FAKULTA MATEMATIKY, FYZIKY
A INFORMATIKY UNIVERZITY KOMENSKÉHO
V BRATISLAVE

# Diplomová práca

Oľga Žilková

Bratislava 2006

FAKULTA MATEMATIKY, FYZIKY
A INFORMATIKY UNIVERZITY KOMENSKÉHO
V BRATISLAVE
Katedra aplikovanej matematiky a štatistiky

# Risk Based Pricing

Diplomová práca

**Diplomant** Oľga Žilková
**Vedúci diplomovej práce**: RNDr. Jiří Kottas, CSc.

Bratislava 2006

Prehlasujem, že túto prácu som vypracovala sama, iba s použitím uvedenej literatúry a s pomocou diplomového vedúceho.

Ďakujem svojmu diplomovému vedúcemu RNDr. Jiřímu Kottasovi, CSc.

# Obsah

# Úvod

Jednou z významných činností bánk prinášajúcich zisk je poskytovanie úverov. Úver možno charakterizovať ako dočasné zapožičanie peňazí - finančných prostriedkov právnickým alebo fyzickým osobám za vopred dohodnutých podmienok. Osoby uchádzajúce sa o úver obyčajne prechádzájú ponukami všetkých finančných inštitúcií poskytujúcich úver, za účelom dohodnúť čo najlepšie možné podmienky a cenu. Poskytovateľ takéhoto úveru, často banka[1], sa snaží rozumne profitovať z obchodu. Zisk z úveru možno rozdeliť na dve časti. Prvou a základnou časťou sú úroky. S poskytnutím úveru je však spojené riziko, že klient úver nebude schopný splatiť. Preto okrem štandardných úrokov musí byť do ceny úveru započítaná aj riziková marža, ktorá toto riziko zohľadňuje. Je prirodzené, že oba tieto aspekty sú navzájom prepojené. Čím väčšie je riziko, že poskytnutý úver nebude splatený, tým vyššiu úrokovú sadzbu si tento úver vyžaduje. Banka by mala byť v čase poskytovania úveru schopná určiť jeho kvalitu na základe analýzy úverovej spôsobilosti žiadateľa, a tak kvantifikovať úverové riziko spojené s obchodom. Ak banka nadhodnotí toto riziko, cena úveru opierajúca sa o túto analýzu bude príliš vysoká v porovnaní s cenami konkurenčných bánk a potencionálny dlžník sa bude uchádzať o úver niekde inde. Na druhej strane, ak banka podhodnotí toto riziko a bude schvaľovať úvery pochybným dlžníkom za cenu, ktorá je príliš nízka v porovnaní s rizikom spojeným s obchodom, výrazne tým prispeje ku znehodnoteniu portfólia a teda aj nižším ziskom. Preto opatrné a dôkladné zhodnocovanie úverového portfólia (credit risk assessment) je v bankovom sektore veľmi dôležité.

Cieľom diplomovej práce je predstaviť vhodný postup pre stanovenie cien úverov, ktoré budú čo najpresnejšie zohľadňovať bonitu klienta a s ním spo-

[1] v ďalšom texte bankami budeme nazývať všetky finančné inštitúcie poskytujúce pôžičky, i nebankové subjekty ako napríklad splátkové spoločnosti

jené úverové riziko pre banku. V literatúre je tento prístup nazývaný risk based pricing. Navrhnutý model v diplomovej práci sa skladá z troch na seba nadväzujúcich modelov. Skóringový model kvantifikuje riziko klienta, ktoré nám ďalej slúži ako vstup do druhého modelu - modelu segmentácie. Tento model zatrieďuje klientov do homogenných rizikových tried, pre ktoré tretí model, model rizikových nákladov, výpočíta očakávané náklady spojené s rizikom nesplatenia úveru alebo jeho časti. Tieto náklady sa vyjadrujú vo forme nutnej prirážky k základnej cene úveru.
Diplomová práca je rozdelená do troch častí. V prvej časti je vysvetlény model cenotvorby v bankách a sformulovný problém v praxi. Druhá časť je venovaná postupu zhotovenia risk based pricing a je rozdelená do troch hlavných celkov analyzujúce jednotlivé modely. Posledná časť je venovaná záverom a postrehom. Na koniec je priložený algoritmus na výpočet risk based pricing v programovacom prostredí SAS.

# Kapitola 1

# Úverové riziko a ceny úverov

## 1.1 Model cenotvorby v banke

Zmluvou o úvere sa banka ako veriteľ zaväzuje poskytnúť dlžníkovi v jeho prospech peňažné prostriedky do určitej sumy a dlžník sa zaväzuje poskytnuté peňažné prostriedky vrátiť a zaplatiť úroky. Úroky predstavujú cenu úveru, za ktorú veriteľ zapožičiava svoje peniaze. Celková výška úrokov, ktoré banka zinkasuje, musí byť vyššia ako jej náklady na získanie potrebných finančných prostriedkov. Úrokový rozdiel by mal banke priniesť zisk a zohľadniť mieru rizika, komplexnosť a náročnosť služieb, ktoré banka klientovi poskytla. I keď jednému klientovi môže každá banka na základe vlastnej cenovej politiky ponúknuť inú úrokovú sadzbu, vo všeobecnosti sa banky riadia rovnakými pravidlami.
Cenotvorbou nazývame proces, v ktorom sa na základe presne definovaných pravidiel určuje cena úveru. Napríklad jednoduchý model cenotvorby môže byť založený na predpoklade, že sa finálna úroková sadzba úveru skladá zo štyroch komponentov:

- **cena zdrojov komerčnej banky** - pokrýva náklady vzniknuté pri zabezpečovaní potrebných finančných prostriedkov. Často sa označuje ako cena finančných zdrojov ale tiež base rate.

- **operačné náklady pri procese úverovania** - pokrýva prevádzkové náklady, ktoré zahŕňajú žiadosti a proces splátok, mzdové náklady a iné výdavky spojené s údržbou. V praxi sa často tieto náklady pokrývajú rôznymi poplatkami spojenými s vedením úverového účtu

- **riziková marža** - pokrýva rizikové náklady spôsobené zo strát zlých obchodov, ktoré by mali byť vykryté príjmami z rizikových nákladov dobrých obchodov
- **zisková marža** - predstavuje ziskovú prirážku, ktorú si banky určujú v závislosti od toho, aké miesto chcú oproti konkurencii zaujať a aký zisk chcú dosiahnuť

Tento jednoduchý model je vhodný pre krátkodobé úvery. Pri stanovovaní základných úrokových sadzieb pre dlhodobé úvery (hypotéky) sa banky obvykle riadia cenami swapov na medzibankovom trhu.
Úverové riziko spojené s poskytnutím úveru sa v jednotlivých prípadoch líši v závislosti od úverového produktu a charakteristík dlžníka. Preto jedným z najdôležitejších aspektov pri tvorbe ceny úverov je správne nastavenie mechanizmu na stanovenie rizikovej marže v závislosti od spominaných charakteristík a produktu. Model cenotvorby, pri ktorom v rámci toho istého úverového produktu sú menej rizikoví klienti odmenení nižšou rizikovou maržou a tým aj nižšou úrokovou sadzbou, a s postupným zvyšovaním stupňa rizika sú pokutovaní za svoje horšie kreditné správanie stanovením vyššej rizikovej marže, nazývame rizikovo založeným, resp. stupňovitým spôsobom cenotvorby (*risk based pricing*).

