnih meril za postopek gradnje odločitvenega drevesa.

V tem poglavju si poglejmo postopek gradnje odločitvenih dreves opisan v [20] in [21].

6.1 Ideja metode rekurzivnega deljenja

V enostavni linearni regresiji je odvisna spremenljivka Y modelirana kot vsota funkcije odvisne spremenjivke X in belega šuma:

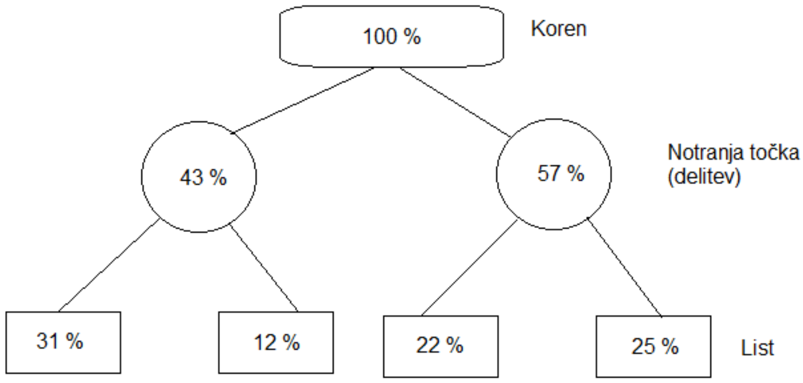
$$Y = \beta_0 + \beta_1 X + \epsilon \tag{6.1}$$

V multipli regresiji dopuščamo, da je odvisna spremenljivka Y modelirana z več neodvisnimi spremenljivkami $X_1, X_2, \ldots, X_p$, ki jih označimo z X. Potem velja:

$$Y = \beta_0 + \beta^T X + \epsilon \tag{6.2}$$

Podatkovna osnova ima pogosto veliko potencialnih pojasnjevalnih spremenljivk, povezave med njimi pa niso povsem jasne, zato je težko najti model, ki bi deloval na celotni podatkovni osnovi. Modeli, ki na takšnih podatkovnih osnovah delujejo, so pogosto zapleteni in težki za interpretacijo. Ena od možnih poenostavitev modeliranja je, da se podatkovno osnovo smiselno razdeli na več disjunktnih množic, za vsako množico pa se razvije svoj model, nato pa se jih glede na prvotno delitev podatkov sestavi skupaj. Takšnemu modeliranju pravimo **rekurzivno deljenje**.

Metodo rekurzivnega deljenja najlažje predstavimo z odločitvenim drevesom, v katerem vsaka notranja točka in listi drevesa predstavljajo korake delitve. Pri tem poljubna podatkovna točka x pripada listu drevesa le, če je razvrščena v ustrezno celico delitve. To najlažje preverimo tako, da se postavimo v koren drevesa, nato pa se zaporedno vprašamo o lastnostih opazovane podatkovne točke x. Zaporedje vprašanj določajo notranje točke odločitvenega drevesa, odgovore pa ponujajo veje drevesa, ki z ponujenimi odgovori na vprašanja povezujejo točke drevesa, kar prikazuje slika 6.1. Koren drevesa vsebuje vse podatkovne točke, listi drevesa pa vsebujejo le tiste podatkovne točke, ki se glede na zaporedje vprašanj in odgovorov uvrstijo v posamezni list drevesa.



Slika 6.1: Deleži podatkovnih točk v delitvah odločitvenega drevesa.

V odločitvenem drevesu je za napoved odvisne spremenljivke Y določen model, ki upošteva lastnosti podatkovnih točk, ki so se uvrstile v posamezno točko odločitvenega drevesa. Zelo pogosto se uporablja odsekoma konstantni model, ki je enak matematičnemu upanju odvisne spremenljivke Y v opazovani točki odločitvenega drevesa.

Definicija 6.1.1. Naj podatkovne točke $(x_1, y_1), (x_2, y_2), \ldots, (x_c, y_c)$ pripadajo listu l odločitvenega drevesa T. Potem velja, da je **odsekoma konstantni model** v listu l enak:

$$\widehat{y} = \frac{1}{c}\sum_{i=1}^{c} y_i \tag{6.3}$$

Odsekoma konstantni model ima v primerjavi z zapletenejšimi modeli kar nekaj prednosti, kot so:

- izdelava napovedi odvisne spremenljivke je hitra in ne zahteva zapletenih izračunov,
- enostavno je razumeti katere spremenljivke so pomembne za izdelavo napovedi odvisne spremenljivke,
- če je kateri od podatkov pomanjkljiv in zato potovanje po odločitvenem drevesu vse do željenega lista ni mogoče, je napoved odvisne spremenljivke še vedno možna z izračunom povprečne vrednosti vseh listov notranjega vozlišča, do katerega smo prišli, in
- gre za hiter in zanesljiv algoritem;

Nekaj podobnih modelov za namen izračuna izgube ob neplačilu na ravni portfelja smo opisali že v poglavjih 4.3.1 in 4.3.2, kjer smo ločili med neplačilnim in časovnim tehtanjem, znotraj njiju pa še tehtanje s številom neplačiln in tehtanje z izpostavljenostjo neplačil.

6.1.1 Kriteriji za ustavitev

Vsak rekurzivni algoritem mora imeti jasno definiran kriterij, ki pove, kdaj naj se algoritem ustavi. Takšnemu kriteriju pravmo **kriterij za ustavitev** (ang. stopping criterion). V primeru gradnje odločitvenega drevesa kriterij za ustavitev pove, kdaj naj algoritem preneha z iskanjem naslednje delitve, zato je izbira ustreznega kriterija ključen korak za izgradnjo kvalitetnega odločitvenega drevesa.

Najbolj očiten kriterij za ustavitev je, ko točka odločitvenega drevsa vsebuje le še eno podatkovno točko, saj nadaljnja delitev ni več možna. Število listov takšnega odločitvenega drevesa je enako številu podatkovnih točk, kar pomeni, da je drevo zelo obsežno. Posledično je malo verjetno, da je odločitveno drevo optimalno. Delno izboljšan zgornji kriterij je, da se gradnja drevesa konča, ko ima točka odločitvenega drevesa manj kot n podatkovnih točk. Vendar tudi takšno drevo zelo verjetno ni optimalno. Ker so modeli v posameznih točkah odločitvenega drevesa dokaj enostavni, je smiselno, da se večina truda vloži v iskanje optimalnega drevesa, kar pomeni, da je potrebno najti ustrezno delitev podatkovnih točk.

Iščemo najboljšo napoved odvisne spremenljivke Y. Cilj je maksimizirati funkcijo $I[C;Y]$, ki za opazovano delitev podaja informacijo o odvisni spremenljivki Y, spremenljivka C pa pove, v katerem listu odločitvenega drevesa se nahajamo. Maksimizacije ni mogoče izvesti v enem koraku, zato začnemo tako, da najdemo binarno vprašanje, ki maksimizira informacijo o odvisni spremenljivki Y. Tako najdemo koren drevesa in dve hčerinski točki. V vsaki od hčerinskih točk ponovimo postopek in iščemo binarna vprašanja, ki dajejo maksimalne informacije o odvisni spremenljivki Y.

Ustreznejši kriterij za ustavitev je, da se gradnja odločitvenega drevesa konča, ko nadaljnja delitev oceno spremenljivke Y izboljša za manj kot vnaprej določen prag.

Definicija 6.1.2. Vsota kvadratov napak drevesa T je enaka:

$$S = \sum_{c \in listi(T)} \sum_{i \in C} (y_i - m_c)^2 \tag{6.4}$$

pri čemer je m_c napoved za list c, ki je enaka:

$$m_c = \frac{1}{n_c} \sum_{i \in c} y_i \tag{6.5}$$

pri čemer je n_c število podatkovnih točk v listu c.

S pomočjo izraza za m_c zgornji izraz poenostavimo:

$$\begin{aligned} S &= \sum_{c \in listi(T)} \sum_{i \in c} (y_i - m_c)^2 \\ &= \sum_{c \in listi(T)} \sum_{i \in c} (y_i - \frac{1}{n_c} \sum_{i \in c} y_i)^2 \\ &= \sum_{c \in listi(T)} n_c V_c \end{aligned} \tag{6.6}$$

pri čemer z V_c označimo varianco lista c.

Cilj je minimizirati vsoto kvadratov napak S drevesa T, kar je tudi naš kriterij za ustavitev gradnje odločitvenega drevesa. Enostavni algoritem za gradnjo odločitvenega drevesa je sledeč:

1. Gradnjo drevesa začni z eno samo točko, ki vsebuje vse podatkovne točke. Izračunaj m_c in S.

2. Če imajo v opazovani točki odločitvenega drevesa vse podatkovne točke isto vrednost za vse vstopne spremenljivke, končaj. Sicer išči po vseh binarnih delitvah vseh spremenljivk, dokler ne najdeš delitve, ki vrednost S čim bolj zmanjša. V kolikor je največje zmanjšanje vrednosti S manjše od vnaprej določenega pragu δ, ali ena od na novo kreiranih točk vsebuje manj kot q podatkovnih točk, končaj. Sicer delitev uporabi in kreiraj novi dve točki odločitvenega drevesa.

3. V vsaki na novo kreirani točki se vrni na prvi korak.

   Opomba: Če je več pojasnjevalnih spremenljivk enako dobrih, je potem izbira katero uporabiti, odvisna od arbitrarne odločitve pri programiranju odločitev.

Slabost zgornjega algoritma je, da se algoritem lahko prehitro ustavi. Če katera spremenljivka ni dovolj informativna, se gradnja odločitvenega drevesa pri njej ustavi, čeprav bi naslednja delitev model mogoče izboljšala. Do te slabosti pride zaradi kriterija ustavitve, da se algoritem ustavi, če je zmanjšanje S manjše od δ, ter zaradi predpostavke, da je minimalno število podatkovnih točk q na vozlišče določeno arbitrarno.

6.1.2 Ideja navzkrižne validacije

Uspešnejši pristop za iskanje regresijskega drevesa je uporaba navzkrižne validacije, pri kateri se podatkovno osnovo naključno razdeli na dve množici, **učno** ter **testno**.

Učna množica je namenjena gradnji odločitvenega drevesa, v našem primeru z zgornjim algoritmom, pri čemer naj bosta $q = 1$ in $\delta = 0$. Pričakovano je, da bo takšno drevo preveliko in se bo preveč prilegalo podatkom učne množice, novim podatkom pa se bo slabo prilegalo (ang. overfit) Iz tega razloga v naslednjem koraku sledi navzkrižna validacija s ciljem obrezati preveliko odločitveno drevo. Za vsak par listov ter za njuno starševsko točko izračunamo napako na testni množici podatkov ter presodimo, ali se napaka zmanjša, če bi se odločitveno drevo končalo že v starševski točki. Če to drži, drevo obrežemo in starševska točka pa postane list drevesa, sicer drevesa ne obrežemo. Tak postopek ponavljamo na vseh parih listov odločitvenega drevesa ter njihovih starševskih točkah, dokler napak na testni množici ni več mogoče zmanjšati.

Obstajajo še drugi načini navzkrižne validacije, na primer način, s katerim se podatkovno osnovo prav tako razdeli na učno in testno množico, nato pa se izmenjujeta postopka gradnje in obrezovanja odločitvenega drevesa ter se menjata vlogi učne in testne množice. Postopek gradnje se konča, ko se velikost drevesa ne spreminja več.

Postopek navzkrižne validacije je ustreznejši od arbitrarnega določanja kriterijev za ustavitev, saj pri navzkrižni validaciji za vsak par listov preverjamo, ali se na podlagi ocene napake dodatna delitev v drevesu sploh izplača, pri arbitrarno določenih kriterijih pa te kontrole ni.

6.2 Rekurzivno deljenje

V tem razdelku bomo idejo rekurzivnega deljenja, ki je bila predstavljena v prejšnjem poglavju opisali natančneje.

6.2.1 Notacija

Naj populacija modela obsega n opazovanj iz C razredov. Model, ki ga bomo opisali v nadaljevanju, bo n opazovanj razdelil v k podskupin ter jih razvrstil v C razredov. Uporabljali bomo naslednje oznake:

- π_i za $i = 1, 2, \ldots, C$ so apriorne (naprejšnje) verjetnosti za vsak razred.
- $L(i, j)$ za $i = 1, 2, \ldots, C$ so matrike izgub iz naslova nepravilnih razrvstitev i kot j. Velja $L(i, i) = 0$.
- Z A označimo točko drevesa; pri tem z A označujemo tako množico podatkovnih točk iz podatkovnega vzorca kot razvrstitveno pravilo prihodnjih podatkov.
- $\tau(x)$ za pravilni razvrstitveni razred vektorja napovednih spremenljivk x,
- $\tau(A)$ za razvrstitveni razred, kamor spada točka A, če bi bila točka A list odločitvenega drevesa,
- n_i za število opazovanj v vzorcu, ki so v razvrstitvenem razredu i,
- n_A za število opazovanj v točki A,
- $P(A)$ za verjetnost A (za prihodnje podatke)

$$\begin{aligned} &= \sum_{i=1}^{C} \pi_i P\{x \in A \,|\tau(x) = i\} \\ &\approx \sum_{i=1}^{C} \pi_i n_{iA}/n_i \end{aligned} \tag{6.7}$$

- $p(i|A)$ za $P\{\tau(x) = i|x \in A\}$ (za prihodnje podatke)

$$\begin{aligned} p(i|A) &= \pi_i P\{x \in A|\tau(x) = i\}/P\{x \in A\} \\ &\approx \pi_i(n_{iA}/n_i)/\sum \pi_i(n_{iA}/n_i) \end{aligned} \tag{6.8}$$

- $R(A)$ za tveganje vozlišča A

$$R(A) = \sum_{i=1}^{C} p(i|A)L(i, \tau(A)) \tag{6.9}$$

kjer je $\tau(A)$ izbran tako, da minimizira $R(A)$.

- $R(T)$ za tveganje modela (ali drevesa) T

$$R(T) = \sum_{j=1}^{k} P(A_j)R(A_j) \tag{6.10}$$

  kjer so A_j listi drevesa.

- če velja $L(i,j) = 1$ za vse $i \neq j$ in so vse apriorne verjetnosti π enake opazovanim frekvencam vzorca, potem je $p(i|A) = n_{i,A}/n_A$ in $R(T)$ je delež napačne razvrstitve.