## 1.2 Risk based pricing

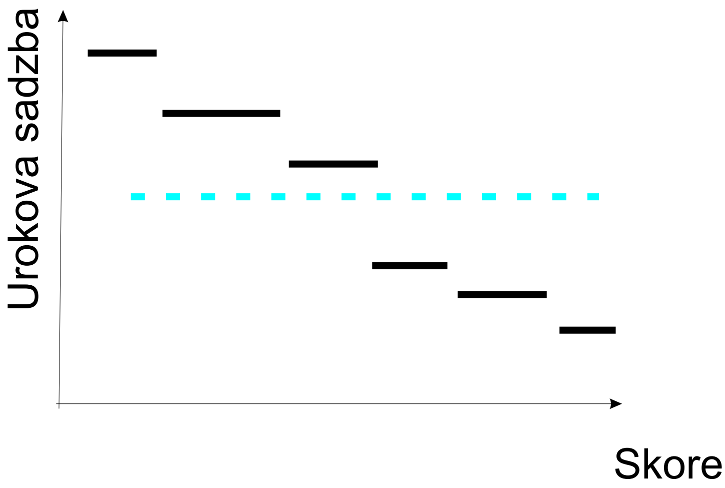
Risk based pricing je spôsob ocenenia úveru, ktorý kladie dôraz na riziko žiadateľa a stanovuje rôzne ceny úveru pre odlišne rizikových uchádzačov. Teda základ metódy tvorií kvantifikácia schopnosti dlžníka splatiť úver v súlade so zmluvnými podmienkami. Skóringová funkcia je nástroj pre túto kvantifikáciu. Je definovaná na množine všetkých možných charakteristík klienta (prípadne úveru). Jej funkčnou hodnotou je reálne číslo, tzv. skóre. Teda skóre je hodnota, ktorá (použitím vhodnej skóringovej funkcie) je prideľovaná klientom v závislosti od stupňa rizika, ktoré predstavujú pre banku. Riaďme sa pravidlom: menej rizikoví klienti dosahujú vyššie skóre v porovnaní s rizikovejšími klientami. Potom metóda risk based pricing stanovuje nižšiu úrokovú sadzbu klientom s vyšším skóre a naopak klienti s nižším

skóre dostávajú vysokú úrokovú sadzbu odrážajúcu možné narastajúce náklady a straty spojené so zapožičaním peňazí.
Stanovenie správnej ceny je pre banku kľúčové, pretože podhodnotenie úveru môže znamenať relevantné finančné straty pre banku a jeho nadhodnotenie stratu zákazníkov a celkového postavenia na trhu voči konkurencii. Správna implementácia metódy v praxi môže priniesť veľa výhod nielen pre banku, ale aj pre zákazníkov. Banka si zvýši svoje objemy a počty poskytnutých úverov, pričom redukuje nepriaznivý vplyv zo zlých obchodov. Dobrí klienti sú odmeňovaní nízkymi úrokovými sadzbami a veľmi rizikoví klienti dostávajú šancu dostať úver, hoci za menej výhodných podmienok.
Na nasledujúcom príklade si znázorníme myšlienku risk based pricing. Uvažujme portfólio úverov s poskytnutou výškou 100 000 korún a s dobou splatnosti 1 rok. V prípade, že banka neuvažuje stupeň rizika jednotlivých klientov, ale len riziko celého portfólia, tak cenu stanovuje rovnakú pre celé portfólio. V tomto prípade lepší klienti doplácajú na platobnú neschopnosť zlých klientov. Teraz zaveďme metódu risk based pricing a rozdeľme portfólio do niekoľkých skupín podľa rizika, ktoré predstavujú pre banku. Klienti v lepších skupinách dosahujú vyššie skóre v porovnaní s klientami v horších skupinách. Pre jednotlivé skupiny sa kvantifikuje toto riziko a určí cena zodpovedajúca tomuto riziku. Cena úverov vzhľadom na skóre by mohla potom vyzerať napriklad ako je znázornené na obrázku 1.1 .

Skôr ako sa pustíme do modelovania cien na základe vyššie uvedených pravidiel, je dobré si uvedomiť, čo vlastne také úverové riziko obnáša a ako toto riziko môžu banky kvantifikovať. Vo všeobecnosti sa banky riadia podľa návrhu nových bankových pravidiel Bazilej II, ktoré uverejnila medzinárodná organizácia centrálnych bánk (The Bank for International Settlements, skr. BIS). Tieto pravidlá sa týkajú kapitálových požiadaviek na riadenie rizík v bankách. V nasledovnej časti si povieme o zložkách úverového rizika v ponímaní Bazilejskej kapitálovej dohody (Basel Capital Accord). V našej práci sa nebudeme zaoberať týmito kapitálovými požiadavkami.

## 1.3 Úverové riziko a očakávané straty

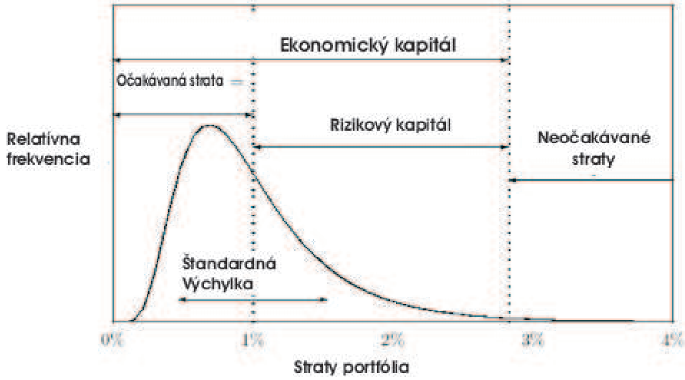
Banky chcú kvantifikovať úverovú spôsobilosť klienta, teda riziko, že úver nebude splatený, aby mohli tomuto riziku prispôsobiť ceny úverov. Riziko plynúce z portfólia bankou schválených úverov podľa Bazilej II možno roz-




Obr. 1.1: Risk based pricing

deliť na dve časti: (1) *očakávané straty* vyjadrené ako priemerné množstvo pozorovaných strát, ktoré budeme očakávať v budúcnosti ("Expected Loss"-EL)[1]; a (2) *maximálnu stratu*, ktorá sa môže vyskytnúť na vopred zadefinovanom intervale spoľahlivosti po tom, ako sú odpočítané očakávané straty ("Unexpected Loss" - UL).
Očakávané straty by mali byť zahrnuté v úrokovej sadzbe ako rizikové náklady a na neočakávané straty by sa malo prihliadať ako na skrytú možnosť podstatnej straty, ktorá sa môže vyskytnúť nečakanými výkyvmi v EL. Skutočné straty môžu byť väčšie, ale aj menšie ako očakávané a preto sa v úrokovej miere neočakávané rizikové náklady neuvažujú. To je úlohou rizikového riadenia banky, ktoré alokuje kapitál tak, aby pokryl neočakávané straty. Graf na obrázku 1.2 znázorňuje typické pravdepodobnostné rozdelenie strát z úverov počas zvoleného obdobia (štandardne jeden rok).

[1]Tento prístup je založený na predpoklade, že kvalita klientov z pohľadu schopnosti splatiť úver sa v čase nemení. Teda tento prístup nezohľadňuje napríklad makroekonomického či mikroekonomického prostredia




Obr. 1.2: Rozdelenie strát z úverov

Očakávané straty dlžníka potom vypočítame použitím formulky:

$$EL \;=\; PD\;.\;EAD\;.\;LGD \tag{1.1}$$

kde pravdepodobnosť zlyhania - defaultu ($PD$) je pravdepodobnosť, že dlžník zlyhá[2] za sledovaný časový horizont(napríklad jeden rok); expozícia v momente zlyhania ($EAD$) je očakávaná celková expozícia alebo nezaplatené čiastky úverov v čase zlyhania; a strata v prípade zlyhania ($LGD$) vyjadruje stratu ako percentuálny podiel z expozície v momente zlyhania. Je to doplnok tzv. miery vyťaženia úveru po nastatí udalosti zlyhania ($REC$).

$$LGD = 1 - REC$$

Na odhad očakávaných strát týmto spôsobom je nutné poznať vyššie uvedené tri elementy a koreláciu medzi nimi. Problému odhadu očakávaných strát bolo venovaných množstvo štúdií a publikovaných niekoľko prác za účelom správneho alokovania kapitálu pre úverové portfólio. Medzinárodna organizácia centrálnych bánk zadefinovala dva hlavné prístupy a to

1. štrukturálny - zlyhanie je spojené s kapitálovou štruktúrou firmy

[2]Podľa Bazilej II, stav zlyhania nastáva, ak nastane aspoň jedna z dvoch nasledovných situácií: prvou je, že veriteľ pokladá za nepravdepodobné, že dlžník splatí všetky svoje dlhy, a druhá situácia nastáva, ak je dlžník 90 dní po dátume splatnosti ľubovolnej splátky

2. reduced form - zlyhanie modeluje ako stochastický proces

Risk based pricing model, ktorému sa budeme venovať je založený na myšlienke reduced form.