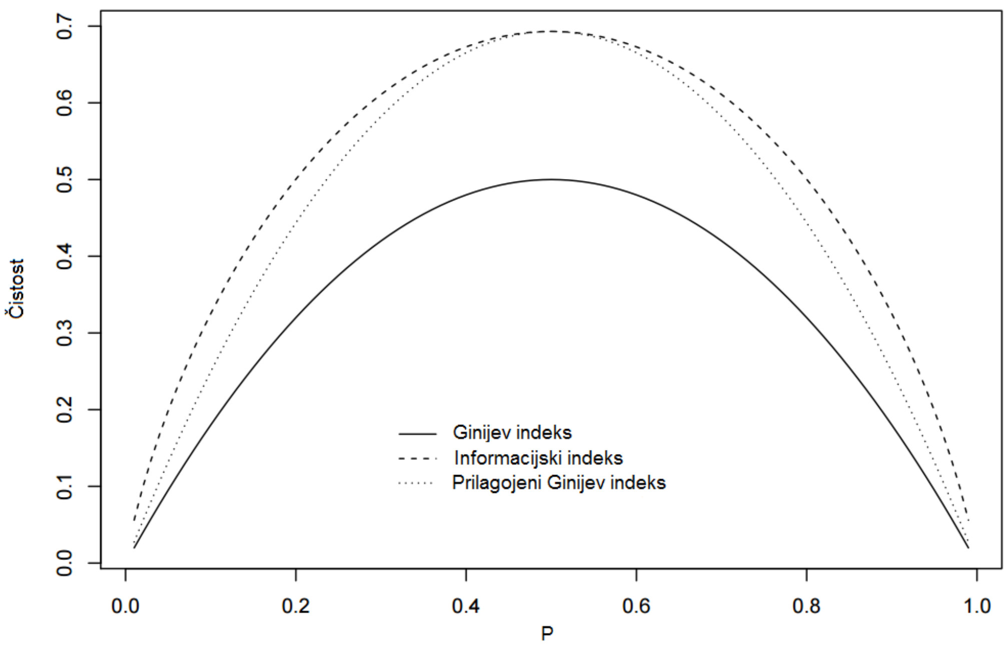
6.2.2 Kriteriji za delitev

Kriterij za delitev definiramo s pomočjo funkcije nečistosti (ang.impurity function):

Definicija 6.2.1. Naj bo f poljubna funkcija nečistosti. Potem

$$I(A) = \sum_{i=1}^{C} f(p_{iA}) \tag{6.11}$$

definira **nečistost točke** A, kjer z p_{iA} označujemo razmerje tistih podatkovnih točk iz A, ki pripadajo razredu i v prihodnjih vzorcih.



Slika 6.2: Mera čistosti - primerjava Ginijevega indeksa in informacije o nečistosti.

Želimo najti točko A, za katero velja, da je nečistost $I(A) = 0$. Da bi to dosegli, mora biti funkcija nečistosti f konkavna z $f(0) = f(1) = 0$.

Glede na sliko 6.2 sta kandidata za funkcijo nečistosti f:

1. Informacijski indeks $f(p) = -p\log(p)$

2. Ginijev indeks $f(p) = p(1-p)$

To delitev uporabimo pri maksimalnem zmanjšanju nečistoč, pri čemer podatkovne točke iz točke A razdelimo v dve hčerisnki točki: levo A_L in desno A_D. Potem velja:

$$\Delta I = p(A)I(A) - p(A_L)I(A_L) - p(A_R)I(A_R) \tag{6.12}$$

Na sliki 6.2 sta narisani obe predlagani funkciji nečistosti ter prilagojeni Ginijev indeks.

Na primeru dveh razredov ($C = 2$), se meri skoraj ne razlikujeta in skoraj vedno izbereta isto točko delitve.

6.2.3 Vključitev funkcije izgube

Tveganje lahko zmanjšamo tudi z vključitvijo funkcije izgube, ki v funkciji nečistosti ni upoštevana. Sledi opis dveh možnih razširitev funkcije nečistošti, to sta poplošen Ginijev indeks in spremenjene apriorne verjetnosti.

1. **Posplošen Ginijev indeks**

   Naj bo objekt naključno izbran iz enega od C razredov z verjetnostmi $(p_1, p_2, \ldots, p_C)$ in je z enako verjetnostjo naključno razporejen v enega od naštetih razredov. Verjetnost napačne razporeditve je enaka:

   $$\sum_i \sum_{j \neq i} p_i p_j = \sum_i \sum_j p_i p_j - \sum_i p_i^2 = \sum_i 1 - p_i^2 \tag{6.13}$$

   kar je **Ginijev indeks** za p_i.

   Naj bo $L(i,j)$ izguba ob razporeditvi v razred j, čeprav objekt spada v razred i. Pričakovan strošek napačne razporeditve je $\sum_i \sum_j L(i,j) p_i p_j$.

   Od tod definiramo **posplošen Ginijev indeks nečistosti** kot:

   $$G(p) = \sum_i \sum_j L(i,j) p_i p_j \tag{6.14}$$

   Ginijev indeks ima tudi nekaj pomanjkljivosti. Funkcija $G(p)$ ni nujno konkavna, kar je bil naš pogoj pri iskanju funkcije nečistosti. Druga pomanjkljivost je, da $G(p)$ v izračunu uporablja simetrično matriko $L(i,j)$, saj je izračun sledeč:

   $$G(p) = \frac{1}{2} \sum_i \sum_j [L(i,j) + L(j,i)] p_i p_j \tag{6.15}$$

   Zlasti so težave na primeru dveh razredov, saj izračun $G(p)$ ignorira matriko izgub iz naslova nepravilne razvrstitve.

2. **Spremenjene apriorne verjetnosti**

   Spomnimo se definicije tveganja vozlišča A:

$$\begin{aligned} R(A) &= \sum_{i=1}^{C} p_{i,A} L(i, \tau(A)) \\ &= \sum_{i=1}^{C} \pi_i L(i, \tau(A))(n_{iA}/n_i)(n/n_A) \end{aligned} \tag{6.16}$$

   Predpostavimo, da obstajata $\widetilde{\pi}$ in $\widetilde{L}$, tako da velja naslednje:

$$\widetilde{\pi}_i \widetilde{L}(i,j) = \pi_i L(i,j) \qquad \text{za vsak } i,j \in C \tag{6.17}$$

   Potem se ob novih izgubah in apriornih verjetnostih tveganje $R(A)$ ne spreminja. Če je $\widetilde{L}$ proporcionalna matriki izgub z enicami in nulami, potem morajo biti apriorne verjetnosti $\widetilde{\pi}$ uporabljene v kriteriju za delitev. To je mogoče le, če je L oblike:

$$L(i,j) = \begin{cases} L_i & i \neq j \\ 0 & i = j \end{cases}$$

   V tem primeru je:

$$\widetilde{\pi} = \frac{\pi_i L_i}{\sum_j \pi_j L_j} \tag{6.18}$$

   To je vedno možno za $C = 2$, zato so spremenjene apriorne verjetnosti obeh razredov enake.

   Druga utemeljitev za uporabo spremenjenih apriornih verjetnosti je naslednja: Nečistost $I(A) = \sum f(p_i)$ svoj maksimum doseže v $p_1 = p_2 = \ldots = p_C = \frac{1}{C}$. Če bi problem imel, na primer, strošek napačne razporeditve za razred 1 dvakratnik stroška za razred 2 ali 3, potem bi si želeli, da ima $I(A)$ maksimum v $p_1 = \frac{1}{5}, p_2 = p_3 = \frac{2}{5}$, saj je to niz najslabših možnih razmerij verjetnosti, ki določajo razred. Metoda spremenjenih verjetnosti s prestavljanjem p_i dela ravno to.