## 1.4 Formulácia problému

Myšlienka risk based pricing modelu nie je nová. V praxi je bežne využívaná v poisťovníctve, a niekoľko rokov je model zaužívaný pri poskytovaní úverov korporátnym klientom (využívajú sa známe metódy ako creditmetrics, Moody's KMV model, option based model). Na Slovensku pomerne málo používaným modelom je rizikovo založený model cenotvorby pre spotrebiteľské úvery.[3] V diplomovej práci sa budeme venovať tejto problematike a budeme sa snažiť vyvinúť model vhodný pre toto prostredie. Pri tvorbe modelu musíme hlavne prihliadať na kvalitu dát, ekonomickú interpretáciu a relatívne ľahkú implementáciu. Základom modelu tvorí analýza predošlých úverov a predpoklad, že budúce úvery sa budú správať rovnako. Na Slovensku sa výrazne zvýšil dopyt po úveroch až v posledných rokoch. Preto sa v dátových skladoch bánk nachádza relatívne malé množstvo dát o poskytnutých úveroch vhodných na dôkladnú analýzu. Nízka kvalita dát (množstvo chýbajúcich, či nesprávne zadaných údajov), krátka história a rôzne staré portfólia sú problémy, na ktoré musíme pri tvorbe modelu prihliadať. Historicky odsledované správanie sa úverov je závislé aj od typu úveru a od podmienok, za akých bol úver poskytnutý. Teda každý typ úveru si vyžaduje vlastnú analýzu a model. My sa budeme snažiť zostaviť model aplikovateľný v praxi s relatívne ľahkou implementáciou. Vzhľadom na spomenuté obmedzenia sa budeme opierať o niekoľko predpokladov zjednodušujúcich situáciu. Jedným z najdôležitejších je predpoklad, že správanie úverovaných klientov sa v čase nemení, t.j. vypočítaná miera zlyhania klienta bude na jednotlivých splátkach rovnaká.
Úvery sa vo väčšine prípadov splácajú anuitným spôsobom, čiže rovnakými mesačnými splátkami. My sa tiež v práci obmedzíme na model vhodný pre splátkové úvery určené spotrebiteľom.

[3] Existujú prístupy kategorizácie klientov podľa ich kvality splácať úver, cena úveru však nie je exaktne určená podľa korektnej rizikovej marže.

# Kapitola 2

# Risk Based Pricing Model

## 2.1 Skóringový model

Skóringové modely v bankovníctve slúžia na rozlíšenie rizikových klientov žiadajúcich o úver využijúc rôzne dostupné informácie o klientovi. Výsledkom je skóre, ktoré zodpovedá úverovej spôsobilosti klienta. Hlavná idea skóringového modelu bola po prvý krát spomenutá už v roku 1936, kedy Fisher vo svojom článku vysvetlil možnosť rozlišovania skupín na základe ich odsledovaných charakteristík. Dlhé roky skóringové modely slúžili hlavne pri rozhodovaní, či skúmaný úver schváliť alebo zamietnuť, pričom tradičné metódy rozhodovania boli robené ľudským úsudkom o rizikovosti klienta, vychádzajúceho zo skúseností z predchádzajúcich rozhodnutí. Narastajúci dopyt po úveroch, rozšírenie konkurenčného trhu a objavenie nových výpočtových technológií viedlo k vývinu sofistikovanejších štatisticko - matematických skóringových modelov. Dnes skóringovou funkciou nazývame funkciu, ktorá dokáže predikovať, či potencionálny klient (v prípade aplikačného skóringu) alebo už existujúci dlžník (v prípade behaviorálneho skóringu) bude schopný plniť svoje záväzky voči banke poskytujúcej úver, alebo či sa stane pre ňu zlyhávajúcim klientom. Nech $Y$ je skúmaná náhodná veličina, ktorá nadobúda hodnotu 1, ak klient zlyhal za sledované obdobie, inak hodnotu 0. Potom skóringová funkcia je zobrazenie $f:\ X \subset R^n \rightarrow [0,1]$, ktoré v istom zmysle dokáže, čo najlepšie odlíšiť dobrého klienta od zlého klienta. Jednoznačná identifikácia dobrého a zlého (zlyhaného) klienta je pre jednotlivé finančné inštitúcie individuálna. Buď sa volí štandartná definícia zlyhania podľa Bazilej II alebo sa zadefinuje vlastná definícia podľa dostupnosti a povahy skúmanej vzorky úverov. Prvkami z množiny $X = \{x_1, \cdots, x_n\}^T$ môžu byť

- pomerové ukazovateľe skonštruované z finančných výkazov [1]
- osobné údaje o klientovi ako napríklad príjem, vek, zamestnanie, vzdelanie, atď.
- informácie z výpisu registra úverov alebo z bankových záznamov (ak žiadateľ je klientom banky) majúce vplyv na bonitu klienta.

Skóringová funkcia je zostavená podľa historicky napozorovaného správania sa vzorky úverov toho istého charakteru, pre ktorý je model zostavovaný. Predpokladáme, že povaha dát, na ktorých je skóringová funkcia vyvinutá, bude konzistentná s povahou budúcich úverov, na ktoré bude model aplikovaný.

## Stručný prehľad skóringových funkcií

Medzi najrozšírenejšie skóringové funkcie patria: lineárny pravdepodobnostný model, regresné modely, logistická regresia, modely založené na diskriminančnej analýze, neurónové siete, suport vector machines, matematické programovanie, rozhodovacie stromy, model k-teho najbližšieho suseda, analiticko hierarchické procesy, experné systémy, genetické algoritmy a ďalšie. Je ťažké s istotou povedať, ktorá metóda je najlepšia. Pri výbere funkcie musíme brať ohľad na povahu dát, t.j. lineárnosť, prípadne nelineárnosť, robustnosť k outlayerom, citlivosť k extrémnym hodnotám, robustnosť v čase, interpretovateľnosť výsledkov pre užívateľa a pod. Modely ako lineárny pravdepodobnostný model a regresné modely sú aplikovateľné len za silných predpokladov o vstupných dátach, ktoré v praxi často nie sú splnené. Neurónové siete majú oproti regresným modelom niekoľko výhod. Nevyžadujú si formuláciu predpokladov o rozdelení vstupných informácií, zároveň sú tolerantné k nekompletným a atypickým dátam a dokážu aproximovať ľubovolné nelineárne zobrazenie. Avšak vyžadujú veľké množstvo dát na tvorbu modelu. Ten nie je interpretovateľný, pretože vnutorné vrstvy sami o sebe nemajú

[1] v prípade, že skórujeme právnické osoby

špeciálny význam a tvoria tzv. čiernu skrinku. Novou metódou, ktorá rovnako ako neurónové siete, v sebe zahŕňa flexibilitu nelineárnych systémov je metóda podporných vektorov (support vector machines). Hlavná myšlienka tejto metódy spočíva v zobrazení vstupných dát prostredníctvom nelineárnej funkcie do vysoko dimenzionálneho priestoru, na ktorom sa pokúša nájsť nadrovinu, ktorá bude najlepšie klasifikovať klientov na skupinuy dobrých a zlých klientov. Výhodou je, že nie je nutné explicitne hľadať toto nelineárne zobrazenie, ale použijeme funkciu jadra zjednodušujúcu celý klasifikačný problém. Najnovšie štúdie sú venované práve tejto metóde a hľadaniu najvhodnejšej funkcie jadra.
Napriek novým štúdiám stále veľmi obľúbenou a najvhodnejšou metódou v praxi je logistická regresia. Jej výhodou, ktorá sa v praxi veľmi oceňuje, je fakt, že táto metóda priamo produkuje pravdepodobnosti zlyhania pre jednotlivých klientov a je veľmi ľahko interpretovateľná. Ďalšou jej výhodou je, že si dokáže ľahko poradiť s premennými spojitého aj kategorického typu. Na druhej strane však vyžaduje predpoklady o vstupných dátach (nesmie byť prítomná multikolinearita, outleiere, atď.) a preto je pri používaní metódy nutné sledovať splnenie uvedených predpokladov.

### 2.1.1 Logistická regresia

Skôr ako budeme aplikovať logistickú regresiu na náš klasifikačný problém, v krátkosti si ju predstavme.
Regresná analýza umožňuje charakterizovať vzťah medzi vysvetľovanou a vysvetľujúcimi premennými. Kým v lineárnej regresii je vysvetľovaná premenná spojitá, v logistickej regresii je kategorická. Logistická regresia pomocou vysvetľovacích premenných, ktoré môžu byť spojité alebo tiež kategorické, modeluje pravdepodobnosť $p_i$ špecifikovaného výstupu pre $i$-ty prípad. Vzťah medzi pravdepodobnosťou a vysvetľujúcou premennou je nelineárny a preto sú pravdepodobnosti transformované do tzv. logitov

$$logit(p_i) = \log\Big(\frac{p_i}{1-p_i}\Big)$$

Táto transformácia zároveň garantuje, že odhadnuté pravdepodobnosti sú v rozsahu medzi 0 a 1 tak, ako sa od definície pravdepodobnosti očakáva.
My sa snažíme vytvoriť skóringový model, ktorý bude vedieť odlíšiť dve skupiny, a to zlých klientov od dobrých. Teda vysvetľovaná premenná je dichotomická, t.j. môže nadobúdať len dve hodnoty. Hodnotu 0, ak je klient dobrý

a hodnotu 1, ak je klient zlyhaný. A logistickou regresiou chceme odhadnúť, s akou pravdepodobnosťou má klient tendenciu zlyhať. Vhodným logistickým modelom pre dichotomickú vysvetľovanú premennú $y_i \in \{0, 1\}$ je binárna logistická regresia. Označme závislé premenné ako $X_i = \{x_i^1, x_i^2, \cdots, x_i^n\}$ a odhadnutú pravdepodobnosť $p_i$. Z tvaru logistickej regresie pre $i$-ty prípad platí:

$$\log(\frac{p_i}{1-p_i}) = x_i^T\beta \;\Rightarrow\; p_i = \Pr(y_i = 1|x_i) = \;\Lambda(x_i^T\beta) = \frac{e^{x_i^T\beta}}{1+e^{x_i^T\beta}}$$

kde $X_i = \{x_i^1, \cdots, x_i^n\}$ sú vstupné dáta o klientovi, $y_i$ je náhodná veličina detekujúca zlyhanie u klienta a $\beta = (\beta_1, \beta_2, \cdots, \beta_n)^T$ je vektor odhadnutých parametrov udávajúce váhy (dôležitosť) jednotlivým vstupným hodnotám $x_i$. Skalárny súčin $x_i^T\beta$ predstavuje skóre. Na odhad parametrov sa požíva metóda maximálnej vierohodnosti, ktorá vypočíta také hodnoty parametrov, pri ktorých pravdepodobnosť nastatia skutočnej napozorovanej hodnoty, je najvätčšia možná. Za predpokladu nezávislosti medzi vysvetľujúcimi premennými, funkcia maximálnej vierohodnosti $L$ pre $N$ pozorovaní $y_1, \cdots, y_N$ s pravdepodobnosťami $p_1, \cdots, p_N$ má tvar:

$$L = \prod_{\{i|y_i=0\}} [1-\Lambda(x_i^T\beta)] \prod_{\{i|y_i=1\}} \Lambda(x_i^T\beta) \;= \prod_{i=1}^{N} [1-\Lambda(x_i^T\beta)]^{1-y_i}\Lambda(x_i^T\beta)_i^y$$

Logaritmus vierohodnej funkcie má tvar

$$\ln L = \sum_{i=1}^{N} [y_i \ln \Lambda(x_i^T\beta) + (1-y_i)\ln 1 - \Lambda(x_i^T\beta))]$$

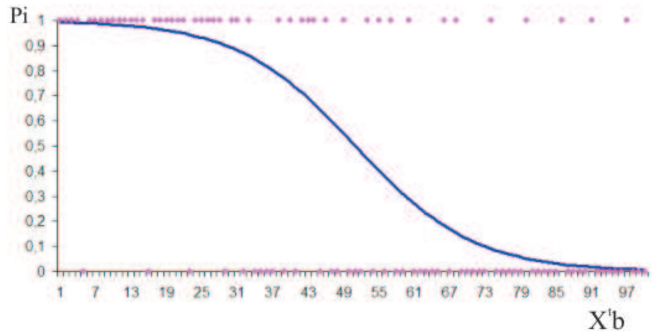
Pri hľadaní maxima spočítame deriváciu funkcie $\ln L$ vzhľadom na $\beta$. Riešime

$$\frac{\delta lnL}{\delta\beta} = \sum_{i=1}^{N} [y_i \frac{f(x_i^T\beta)}{\Lambda(x_i^T\beta)} + (1-y_i)\frac{-f(x_i^T\beta)}{1-\Lambda(x_i^T\beta)}]x_i^T = 0 \;,$$

kde $f(z) \equiv \delta\Lambda(z)/\delta z$. Po zjednodušení dostaneme

$$\frac{\delta lnL}{\delta\beta} = \sum_{i=1}^{N} (y_i - \Lambda(x_i^T\beta))x_i^T = 0 \tag{2.1}$$

Logaritmická vierohodnostná funkcia je globálne konkávna a preto riešením rovnice 2.1 je globálne maximum. Výraz v rovnici 2.1 je nelineárny vzhľadom




Obr. 2.1: Logistická regresia

na $\beta$ a preto riešenie vyžaduje špeciálne iteračné metódy (napr. Newton-Raphsonov typ algoritmu). Riešenie označme $\hat{\beta}$. Pripomeňme si, že funkcia $\Lambda(x_i^T\beta)$ predstavuje pravdepodobnosť zlyhania $p_i$ a nadobúda svoj obor hodnôt na intervale [0,1]. Graf na obrázku 2.1 znázorňuje priebeh logistickej funkcie vzhľadom na skóre $x_i^T\beta \in R^1$. Takto definovaná logistická funkcia zodpovedá bankovej konvencii, ktorej cieľom je lepšiemu klientovi prideliť vyšší počet bodov - vyššie skóre. Funkcia $\Lambda(x_i^T\beta)$ predstavuje pravdepodobnosť zlyhania.

V praxi je často vhodné, aby sa hodnota skóre pohybovala v určitom bankou stanovenom rozmedzí. Skalárny súčin $x_i^T\beta$ je možné preškálovať napríklad na stupnicu 0 až 100 a budeme ho nazývať funkciou skóre daného klienta. Označme pôvodné skóre $s_{pov}$. Platí $s_{pov} = x_i^T\beta$. Nech $s_{min}$ je minimálna hodnota spomedzi všetkých pôvodných skóre a $s_{max}$ je takáto maximálna hodnota. Potom výsledné skóre $S \in [0, 100]$ získame nasledovnou transformáciou:

$$S(s_{pov}) = \frac{100 s_{pov}}{s_{max} - s_{min}} - \frac{100 s_{min}}{s_{max} - s_{min}} \tag{2.2}$$

Postupom uvedeným v tejto časti, dostávame skóre pre jednotlivých dlžníkov, ktoré odráža ich úverové riziko pre banku. Odhadujeme parametre $\beta_i$ pre jednotlivé vysvetľujúce premenné $X_i$ zahrnuté do modelu. Je nereálne, aby množinu vysvetľujúcich premenných tvorili všetky dostupné informácie o klientovi. Je dôležité vedieť, ktoré charakteristiky klienta, prípadne úveru použiť ako nezávislé premenné a v akom tvare, aby model dokázal čo najlepšie predvídať zlyhanie klienta a zároveň chceme dosiahnuť spoľahlivé výsledky. Aby sme sa vyhli skresleným odhadom, musí byť splnených niekoľko predpokladov, ktoré si logistická regresia vyžaduje (nesmie byť prítomná multikolinearita medzi vysvetľujúcimi premennými, nesmú byť v dátach outleiere, musí byť splnená linearita medzi vysvetľovanou premennou a logitovou transformáciou pravdepodobnosti nastatia pozorovaného javu). Preto skôr ako budeme vytvárať model pre rôzne vysvetľujúce premenné, musíme ich rozanalyzovať.

### Výber premenných do modelu

V tejto sekcii práce sa venujeme analýze a výberu vhodných premenných - potencionálnych vstupov do skóringovej funkcie, vo vhodnom tvare. Pripomeňme si, že premennými môžu byť rôzne dostupné informácie o klientovi (vek, prijem, vzdelanie, počet detí, koľko klient vlastní aút, domov, úverová spôsobilosť klienta na prípadných predošlých úveroch, ... ). Niektoré premenné sami o sebe nedokážu dobre vysvetliť skúmaný výstup, avšak pod vplyvom inej premennej sa ich predikčná schopnosť zvyšuje. Preto sledujeme takéto vzájomné pôsobenie jednotlivých premenných a do modelu zahrňujeme rozličné, avšak zmysluplné kombinácie (slobodní stredoškoláci, 30 ročný s tromi deťmi, a pod.).

Predpokladom logistickej regresie je lineárny vzťah medzi spojitou vystvetľujúcou premennou a logitovou transformáciou pravdepodobnosti. Ak predpoklad linearity v logite je vychýlený, logistická regresia podhodnocuje stupeň asociácie medzi vysvetľovanou a vysvetľujúcimi premennými. Jedným z možných riešení, aby sme zabránili resp. zmiernili tento nedostatok linearity, je rozbitie spojitej náhodnej premennej do niekoľkých kategórií, čím dostaneme separované logity pre rôzne stupňe vysvetľujúcej premennej.

Ďalším predpokladom logistickej regresie je fakt, že nesmie byť prítomná multikolinearita. Inými slovami, vysvetľujúca premenná zahrnutá v modeli nesmie byť lineárnou kombináciou inej vysvetľujúcej premennej, ktorú tiež uvažujeme v modeli. Multikolinearita nemení odhady parametrov, len ich spoľahlivosť. Vysoké štandardné odchýlky detekujú prítomnosť multikolinearity v modeli. Na zmeranie korelácie medzi vysvetľujúcimi premennými možno použiť jednu z nasledovných troch testov a príslušných štatistík [1]

1. Pearsonovu $\chi^2$ štatistiku, ak skúmame asociáciu medzi dvomi nominálnymi kategorickými premennými. Veľkosť sily asociácie možno zmerať Cramerovou V štatistikou
2. Mantel-Haenszel $\chi^2$ , ak skúmame asociáciu medzi dvomi ordinálnymi kategorickými premennými. Veľkosť sily asociácie môže byť meraná Spearmanovým korelačným koeficientom
3. Analýzu variancie (jednostranný test ANOVA), ak skúmame asociáciu medzi kategorickou a spojitou premennou. Na určenie veľkosti asociácie možno použiť Kruskal-Wallisovu H štatistiku

V prípade identifikácie multikolinearity medzi dvomi premennými, možno k problému pristupovať tak, že z modelu vylučujeme premennú s menšou predikčnou silou.

Ďalším problémom, ktorý môže spôsobiť skreslené vystupy, je výskyt outlierov vo vstupných dátach. Outliere sú atypické, extrémne hodnoty. Tieto extrémne hodnoty treba vhodnou analýzou identifikovať a následne ich buď nahradiť inou hodnotou, alebo tieto pozorovania jednoducho nezahrnúť do vývoja modelu alebo ich analyzovať osobitne.

Na vykrytie nedostatkov v rozdelení vstupných dát, navrhol Falkenstein[3] niekoľko rôznych transformácii. Medzi inými napríklad

1. nahradenie pozorovaných premenných ich percentilmi (pre premenné prevažne spojitého typu)
2. aplikovanie sigmoidovej funkcie $\frac{1}{1+e^{-x}}$
3. využitie neparametrického odhadu zlyhania pre jednotlivé premenné

Pri menej kvalitných vstupných dátach je tiež potrebné si vedieť poradiť s chýbajúcimi hodnotami a nekorektne zadanými údajmi.

Po úprave vstupných dát získavame množstvo premenných, ktoré by mohli byť v modeli použité. Nás zaujíma, ako jednotlivé premenné dokážu vysvetľovať skúmaný výstup, v našom prípade zlyhanie dlžníka. Pre lepšiu predstavu je vhodné skúmať predikčnú silu jednotlivých premenných. Jednou z možností, ako odmerať túto silu, je spustenie jednorozmernej logistickej regresnej analýzy pre každú vysvetľujúcu premennú.

Na výber premenných po základnej jednorozmernej analýze možno použiť automatický proces - viackrokovú metódu logistickej regresie. Tento proces je v praxi veľmi obľúbeným a často využívaným.

### Viackroková metóda logistickej regresie

Logistická regresia volí množinu náhodných premenných, ktoré najlepšie vysvetľujú správanie klienta, aplikovaním viackrokovej metódy. Táto viackroková procedúra vyberá vhodných kandidátov (pri napredujúcej metóde), resp. vylučuje zlých kandidátov (spätná metóda) kontrolou ich štatistickej významnosti pre model. Túto štatistickú významnosť meria použitím vierohodnostného pomeru (Likelihood Ratio). LR štatistika je definovaná ako dvojnásobok logaritmu vierohodnostných pomerov dvoch modelov, pričom model 1 predstavuje model so zahrnutím skúmanej vysvetľujúcej premennej a model 2 túto premennú neuvažuje.

$$LR = -2\ln\Big(\frac{L(model2)}{L(model1)}\Big)$$

LR má asymptotické $\chi^2$-rozdelenie, kde počet stupňpv voľnosti je rovný rozdielu medzi počtom odhadovaných parametrov v daných dvoch modeloch. Celý krokový proces končí, keď by už žiadna nová pridaná vysvetľujúca premenná model už významne nezmenila. Pri výbere finálneho modelu spomedzi modelov vytvorených počas jednotlivých krokov volíme jednu z nasledovných stratégií

1. volíme **posledný model**, t.j. model vygenerovaný v poslednom kroku

2. volíme **model s najnižšou AIC**. Akaike Information Criterion (AIC) je jednou zo štatistík vypovedajúcich o efektivite a schopnosti modelu vysvetliť skúmanú závislú premennú. Čím nižšiu hodnotu dosahuje táto štatistika, tým máme lepší model.

$$AIC = -2\ln L + 2 * k$$

kde $k$ je počet koeficientov v skúmanom modeli a $\ln L$ je logaritmus vierohodnej funkcie

3. volíme **model s najnižšou BIC**. Bayesian Information Criterion (BIC) je tiež štatistikou používanou pri porovnávaní vhodnosti modelu. Tiež platí pravidlo, čím nižšia hodnota štatistiky, tým lepší model.

$$BIC = -2 \ln L + k * ln(n)$$

kde $n$ je celkový počet pozorovaní v modeli

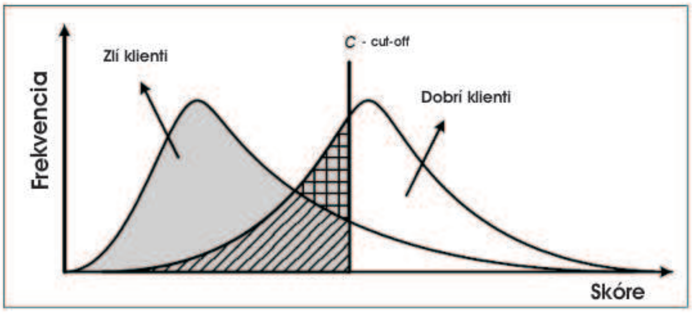
V praxi sa nakoniec z množiny vybratých parametrov odstraňujú tie, ktoré nezodpovedajú vhodnej ekonomickej interpretácii, napríklad majú opačné znamienko korelácie s vysvetľovanou premennou, akoby bolo pre ekonomickú interpretáciu prijateľné. Na záver musíme ešte určiť kvalitu modelu v zmysle schopnosti separácie dobrých klientov od zlých klientov, na základe čoho sa rozhodujeme či vytvorený model je vhodným nástrojom pre rozhodovanie v banke.

### 2.1.2 Posudzovanie kvality modelu a štatistické vlastnosti

Ako sme už načrtli, kvalitu vyvinutého skóringového modelu posudzujeme podľa diskriminančnej sily: schopnosti modelu separovať dobrých klientov od zlých. Túto diskriminančnú silu potrebujeme kvantifikovať. Aby sme sa vyhli vychýleným a skresleným výsledkom, validácia by sa mala robiť na dátach (tzv. testovacia vzorka), ktoré by nemali obsahovať údaje použité pri vývine modelu (out of sample) a mali by sa aj časovo líšiť[2] (out of universe). Pri nedostatku dát na testovanie je vhodnou alternatívou použitie bootstrap rozdelenia. Bootstrap je rozdelenie skúmanej štatistiky - v našom pripade diskriminančnej sily, ktoré vznikne niekoľkonásobným prevzorkovaním pôvodnych dát. Bližšie sa tejto metóde budeme venovať v tretej časti modelu pri určovaní rizikových nákladov.

Prehľad o kvalite modelu si môžeme spraviť nanesením do jedného grafu (1) empirické rozdelenie dobrých dlžníkov a (2) empirické rozdelenie zlých vzhľadom na skóre. Model dokáže tým lepšie rozlišovať medzi týmito dvomi skupinami klientov, čím menej sa prekrývajú tieto dva rozdelenia. Prekrývajúcou sa populáciou dobrých a zlých klientov vznikajú dva druhy chýb

[2] ak chceme testovať aj stabilitu v čase




Obr. 2.2: Empirické rozdelenie dobrých a zlých klientov

vzhľadom na určitú hranicu skóre, a to chyba I. druhu (kedy je dobrý klient identifikovaný ako zlyhaný) a chyba II. druhu (kedy je zlyhaný klient považovaný za dobrého). V praxi si rizikový menežment na základe miery rizika, ktoré je ochotný akceptovať pevne stanoví hranicu (tzv. cut-off) a klientov s dosiahnutým skóre pod túto hranicu bude zamietať.

Na zmeranie diskriminančnej sily modelu bolo vyvinutých niekoľko metód, ako Kolmogor-Smirnov test, ROC krivka, Gini index, CAP krivka, Informančná entropia, Hosmer-Lemeshov test. Vysvetlenie jednotlivých testov je nad rámec tohoto textu.[3] Pretože vypovedacie vlastnosti jednotlivých metód sú približne na rovnakej úrovni, voľba metódy nie je kľúčová. My sme si na analyzovanie kvality modelu zvolili ROC krivku s GINI koeficientom, ktoré si teraz v krátkosti popíšeme.