Opomba 6.2.2. Kadar so uporabljene spremenjene apriori verjetnosti, te vplivajo samo na izbiro delitve v nasprotju z običajnimi apriori verjetnosti, ki se uporabljajo za izračun tveganja točke odločitvenega drevesa. Spremenjene apriori verjetnosti nastopajo le kot pomoč pravilu nečistosti pri izbiri ustrezne delitve.

Opomba 6.2.3. Argument za spremenjene apriori verjetnosti velja tako za Ginijev indeks kot za informacijska delitvena pravila.

6.3 Obrezovanje drevesa

6.3.1 Definicije

Zgrajeno drevo je sprva zelo verjetno preveliko ali komepleksno, zato se je potrebno odločiti, koliko drevesa je smiselno ohraniti. V stopenjski regresiji se tak problem reši z zaporedno obravnavo, prilagajanje pa se ustavi, ko $F-$ test pade pod vnaprej določeno mejo α.

Definicija 6.3.1. Naj bodo $T_1, T_2, \ldots, T_k$ točke drevesa T. Potem:

- s $|T|$ označimo **število listov drevesa**,
- s $R(T) = \sum_{i=1}^{k} P(T_i)R(T_i)$ določimo **tveganje drevesa** T,
- $\alpha \in (0, \infty)$ je **parameter kompleksnosti**, ki meri strošek dodatne spremenljivke v modelu, ter
- $R(T_0)$ določa **tveganje drevesa brez delitev**,

Definicija 6.3.2. Strošek drevesa definiramo kot:

$$R_\alpha(T) = R(T) + \alpha|T| \tag{6.19}$$

S T_α označimo poddrevo celotnega modela, ki ima minimalne stroške. Očitno je potem T_0 enak celotnemu modelu, T_∞ pa je model brez delitev.

V [21] so predstavljeni naslednji rezultati:

1. Če sta T_1 in T_2 poddrevesa drevesa T, tako da velja $R_\alpha(T_1) = R_\alpha(T_2)$, potem je ali T_1 poddrevo drevesa T_2 ali je T_2 poddrevo drevesa T_1. Zato velja ali $|T_1| < |T_2|$ ali $|T_2| < |T_1|$.
2. Če je $\alpha > \beta$, potem je ali $T_\alpha = T_\beta$ ali je T_α strogo poddrevo drevesa T_β
3. Naj bo $\alpha_1, \alpha_2, \ldots, \alpha_m$ množica števil. Velja, da so tako $T_{\alpha_1}, T_{\alpha_2}, \ldots, T_{\alpha_m}$ kot $R(T_{\alpha_1}), R(T_{\alpha_2}), \ldots, R(T_{\alpha_m})$ lahko učinkovito izračunani.

Z uporabo prvega rezultata lahko enolično definiramo T_α kot najmanjše drevo T, za katerega je $R_\alpha(T)$ minimiziran.

Ker ima poljubno zaporedje vgnezdenih dreves drevesa T največ |T| listov drevesa, drugi rezultat pomeni, da so vse možne vrednosti α lahko grupirane v m intervalov, $m \leq |T|$

$$\begin{aligned} I_1 &= [0, \alpha_1] \\ I_2 &= (\alpha_1, \alpha_2] \\ &\vdots \\ I_m &= (\alpha_{m-1}, \infty] \end{aligned} \tag{6.20}$$

kjer si vsi $\alpha \in I_i$ delijo isto minimizirano poddrevo.

6.3.2 Navzkrižna validacija

Navzkrižno validacijo uporabljamo za določitev najbolj ustrezne vrednosti za α po naslednjih korakih:

1. Prilagodi model na podatkovno osnovo

   Izračunaj $I_1, I_2, \ldots, I_m$ Nastavi:

$$\begin{aligned} \beta_1 &= 0 \\ \beta_2 &= \sqrt{\alpha_1 \alpha_2} \\ \beta_3 &= \sqrt{\alpha_2 \alpha_3} \\ &\vdots \\ \beta_{m-1} &= \sqrt{\alpha_{m-2} \alpha_{m-1}} \\ \beta_m &= \infty \end{aligned} \tag{6.21}$$

   pri čemer je vsak β_i značilna vrednost za pripadajoči I_i.

2. Množico podatkov razdeli v s skupin $G_1, G_2, \ldots, G_s$, vsako velikosti $\frac{s}{n}$ in za vsako skupino ločeno:

   - celoten model prilagodi na množico podatkov na vse razen G_i' in določi $T_{\beta_1}, T_{\beta_2}, \ldots, T_{\beta_m}$, za pripadajočo zmanjšano množico podatkov,
   - izračunaj napovedan razred za vsako opazovanje iz G_i in za vsak model T_{β_j} za $1 \leq j \leq m$,
   - izračunaj tveganje vsakega subjekta posebej.

3. Seštej po G_i da dobiš oceno tveganja za vsak β_j. Tista vrednost β, ki ima najmanjše tveganje za T_β na celotni množici podatkov, je izbrana kot najboljše obrezano drevo.

Poglavje 7

Primer gradnje odločitvenega drevesa

V članku [22] je opisana študija, ki izgubo ob neplačilu ocenjuje z neparametričnim regresijskim drevesom. Napovedna moč modela je ocenjena na več časovnih horizontih poplačila v obdobju po nastanku dogodka neplačila, med drugimi tudi za obdobje enega leta. Model temelji na zgodovinskih podatkih o realiziranih izgubah, rezultati pa kažejo, da je modeliranje izgube ob neplačilu z odločitvenimi drevesom ustrezen pristop.

7.1 Podatki

V študji primera iz [22] so uporabljeni podatki portugalske banke Banco Comercial Português. Podatkovna osnova je sestavljena iz 374 posojil malih in srednje velikih podjetij (SME), ki so postali neplačniki v obdobju med junijem leta 1995 in decembrom leta 2000. Vsa podjetja imajo promet na letni ravni večji od 2,5 milijonov EUR. Dolg pri približno polovici posojil je manjši od 50.000 EUR. Na sliki 7.1 je zbirni pregled podatkovne osnove - število opazovanj, višina posojil ter povprečna višina posojil ločeno glede na leto zbiranja podatkov. Število neplačnikov je bilo v letih 1995 in 1996 nekoliko višje od preostalih let, čeprav v tem obdobju recesija ni bila zaznana.

Leto	Št. neplačil	Znesek (EUR)	Povprečni znesek (EUR)
1995*	65	5.950.097	91.540
1996	89	15.945.725	179.165
1997	59	12.479.614	211.519
1998	57	5.463.170	95.845
1999	47	7.013.532	149.224
2000	57	5.834.894	102367
Skupaj	374	52.687.032	140.874

*Druga polovica leta

Slika 7.1: Predstavitev podatkov. Tabela povzeta iz [22].

Podjetja so glede na dejavnost razdeljena v štiri skupine: gradbena, proizvodna, trgovina in storitve.