## ROC krivka a GINI koeficient

ROC (Reciever Operating Characteristics) nanáša do grafu kumulatívne percento dobrých klientov ku kumulatívnemu percentu zlých klientov vzhľadom na meniace sa hranice skóre. Ekvivalentom je zobrazenie chyby I.druhu ku 1 mínus chybe II.druhu vzhľadom na rôzne nastavenia hraníc skóre. Čím je ROC krivka bližšie k bodu (0,1), tým lepšie dokáže model rozlišovať me-

[3]Rozpracované validačné metódy možno nájsť v publikácii Basel Commitee on Banking Supervision: Studies on the Validation of Internal Rating Systems

dzi dobrými a zlými klientami. ROC krivka ideálneho skóringového modelu (bez výskytu chyby I.druhu) je nanesená pozdĺž vertikálnej osi do bodu (0,1) a potom priamo do bodu (1,1). V prípade modelu, ktorý nedokáže rozlišovať medzi klientami a skóre prideľuje náhodne, ROC krivka kopíruje diagonálu.(Pomer dobrých a zlých klientov je vzhľadom na meniacu sa hranicu skóre stále rovnaký.) Preto kvalitu modelu určuje oblasť AUC, oblasť medzi ROC krivkou a diagonálnou krivkou. Viac intuitívna interpretácia oblasti pod ROC krivkou je nasledovná. Nech $N_D$ je celkový počet zlyhaných klientov; $N_{ND}$ je celkový počet dobrých klientov. Vytvoríme $N_D * N_{ND}$ dvojíc: každý dobrý klient je spárovaný s každým zlým klientom. Naša skúmaná štatistika je proporcionálna časť správnych dvojíc (lepší klient má vyššie skóre ako horší klient) ku všetkým dvojiciam[4].
Nevychýlený odhad $\hat{U}$ pre AUC je:

$$\hat{U} = \frac{1}{N_D.N_{ND}} \sum_{(D,ND)} u_{D,ND}$$

kde

$$u_{D,ND} = \{ \begin{array}{cc} 1 & \text{ak} S_D < S_{ND} \\ \frac{1}{2} & \text{ak} S_D = S_{ND} \\ 0 & \text{ak} S_D > S_{ND} \end{array} \tag{2.3}$$

kde $S_D$ je skóre zlyhaného klienta; a $S_{ND}$ je skóre dobrého klienta. Veľkosť AUC dosahuje hodnoty medzi 0,5 a 1. Preškálovaním dostávame odhad Giniho koeficienta.

$$\widehat{GINI} = 2 * AUC - 1$$

Vzťah 2.3 možno prepísať do nasledovného tvaru, ktorý využívame pri posudzovaní kvality modelu.

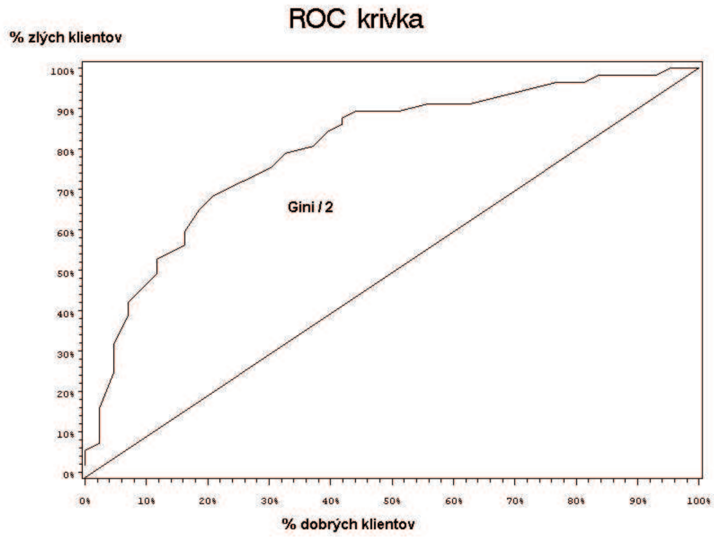
$$\hat{U} = \frac{X_{ND} - N_{ND}(N_{ND}+1)/2}{N_D.N_{ND}}$$

Potom

$$\widehat{GINI} = 2 * \frac{X_{ND} - N_{ND}(N_{ND}+1)/2}{N_D.N_{ND}} - 1 \tag{2.4}$$

kde $N_D$ je celkový počet zlyhaných klientov; $N_{ND}$ je celkový počet dobrých klientov; a $X_{ND}$ je súčet poradových čísel dobrých klientov zostupne zoradených podľa dosiahnutého skóre.

[4] Mann-Whitneyho U štatistika




Obr. 2.3: ROC krivka a GINI index

Hodnota AUC ako aj GINI index sú ovplyvnené náhodnými výkyvmi vo vstupných dátach. Aby sme zachytili výchylky spôsobené týmito výkyvmi, je vhodné zostrojiť aj intervaly spoľahlivosti pre lepšiu predstavu kvality modelu. Vhodným spôsobom na zostavenie takýchto intervalov spoľahlivosti je bootstrap metóda, ktorej sa bližšie budeme venovať v časti 2.3.1.

Na to, aby mal model dostatočnú diskriminačnú silu by bolo vhodné, aby gini dosahoval aspoň hranicu okolo 70%, ale v praxi to býva ťažko dosiahnuteľné kvôli rôznej kvalite dostupných dát. Každá banka si určí vlastnú hranicu giniho, pri ktorej bude model pokladať za vierohodný a implementovateľný ako vhodný nástroj pre posúdenie žiadateľov o úver. Teda v takomto prípade hodnotu skóre považujeme za spoľahlivú informáciu o úverovej spôsobilosti klienta. Na základe tejto hodnoty skóre, sa budeme snažiť rozdeliť celé skúmané portfólio do niekoľkých pásiem, tzv. segmentov tak, aby klienti v rámci jednotlivých segmentov znamenali pre banku približne rovnakú mieru rizika. Toto rozdelenie portfólia na základe skóre je úlohou druhého modelu procesu risk based pricing.

## 2.2 Model segmentácie

V tejto časti modelu budeme zostavovať rizikové triedy, do ktorých budú začleňovaní dlžníci homogénni v rámci tej istej skupiny vzhľadom na očakávané zlyhanie. Vstupom do modelu je odhadnutý stupeň očakávaného zlyhania (hodnota skóre dlžníka) a skutočná pozorovaná schopnosť klienta splatiť úver (nula jednotkový identifikátor, či klient zlyhal alebo nie). Na konštrukciu takéhoto systému tried budeme hľadať hraničné hodnoty skóre, ktoré budú optimálnym rozhodovacím pravidlom pre zaradenie klientov do jasne odlišiteľných tried. Pri segmentácii je dôležitá správna voľba počtu stupňov, ich definícia, rozsah pravdepodobnosti zlyhania ako aj počet dlžníkov v triede. Vzhľadom na odhadnutú pravdepodobnosť zlyhania (PD), s narastajúcim počtom stupňov sa zmenšuje interval PD, čím dostávame viac homogénne stupne. Avšak čím menší je počet dlžníkov v triede, tým väčší je odhad chýb pre PD. Preto je potrebné nastaviť toľko tried a zadefinovať ich takým spôsobom, ktorý minimalizuje sumu odchýlok.

### 2.2.1 Konštrukcia rizikových tried

Pri konštrukcii homogénnych rizikových tried je nutné dodržať niekoľko pravidiel:

1. miera zlyhania (default rate) pre jednotlivé triedy by mala mať klesajúcu tendenciu v smere od najhorších stupňov klientov k najlepším (v našom prípade od najnižšieho skóre k najvyššiemu),

2. suma štvorcov chýb odhadnutej miery zlyhania od skutočného napozorovaného stavu klienta pozdĺž celej škály skóre by mala byť minimálna

3. každá trieda by mala obsahovať primeraný počet dlžníkov, koncentrácie klientov v triedach by sa nemali veľmi líšiť a mali by zodpovedať pravidlám vytvoreným bankou