Vsi posli imajo bonitetno oceno, ki se dodeli na podlagi bančnega notranjega sistema za ocenjevanje strank ter njihovih poslov. Odvisen je od verjetnosti neplačila stranke ter od kvalitete ter višine zavarovanja posla. V obravnavani portugalski banki je v uporabi sedem bonitetnih razredov, A, B, C1, C2, C3, D in E, pri čemer je A najboljši, E pa najslabši. Skoraj polovica poslov iz podatkovne osnove nima dodeljene bonitetne ocene, kar bi lahko pomenilo, da je te posle potrebno izločiti iz podatkovne osnove. V izogib izločitvi, je tem poslom dodeljena povprečna bonitetna ocena preostalega dela podatkovne osnove. Poleg tega je podatkovni osnovi dodana slamnata spremenljivka z vrednostjo 1, če posel ni imel bonitetne ocene in vrednost 0, če je bonitetna ocena bila dodeljena.

Na delež izgube ob neplačilu (LGD) pomembno vplivata višina ter kvaliteta zavarovanja. V podatkovni osnovi je 58 % poslov zavarovanih z osebnim poroštvom, 15% z različnimi oblikami zavarovanj, kot so nepremičnine, zaloge in depoziti, 36% poslov pa je nezavarovanih.

Podatkovna osnova za vsak posel vključuje še podatke o pogodbeni obrestni meri, starosti podjetja ter o dobi sodelovanja podjetja z banko. Povprečna starost podjetij znaša 17 let, povprečna doba sodelovanja podjetja z banko pa 6 let.

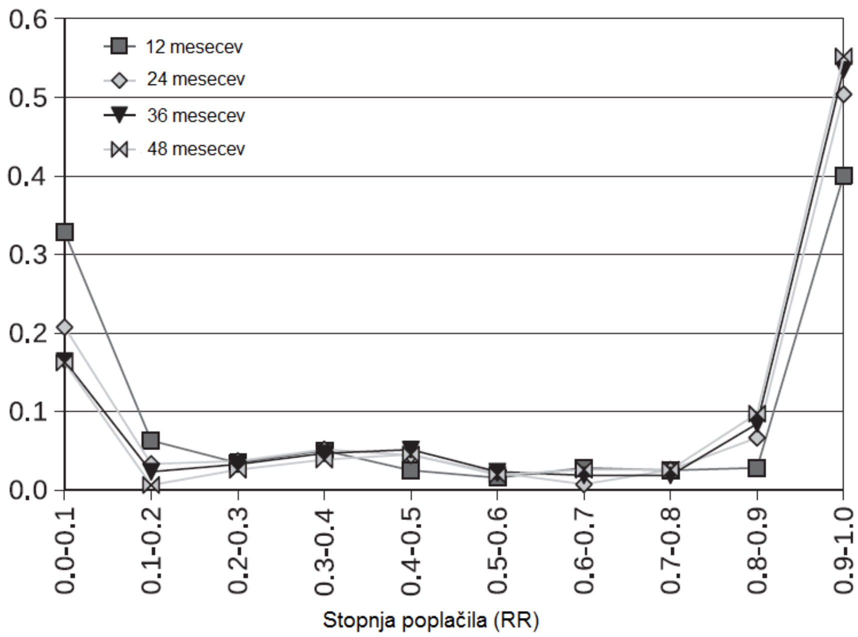
Ker bomo v modelu obravnavali obdobja 12, 24, 36 in 48 mesecev od neplačila, se časovno obdobje pri podatkih, kot sta starost podjetja in doba sodelovanja podjetja z banko, s časom povečujejo, podatki o bonitetnih ocenah ter zavarovanjih pa so bili zbrani ob datumu neplačila in se s časom ne spreminjajo.

V podatkovni osnovi so za vsak posel vključeni tudi denarni tokovi na meseči ravni, do katerih je prišlo po nastanku dogodka neplačila, sestavljeni pa so iz prejemkov plačil ter iz prejemkov iz naslova unovčitve zavarovanj. Za posle z dogodkom neplačila v juniju 1995 je na voljo dolga časovna vrsta denarnih tokov, za posle z dogodkom neplačila v letu 2000 pa je na voljo le kratka časovna vrsta denarnih tokov. Podatek o dejanskih odpisih ni na voljo.

Za izračun neto sedanje vrednosti denarnih tokov je uporabljena diskontna stopnja, ki je enaka pogodbeni obrestni meri. Ta morda ne zajame celotnega tveganja podjetja po nastanku neplačila, vendar pa je večji del poplačil izveden v prvih mesecih po dogodku neplačila, zato tak način diskontiranja bistveno ne odstopa. Stroški izterjevalnega procesa v izračunu denarnih tokov niso upoštevani.

Na sliki 7.2 je prikazana porazdelitev stopnje poplačila za obdobje poplačil 12, 24, 36 in 48 mesecev po nastanku dogodka neplačila, iz katere je razvidna značilna bimodalna porazdelitev stopnje poplačila. Obliko porazdelitve pojasnjujejo na eni strani veliko število posojil z nizko stopnjo poplačila ter na drugi strani veliko število posojil s celotnim poplačilom obveznosti. Frekvenca poplačil se s časom povečuje, saj se s podaljševanjem obdobja poplačevanja pričnejo vnovčevati tudi zavarovanja, ki pomembno vplivajo na višino denarnih tokov.

V tabeli 7.1 je predstavljenih nekaj osnovnih statističnih podatkov kumulativne porazdelitve stopnje poplačila (RR). V prvi vrsti tabele je navedeno število posojil, za katere je v opazovanem obdobju na voljo celotna časovna vrsta poplačil. V številu posojil za 12 mesecev niso upoštevani posli, pri katerih je prišlo do neplačila v zadnjem letu, v letu 2000. Podobno velja tudi za druga časovna obdobja tabele. Pričakovano se povprečna stopnja poplačila povečuje z daljšanjem časovnega obdobja, saj se s časom povečuje znesek poplačil. Mediana stopnje poplačila (RR) ima velik preskok iz 12 na 24 mesecev, kar je posledica velikega prirastka v številu skoraj



Slika 7.2: Porazdelitev stopnje poplačila za obdobja 12, 24, 36 in 48 mesecev. Slika povzeta iz [22].

popolnega poplačila in velikega zmanjšanja števila majhnih poplačil.

	12 mesecev	24 mesecev	36 mesecev	48 mesecev
Število posojil	317	270	213	154
Mat. upanje	0,503	0,646	0,694	0,714
Mediana	0,493	0,907	0,946	0,95
Standardni odklon	0,437	0,411	0,385	0,375

Tabela 7.1: Statistični podatki porazdelitve stopnje poplačila. Tabela povzeta iz [22].

Slika 7.2 in tabela 7.1 potrjujeta, da ni bistvenih razlik med porazdelitvami stopnje poplačil za obdobja 24, 36 in 48 mesecev, kar je posledica frekvence poplačil, ki se po nastanku neplačila hitro zmanjšuje. Opazimo tudi, da se s časom povečuje standardni odklon porazdelitve stopnje poplačil.

7.2 Odločitveno drevo

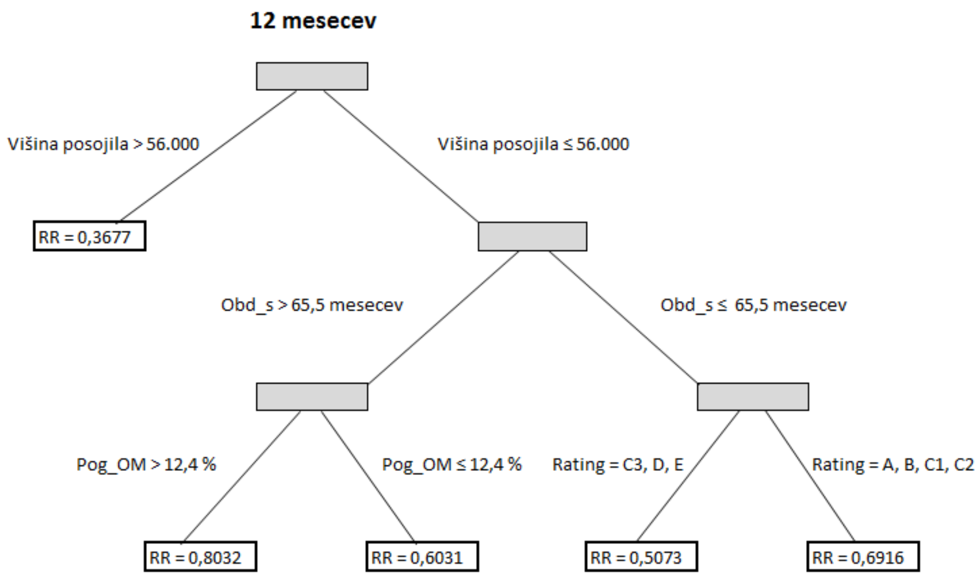
Regresijska drevesa so napovedni modeli, pri katerih se podatkovna osnova z algoritmom iskanja rekurzivno razdeli v manjše, med seboj izključujoče, množice. Podatkovno osnovo sestavljajo posojila, ki so opisana z atributi, opisanimi v razdelku 7.1 in ciljno spremenljivko, stopnjo poplačila (RR), ki je komplement izgube ob neplačilu (LGD).

Algoritem iskanja gradnjo drevesa začne v korenu drevesa, ki vsebuje vse podatkovne točke. Potem izmed vseh možnih binarnih delitev atributov poišče tisto, ki v novo nastalih točkah drevesa minimizira variacijo ciljne spremenljivke. Na ta način je ciljna spremenljivka v vsaki hčerinski točki bolj homogena kot v starševski točki. Tak postopek se rekurzivno ponavlja, dokler nadaljnja delitev ni več mogoča. Zmanjšanje variance ciljne spremenljivke merimo s standardnim odklonom zmanjšanja (7.1), kjer s T označujemo množico podatkovnih točk v starševski točki, T_1 in T_2 pa sta množici podatkovnih točk v hčerinskih točkah, ki sta posledica binarne delitve starševske točke. Operatorja $E(.)$ in $\sigma(.)$ predstavljata matematično upanje in standardni odklon ciljne spremenljivke na množici opazovanj.

$$SDR = \sigma(T) - \frac{E(T_1)}{E(T)}\sigma(T_1) - \frac{E(T_2)}{E(T)}\sigma(T_2) \tag{7.1}$$

Vse podatkovne točke so z začetkom v korenu porazdeljene do listov drevesa. Delitev določa pot do listov drevesa, v katerih se nahajajo napovedi ciljne spremenljivke, ki so definirane kot povprečje vrednosti ciljne spremenljivke podatkovnih točk iz opazovanega lista drevesa. Vrednosti ciljne spremenljivke, stopnje poplačila (RR), se gibljejo med 0 in 1.

Odločitvena drevesa so zelo pogosto precej velika in se preveč prilegajo. Takšna drevesa dajejo dobre informacije v zvezi s podatki, ki so uporabljeni za gradnjo drevesa, dajejo pa nezanesljive napovedi. Model izboljšamo z obrezovanjem odločitvenega drevesa. Postopek obrezovanja se začne v listih drevesa, ocenjujejo pa se pričakovane variance ciljne spremenljivke v hčerinskih in starševskih točkah odločitvenega drevesa. Če je varianca v hčerinski točki večja od variance v starševski točki, potem se lista drevesa obrežeta in starševska točka postane list odločitvenega drevesa. Ta proces se ponavlja, dokler izboljšanje ni več mogoče.



Slika 7.3: Odločitveno drevo za obdobje 12 mesecev. Slika povzeta iz [22].

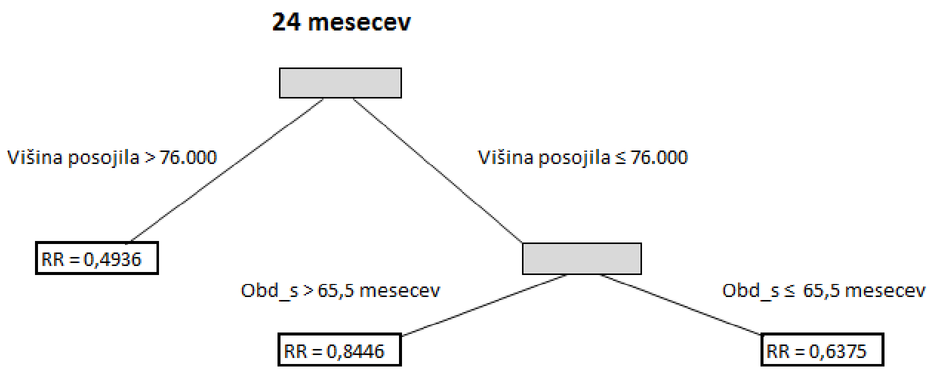
Na sliki 7.3 je regresijsko drevo za obdobje 12 mesecev po nastopu neplačila. S sledenjem poti od korena proti listom odločitvenega drevesa, se najprej vprašamo ali je višina posojila večja od 56.000 EUR? Če je odgovor pritrdilen, potem je pričakovana stopnja poplačila RR enaka 36,77 %. Veja drevesa se v tej točki konča. Če je odgovor nikalen, se nadalje vprašamo, ali je obdobje sodelovanja z banko Obd_s krajše od 65,5 mesecev. Če je odgovor pritrdilen in je bonitetna ocena enaka A, B, C1 ali C2, potem je pričakovana stopnja poplačila RR enaka 69,16 %. Če je bonitetna ocena C3, D, ali E, potem je pričakovana stopnja poplačila RR 50,73 %. Če je obdobje sodelovanja z banko Obd_s daljše od 65,5 mesecev in je pogodbena obrestna mera Pog_OM manjša od 12,4%, potem je pričakovana stopnja poplačila RR enaka 60,31 %. Če je pogodbena obrestna mera Pog_OM višja kot 12,4%, potem je pričakovana stopnja poplačila RR enaka 80,32 %.

Iz odločitvenega drevesa za obdobje 12 mesecev sklepamo naslednje:

1. Pri posojilih z višino posojila nad 56.000 EUR je pričakovano nižje poplačilo, saj je njihova stopnja poplačila najnižja od vseh stopenj poplačil v odločitvenem drevesu.

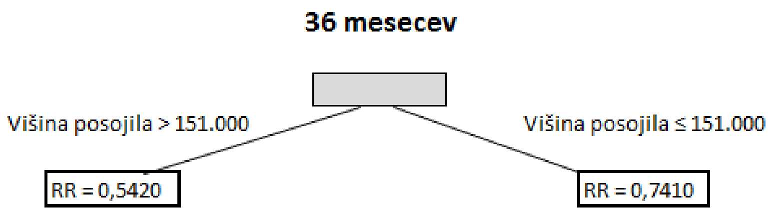
2. Če je višina posojila nizka (pod 56.000 EUR) in je obdobje poslovanja z banko dolgo (nad 65,5 mesecev), potem je tudi stopnja poplačila višja, če je pogodbena obrestna mera višja (nad 12,4 %).