Jeden z možných prístupov pre stanovenie hraníc medzi jednotlivými pásmami, algoritmus využívaný v praxi, je izotónna regresia nastavená Pavlom Charamzom. Izotónna regresia je krivka preložená cez nula jednotkové pozorovaniamia vzhľadom na skóre tak, aby (1) miera zlyhania skupiny s vyšším skóre nebola vyššia ako miera zlyhania skupiny s nižším skóre a (2) miera

zlyhania triedy bola maximálna vzhľadom na zostávajúcu škálu skóre. Výstupom izotónnej regresie je monotónna funkcia (v našom prípade nerastúca - s rastom skóre pravdepodobnosť zlyhania klesá), ktorá minimalizuje súčet štvorcov odchýlok od zistených nula - jednotkových pozorovaní identifikujúcich dobrých a zlých klientov.
Postup pre určenie hodnoty izotónnej regresie v bode $j$ je nasledovný. V prvom kroku zoradíme dlžníkov vzostupne podľa skóre, ktoré im prislúcha. Označme
$X_j$ - maximálne poradie dlžníka majúceho skóre $j$
$D_{ij}$ - počet zlyhaných klientov so skóre spadajúceho do intervalu $[i, j]$
Potom maximálna frekvencia zlyhania pre skupinu klientov s minimálnym dosiahnutým skóre $j$ je daná vzťahom

$$\max_{i \geq j} \Big\{ P_j \Big| P_j = \frac{D_{ji}}{X_i - X_{j-1}} \Big\}$$

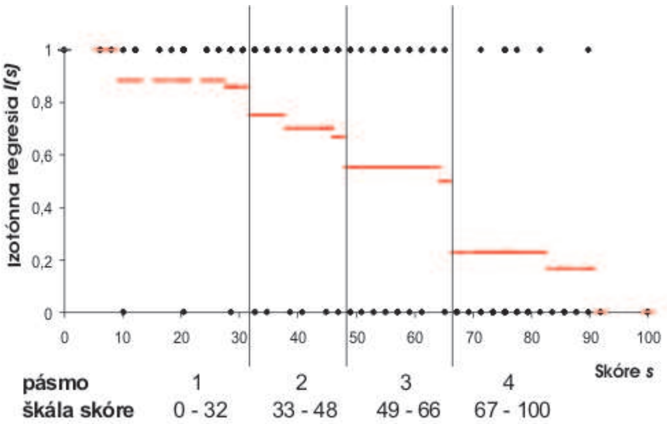
Vzhľadom na pravidlo (1), hodnota izotónnej regresie v bode $j$ je daná rekurzívnym vzťahom:

$$I_j = \min\Big( I_{j-1}, \max_{i \geq j} \Big\{ P_j \Big| P_j = \frac{D_{ji}}{X_i - X_{j-1}} \Big\} \Big) \tag{2.5}$$

kde $I_0 = 1$ a $j = 1, \cdots, 100$
Týmto postupom dostaneme pre každého klienta odhadnutú pravdepodobnosť zlyhania ako hodnotu izotónnej regresie v bode zodpovedajúcemu skóre klienta.

Vynesením izotónnej regresie do grafu si môžeme urobiť vizuálne predstavu o vhodnosti počtu skóringových pasiem a hodnote hraníc medzi týmito pásmami. Na grafe je znázornená ukážka izotónnej regresie, kde vertikálne čiary predstavujú možné rozhodovacie pravidlo pri určovaní homogénnych rizikových tried.

Hranice ukladáme do miest najväčších zlomov, na hodnotu skóre, pri ktorej sa miera zlyhania líši maximálne. Za predpokladu, že takáto skupina obsahuje dostatočný počet pozorovaní, hranice môžeme použiť na konštrukciu rizikových tried.




Obr. 2.4: Izotónna regresia

Špecifickou vlastnosťou izotónnej regresie je, že detekuje iba zlomy. Stanovenie hraníc je pomerne subjektívne. Navyše, ak izotónna regresia neposkytne vo funkčných hodnotách dostatočné rozdiely, odporúča sa použiť vhodnú modifikáciu prístupu.

## 2.3 Model rizikových nákladov

Model rizikových nákladov je posledným modelom potrebným na určenie rizikovej marže. Pripomeňme si, že riziková marža stanovená pre dlžníka je časť ceny úveru kompenzujúca riziko skupiny klientov, do ktorej je dlžník zaradený (použitím pravidla vychádzajúceho z modelu segmentácie) na základe prideleného skóre (použitím skóringovej funkcie). Analýzou správania historicky poskytnutých úverov podľa ich štrukturálnych vlastností (doba splatnosti úveru, výška poskytnutého úveru, výška zabezpečenia úveru) sa ľahko presvedčíme o ich vplyve na úverové riziko. Rozanalyzujme si jednotlivé vlastnosti:

**Doba splatnosti úveru.** S narastajúcou dĺžkou časového horizontu, na ktorý sú peniaze požičané, narastá neistota voči dlžníkovi, či bude schopný splatiť úver.

**Výška poskytnutého úveru.** Klienti žiadajúci o väčší objem pôžičky by mali dostávať vyššie úrokové sadzby ako dôsledok vyššej expozície v okamihu zlyhania.

**Zabezpečenie úveru.** Ide o akúsi zábezpeku obchodu požadované bankou pre prípad zlyhania klienta. Hlavnou funkciou zábezpeky je presunúť časť úverového rizika na zákazníka alebo na tretiu osobu. Keď zákazník založí zásoby, zariadenia alebo iné aktíva, alebo keď má ručiteľa pri získaní úveru z banky, banka prenáša časť rizika zo straty pri zlyhaní na zákazníka alebo ručiteľa.

V prvých dvoch prípadoch sme si uviedli, že cena úveru s dlhšou dobou splatnosti, resp. s vyšším poskytnutým objemom by mala byť vyššia. Toto však nemusí byť pravidlom. Analýzou akéhosi úverového portfólia sa môže ukázať, že úvery poskytnuté na dlhšiu dobu sa správajú lepšie než krátkodobé úvery. Analogicky nevieme s istotou posúdiť ani vplyv výšky poskytnutého úveru. Je potrebné rozdeliť skúmané portfólio na ešte dosť veľké, ale rozdielne skupiny podľa týchto dvoch faktorov. Analýzu rizikových nákladov potom robíme pre jednotlivé skupiny vzhľadom na skóre, dobu splatnosti a výšky poskytnutej pôžičky.

Akonáhle sme si portfólio rozdelili do jednotlivých tried, potrebujeme si vyjadriť rizikové náklady a rizikovú maržu jednotlivých tried. Rizikové náklady, označme $RC_{t_i}$ v danom časovom okamihu $t_i$ chápeme ako časť platieb, ktorá nebola zaplatená dôsledkom zlyhania klienta. Napríklad, nech riziková trieda pozostáva z dvoch úverov na 100 000 korún. Od dlžníkov sa očakávajú platby vo výške 1000 korún na mesačnej báze. Povedzme, že na druhej splátke zlyhá jeden úver a teda dostanem splátku iba od jedného dlžníka. Potom rizikové náklady na druhej splátke budú rovné

$$RC_2 = 1/2 * 1000 = 500 \text{ korún}$$

V prípade, že časť obchodu je zabezpečená, tak straty spôsobené nastatím udalosti zlyhania, sa znižujú o túto časť. Potom finálne rizikové náklady $FRC_{t_i}$ v čase $t_i$ dostaneme ako

$$FRC_{t_i} = RC_{t_i} * \frac{\text{nezaistená časť úveru}}{\text{výška poskytnutého úveru}}$$

Nech v našom príklade je zlyhaný úver Nech v našom príklade je zlyhaný úver istený vo výške 50 000 korún. Potom finálne rizikové náklady rizikovej skupiny na druhej splátke sú

$$FRC_2 = RC_2 * 50000/100000 = 500 * 1/2 = 250 \text{ korún}$$

Pre jednoduchosť, budeme ďalej v práci predpokladať, že úver je nezabezpečený. Teda $FRC_{t_i} = RC_{t_i}$.
V nasledovnej časti si uvedieme postup na výpočet rizikových nákladov. Uvedený postup po aplikácii na jednotlivé skupiny klientov, dostávame, čo sme chceli.

### 2.3.1 Model rizikových nákladov

Skôr než predstavíme model, stručne si popíšme priebeh úverového procesu pri splátkových úveroch. V čase $t_0$ veriteľ požičia finančné prostriedky vo výške $V$, ktoré dlžník po zmluvnej dohode bude postupne splácať v určitých časových okamihoch $t_1 < t_2 < \cdots < t_n$, kde $n$ je dohodnutý počet splátok a $t_n = T$ je maturita úveru. V prípade jednorázovo splatného úveru $t_n = t_1 = T$. V praxi sa obyčajne v čase schvaľovania úveru stanovuje pevná výška splátky a pevne stanovené časové body, v ktorých sa očakáva prítok peňazí od dlžníka. O túto skutočnosť sa opierame aj v našom modeli. Očakávame prítok peňazí od dlžníka podľa fixného amortizačného plánu. Nech $V$ je poskytnutý objem, $P$ fixná výška splátky a $n$ počet časových periód dohodnutých na splácanie. Potom platí:

$$V = \sum_{j=1}^{n} \frac{P}{(1+i_m)^j}$$

kde $i_m$ je pevne stanovená cena úveru za jednu periódu.
Cela situácia môže byť interpretovaná aj z iného pohľadu: dlžníkov dlh v čase $t_0$ je $V_0$ (t.j. objem, ktorý mu veriteľ poskytol). V čase $t_1$ dlžník zaplatí úroky za svoj dlh a výšku istiny (=výška dohodnutej splátky znížená o výšku úrokov) a tým sa mu dlh zníži o túto výšku istiny. Dlžník v čase $t_1$ akoby svoj dlh splatil a vzniká mu nový dlh $V_1$, ktorý musí zaplatiť. Celý proces trvá, kým nebude dlh nulový. (držiac sa zmluvnej dohody $V_n = 0$). Takto interpretujeme nezaplatenú časť dlhu v čase $t_i$ ako $i$-ty dlh dlžníka $V_i$ . Cieľom modelu rizikových nákladov je odhadnúť riziko spojené s nezaplatením dlhu $V_i$, resp. s rizikom zlyhania v čase $t_i$. Čiastočným výstupom modelu bude odhad proporcionálnej časti dlhov, ktoré nebudú splatené v danom časovom okamihu.