3. Če je višina posojila nizka (pod 56.000 EUR) in je obdobje poslovanja z banko kratko (pod 65,5 mesecev), potem je stopnja poplačila višja, če je bonitetna ocena dobra (A, B, C1 ali C2).



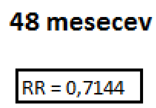
Slika 7.4: Odločitveno drevo za obdobje 24 mesecev. Slika povzeta iz [22].

Odločitveno drevo za obdobje 24 mesecev, ki je prikazano na sliki 7.4 je enostavnejše. Podobno kot v odločitvenem drevesu za obdobje 12 mesecev, je prva delitev odvisna od višine posojila. Mejna vrednost je v primerjavi s prejšnjim odločitvenom drevesom je višja ter znaša 76.000 EUR. Če je posojilo višje od 76.000 EUR je stopnja poplačila RR enaka 49,36 %. Če je posojilo nižje od 76.000 EUR in je obdobje sodelovanja z banko Obd_s daljše od 65,5 mesecev, je stopnja poplačila RR enaka 84,46 %. Če pa je obdobje poslovanja krajše od 65,5 mesecev, potem je stopnja poplačila RR enaka 63,75 %.



Slika 7.5: Odločitveno drevo za obdobje 36 mesecev. Slika povzeta iz [22].

Odločitveno drevo za obdobje 36 mesecev je še enostavnejše. Prikazano je na sliki 7.5. Zgrajeno je iz ene same binarne delitve, ki je ponovno odvisna od višine posojila. Če je višina posojila višja od 151.000 EUR, potem je stopnja poplačila RR enaka 54,20%, sicer pa je stopnja poplačila RR enaka 74,10 %. Ponovno opazimo, da imajo višja posojila nižjo stopnjo poplačil RR.



Slika 7.6: Odločitveno drevo za obdobje 48 mesecev. Slika povzeta iz [22].

Za obdobje 48 mesecev je odločitveno drevo zmanjšano le na koren drevesa, pričakovana stopnja poplačila RR pa je enaka povprečni stopnji poplačil celotne podatkovne osnove, kar pomeni, da je model odločitvenega drevesa kar enak modelu, ki pričakovano stopnjo izračunava na podlagi povprečnih historičnih podatkov.

Odločitveno drevo se z daljšanjem obdobja napovedi poenostavlja, kar je posledica zmanjševanja podatkovne osnove in povečevanje homogenosti stopnje poplačil (RR).

7.3 Napovedna moč modela

Za merjenje pričakovane napovedne moči zgrajenega modela na novih podatkih uporabljamo naslednja količnika:

- Koren povprečne kvadratne napake (RMSE)

$$RMSE = \sqrt{\frac{1}{n}\sum_i (r_i - \hat{r}_i)^2} \tag{7.2}$$

- Relativno absolutno napako (RAE)

$$RAE = \frac{\sum_i |r_i - \hat{r}_i|}{\sum_i |r_i - \overline{r}_i|} \cdot 100 \tag{7.3}$$

kjer z r_i označujemo dejansko stopnjo poplačila posojila i, z $\hat{r}_i$ pa napovedano stopnjo poplačila posojila i, z n pa označujemo število vseh poslov iz podatkovne osnove.

Dobro napovedno moč modela izkazujejo nizke vrednosti količnikov $RMSE$ in RAE, saj modeli z nizkimi vrednostmi količnikov $RMSE$ in RAE izkazujejo manjše razlike med dejanskimi in predvidenimi poplačili ter natančneje napovedujejo dejanska poplačila. Količnik RAE ocenjuje napako napovedi s primerjavo dejanske stopnje poplačila z relativno preprosto mero, ki je enaka povprečju dejanskih vrednosti. Vrednosti količnika RAE, ki so nižje od 100, pomenijo, da je model v meri absolutne napake boljši od preproste mere, povprečja dejanskih vrednosti.

Zelo pogosto je, da se modeli preveč prilegajo, zato morata biti količnika $RMSE$ in RAE merjena na vzrocu, ki je neodvisen od podatkov, ki so bili uporabljeni pri modeliranju. Uporablja se pristop, s katerim se podatkovna osnova naključno razdeli na dva dela: na učno in na testno množico. S tem se zmanjšuje podatkovna osnova za modeliranje, kar za modeliranje ni pozitivno, saj je za izdelavo optimalnega modela potrebno imeti čim večjo podatkovno osnovo. Prav tako je za natančno oceno napake napovedi potrebno imeti razmeroma velik testni vzorec. Pogoja sta si nasprotna, zato je težko najti optimalno razdelitev.

V izogib iskanju optimalne razdelitve podatkovne osnove je bil model iz študije testiran s pomočjo navzkrižne validacije, s čimer je bil model razvit na večini razpoložljivih podatkov, napovedna moč pa ocenjena na celotnih razpoložljivih podatkih. Na podatkovni osnovi je bila izvedena desetkratna navzrkižna validacija, pri čemer se podatkovna osnova razdeli na deset podvzorcev, devet se jih uporabi za modeliranje, enega pa za testiranje. Tak postopek se ponovi desetkrat, pri čemer se vsak od desetih podvzorcev uporabi za testiranje natanko enkrat. Rezultate desetih ponovitev navzkrižne validacije nato združimo, s čimer dobimo oceno napake na celotni podatkovni osnovi.

Da bi zmanjšali variabilnost, je bil na primeru postopek navzkrižne validacije ponovljen tisočkrat z uporabo naključnih generiranj desetih podvzorcev podatkovne osnove. Tabela 7.2 prikazuje povprečne vrednosti količnikov $RMSE$ in REA v 1000 ponovitvah ločeno za učno in testno množico. Za primerjavo je v zadnji vrstici napoved stopnje poplačila iz enostavnega historičnega povprečja. V oklepajih so navedeni standardni odkloni.

Model	12 mesecev		24 mesecev		36 mesecev		48 mesecev	
	RMSE	RAE	RMSE	RAE	RMSE	RAE	RMSE	RAE
In-sample	0,384	81,1	0,374	85,2	0,371	95	0,374	100
Out of sample	0,407	86,8	0,388	88,6	0,387	97,6	0,374	100
	(0,004)	(1,1)	(0,004)	(1,0)	(0,004)	(1,2)	-	-
Zgodovinsko povprečje	0,437	100	0,41	100	0,384	100	0,374	100

Tabela 7.2: Statistični podatki porazdelitve stopnje poplačila. Tabela je povzeta iz [22].

Napake na učni množici so ponavadi manjše kot na testni množici, saj modeli, ki se preveč prilegajo, dajejo preoptimistične ocene napovedne moči. Opazimo, da so za obdobja 12, 24 in 36 mesecev binarna drevesa boljši model kot preprosto historično povprečje, pri 48 mesecih pa gre za ista modela, saj je binarno drevo zgrajeno samo iz korena drevesa.

Poglavje 8

Zaključek

Upravljanje tveganj je eden izmed ključnih elementov poslovanja banke, saj je učinkovito upravljanje tveganj predpogoj za varno in stabilno poslovanje banke. V sklopu učinkovitega upravljanja tveganj morajo banke zagotoviti razpolaganje z ustrezno višino kapitala, ki nudi zaščito pred tveganji ter pokriva nepričakovane izgube.

Novi kapitalski sporazum prav iz tega razloga strmi k uporabi naprednejših ter na tveganja občutljivejših metod za določanje optimalne višine kapitala za pokrivanje izgub iz kreditnega tveganja. Bankam nudi možnost uporabe naprednega pristopa notranjih bonitetnih sistemov, ki temelji na lastnih modelih za ocenjevanje tveganj, kot sta modela za ocenjevanje verjetnosti neplačila in izgube ob neplačilu. Banke, ki uporabljajo napredni pristop notranjih bonitetnih sistemov, veljajo za bolj kredibilne, saj naprednejše pristope merjenja tveganj smejo uporabljati le tiste banke, ki izpolnjujejo številne pogoje ter so pod stalnim preverjanjem s strani nadzornikov.

Uporaba pristopa notranjih bonitetnih sistemov v prvi vrsti omogoča nižje regulatorne kapitalske zahteve in boljše upravljanje kapitala, saj temelji na lastni bančni oceni tveganja, ki je odvisna od lastnosti bančnega portfelja ter okolja, v katerem je banka pozicionirana.

Eden od pomembnih korakov na poti k uporabi pristopa notranjih bonitetnih sistemov je izgradnja lastnega modela za ocenjevanje izgube ob neplačilu, ki banki omogoča natančneje ocenjevanje pričakovanih izgub iz kreditnega tveganja. V magistrskem delu je predstavljena metoda ocenjevanja izgube ob neplačilu z gradnjo odločitvenega drevesa. Pot od korena do listov odločitvenega drevesa določa optimalno segmentacijo portfelja neplačnikov, delež izgube ob neplačilu pa je enak tehtanemu povprečju izgub ob neplačilu opazovanega segmenta.

Obstajajo še drugi pristopi k modeliranju izgube ob neplačilu. Zelo pogosto je modeliranje z regresijsko analizo, študije takšnega modeliranja pa najdemo v [23], [24] in [25].

Literatura

[1] Banka Slovenije, *"Poročilo o finančni stabilnosti"*, ogled 18. 3. 2016, `https://www.bsi.si/library/includes/datoteka.asp?DatotekaId=6301`.

[2] Banka Slovenije, *"Proces ocenjevanja tveganj (Javni del)"*, ogled 20. 8. 2015, `https://www.bsi.si/library/includes/datoteka.asp?DatotekaId=2437`.

[3] *Zakon o bančništvu (ZBan-2)* (2015) (Uradni list RS, št. 25/15).

[4] C. Marrison, *The Fundamentals of Risk Measurment* (The McGraw-Hill Companies, Inc., Europe, 2002), Pogl. The Basics of Risk Management, str. 1-12, Pogl. Introduction to Credit Risk, str. 226-230, Pogl. Types of Credit Structure, str. 231-253, Pogl. Risk Measurement for a Single Facility, str. 254-268, Pogl. Estimating Parameter Values for Single Facilities, str. 269-293

[5] *Sklep o upravljanju s tveganji in izvajanju procesa ocenjevanja ustreznega notranjega kapitala za banke in hranilnice* (2015) (Uradni list RS, št. 19/15).

[6] Abanka d.d., *"Letno poročilo 2014"*, ogled 20. 12. 2015, `http://www.abanka.si/vlagatelji/letna-porocila`.

[7] *"Informacijski sistem Sisbon"*, ogled 2. 2. 2016, `http://sisbon.si/sl/informacije-o-sistemu-sisbon/namen-sisbon`.

[8] J. Bessis, *Risk management in Banking* (Second Edition, Chicester, John Wiley & Sons Ltd, 2002), Pogl. VaR in Capital, str. 87-97, Pogl. Overview of Credit Risk Models, str. 420-431, Pogl. Credit Risk Drivers, str. 435-442, Pogl. Modelling Credit Risk Correlations, str. 566-579

[9] M. Jovan in M. Šušterič, *Statistično ocenjevanje verjetnosti neplačila za slovenska podjetja*, ogled 11. 9. 2015,`https://www.bsi.si/library/includes/datoteka.asp?DatotekaId=543`

[10] *Uredba (EU) št. 575/2013 Evropskega parlamenta in sveta z dne 26. junija 2013 o bonitetnih zahtevah za kreditne institucije in investicijska podjetja ter o spremembi Uredbe (EU) št. 648/2012* (CRR)

[11] S. Peura in J. Soinnen, *One, two, three, four types of default*, verzija marec 2005, ogled 18.8.2015, `papers.ssrn.com/sol3/Delivery.cfm?abstractid=685881`

[12] T. Schuermann, *What Do We Know About Loss Given Default?* Credit Risk Models and Management 2nd edition (ur. D. Shimko), UK: Risk Books, London, 2004, 249-274.

[13] A. Čargo in M. Štajner, *Minimalne zahteve za uvedbo IRB pristopa*, v: Banka Slovenije, Novi bančni standardi in ERM 2, 10. strokovno posvetovanje o bančništvu (2004) 89-105.

[14] S. Mičkovič in Š. Kern, *Izkušnja modeliranja izgube ob neplačilu*, Bančni vestnik **64** (2015) 19-23.

[15] Basel Committee on Banking Supervision (BSBS), *An Explanatory Note on the Basel II IRB Risk Weight functions*, verzija julij 2005, ogled 12.8.2015, `http://www.bis.org/bcbs/irbriskweight.pdf`.

[16] Banka Slovenije, *Na poti do Basla II*, ogled 2. 2. 2016, `https://www.bsi.si/library/includes/datoteka.asp?DatotekaId=502`.

[17] R. Chalupka & J. Kopecsni, *Modeling Bank Loan LGD of Corporate and SME Segments: A Case Study*, verzija 27/2008, ogled 16.8.2015, `https://www.econstor.eu/dspace/bitstream/10419/83329/1/589166727.pdf`.

[18] K. Ruohonen, *Graph Theory*, ogled 15.2.2015, `math.tut.fi/~ruohonen/GT_English.pdf`.

[19] Š. Kern, *Povezanost grafov*, delo diplomskega seminarja, Fakulteta za matematiko in fiziko, Univerza v Ljubljani, 2012.

[20] C. Shalizi, *Lecture 10: Regression Trees*, ogled 18.2.2015, `http://www.stat.cmu.edu/~cshalizi/350-2006/lecture-10.pdf`.

[21] T. M. Therneau, E.J. Atkinskon in M. Foundation, *An Introduction to Recursive Partitioning Using the RPART Routines*, ogled 14. 1. 2016, `https://cran.r-project.org/web/packages/rpart/vignettes/longintro.pdf`.

[22] J.A Bastos, *Forecasting bank loans loss-given default*, Journal of Banking & Finance **34** (2010) 2510-2517.

[23] S. Firestone in M. Rezende, *Are Banks' Internal Risk Parameters Consistent? Evidence from Syndicated Loans*(2013), ogled 12.3.2016, `http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2139539`.

[24] Marko Košak & Jure Poljšak, *Loss given default determinants in a commercial bank lending: an emerging market case study*(2010), ogled 7.10.2015, `hrcak.srce.hr/file/82434`.

[25] Andreas Wirenhammar, *Modeling Downturn LGD for a Retail Portfolio*(2010), ogled 29.10.2015, `http://www.math.kth.se/matstat/seminarier/reports/M-exjobb11/110222a.pdf`.