### Definícia zlyhania

V modeli rizikových nákladov si zadefinujeme stav zlyhania nasledovne:

**Definícia 2.3.1** *V čase $T > t_i$ hovoríme, že klient zlyhal v čase $t_i$, ak*

1. *je v čase $T$ v omeškaní s platbou svojho $i$-teho dlhu (resp. $i$-tej splátky) viac ako 90 dní*

2. *veľkosť zlyhaného dlhu sa už nemení do konca sledovaného obdobia (do času $T$)*

$$V_{t_i} = V_{t_{i+1}} = \cdots = V_T$$

*Pozn: v časových okamihoch $t_j$, kde $j \geq i$ považujeme klienta za zlyhaného.*

Prvá podmienka je bežná definícia zlyhania, ktorá sa využíva v modeloch kreditného rizika. Zadefinovaním druhej podmienky sme do definície zlyhania zahrnuli aj mieru vymáhateľnosti (tzv. recovery rate). V praxi si banka bežne necháva preplatiť od nedisplinovaného klienta všetky náklady spojené s omeškanými platbami a ich vymáhaním (formou sankčnej úrokovej sadzby, rôznych poplatkov a pokút). Ak sa klient dostal do omeškania s platbou nad 90 dní, ale povedzme po 150-tich dňoch tento splatil, tak podľa našej definície túto nedisciplinovanosť v našom modeli budeme ignorovať.
Pre lepšiu zrozumiteľnosť si uveďme príklad:
Uvažujme tri úvery na 100 000 korún. Po roku sú dva z nich zlyhané, t.j. v omeškaní nad 90 dní, každý s nezaplatenou časťou 70 000 korún. To znamená, že podľa Bazilej II ročná pravdepodobnosť zlyhania bude rovná

$$PD = 2/3$$

a expozícia vystavená riziku v čase zlyhania je

$$EAD = 3 * 70000 \text{ korún}$$

O rok neskôr, je v stave zlyhania už len jeden úver s nezaplatenou časťou 60 000 korún (časť dlhu sa mu znížila o zaplatené platby istiny v priebehu roka) a zvyšné dva úvery boli splatené podľa zmluvnej dohody. To znamená, že z expozície $2 * 70000$ korún, ktorá sa po roku dostala do stavu zlyhania, sa vrátilo $70000 + (70000 - 60000)$ korún. Potom miera vymáhateľnosti je

$$REC = \frac{70000 + 10000}{2 * 70000} = \frac{8}{14}$$

Strata v prrípade zlyhania je daná

$$LGD = 1 - REC = 1 - 8/14 = 3/7$$

pre očakávané straty podľa vzťahu 1.1 platí

$$EL = PD * LGD * EAD \; = \; \frac{2}{3} * 3 * 70000 * \frac{3}{7} = 60000 \text{ korún}$$

V našom modeli, podľa našej definície uvažujeme len jedného zlyhaného klienta. Preto miera zlyhania je $DR = 1/3$. Množstvo expozície vystavené riziku v čase zlyhania

$$EAD = 3 * 60000 \text{ korún}$$

Potom naše očakávané celkové náklady sú

$$RC = DR * EAD \; = \; \frac{1}{3} * 3 * 60000 = 60000 \text{ korún}$$

Miera straty je citlivá na expozíciu v čase zlyhania (v našej interpretácii EAD je rovná výške dlhu v čase zlyhania). Preto budeme v modeli pozorované zlyhania vážiť veľkosťou dlhu v čase zlyhania.

## Predpoklady modelu

Ako bolo spomenuté už v úvodnej časti práce, model, ktorý vysvetľujeme je limitovaný nasledovnými predpokladmi. O rizikových skupinách, vytvorených v prvej časti, budeme predpokladať, že sú homogénne a jednotlivci v rámci skupiny sú na seba nezávislí.

**Predpoklad 1** *Homogénnosť rizikových tried.*

Vstupom do modelu sú historicky dostupné dáta skladajúce sa z úverov s rôznou životnosťou. Staré úvery možno pozorovať na dlhšom časovom horizonte ako nové. Pri úveroch s vyššou maturitou ako sledované obdobie, nie je možné niektoré mesiace vôbec pozorovať. Nakoľko máme k dispozícii úvery s relatívne krátkou životnosťou, budeme predpokladať, že priemerné odpozorované správanie sa skúmaného portfólia bude očakávaným správaním portfólia až kým úver nedosiahne dobu splatnosti.

**Predpoklad 2** *Riziko zlyhania sa v čase nemení.*

Skúmame splátkové úvery, o ktorých predpokladáme, že sú plne načerpané v čase $t_0$ a následne splácané vo fixne daných časoch $t_i$ vo fixnej výške $P$.

**Predpoklad 3** *Každý úver má v čase schvaľovania určený fixný amortizačný plán. Fixné výšky splátok $P = P_i$ a fixne dané časové momenty $t_i$, v ktorých sa očakáva platba splátky.*

## Dvojstavový Markov proces

Základ pre model rizikových nákladov tvorí simulácia nastatia udalosti zlyhania počas jednotlivých splátkových mesiacoch. Táto simulácia produkuje výstupy vo forme Markovovho reťazca.

**Definícia 2.3.2** *Náhodnú postupnosť $\{X_n\}_{n\in T}$ s množinou stavou $S$ nazývame Markovov reťazec, ak*

- *množina $T = \{0, 1, 2, \cdots\}$*
- *platí Markovova vlastnosť pre vetky$i_0, \cdots, i_{n-1}, i, j, \in S$ :*
  $P(X_{n+1} = j|X_n = i, X_{n-1} = i_{n-1}, \cdots, X_0 = i_0) = P(X_{n+1} = j|X_n = i)$

Zadefinujme si náhodnú premennú (s markovovou vlastnosťou) $X_{t_i} \in \{0, 1\}$, ktorá bude priraďovať $i$-temu dlhu $V_i$ hodnotu 0, ak klient zaplatil svoj $i$-ty dlh; a hodnotu 1, ak je klient zlyhaný a výška dlhu $V_i$ znamená pre banku náklady. Diskrétny Markov reťazec modeluje stavy, do ktorých sa n.p. $X_{t_i}$ dostáva v diskrétnych časoch $t_i$. Celý proces možno popísať jednokrokovými pravdepodobnosťami prechodu medzi stavmi 0 a 1. Označme $p_{ab}(t_i) = P[X_{t_i} = b|X_{t_{i-1}} = a]$ pravdepodobnosť prechodu zo stavu $a$ do stavu $b$ v čase $t_i$. Ak prechodové pravdepodobnosti sú konštantné v čase (predpoklad 2 v našom modeli), hovoríme, že Markov proces má stacionárne pravdepodobnosti prechodu.

$$p_{ab}(t_i) = P[X_{t_i} = b|X_{t_{i-1}} = a] = P[X_{t_{i+\tau}} = b|X_{t_{i+\tau-1}} = a] = p_{ab} \ , \quad \tau \geq 0$$

Z defininície zlyhania máme, že stav 1 je absorpčný stav, t.j. stav 0 nie je dosiahnuteľný zo stavu 1. Potom pre prvky matice prechodu $\mathbf{P} = \{p_{ab}\}_{a,b=0,1}$ platí

$$p_{10} = 0 \quad p_{11} = 1$$

Takto špecifikovaný Markov proces je charakterizovaný jediným parametrom $p$, t.j. pravdepodobnosťou prechodu medzi stavom 0 a 1.

$$p = p_{01} = 1 - p_{00}$$

Matica prechodu má tvar

$$\mathbf{P} = \begin{bmatrix} 1-p & p \\ 0 & 1 \end{bmatrix}$$

Nech $V_{ab}$ je objem všetkých dlhov, ktoré prešli zo stavu a do stavu b. Funkcia maximálnej vierodhodnosti špecifikovaného Markovovho procesu pre napozorované dáta je daná v tvare

$$\ln L(\mathbf{P}|\text{ pri prechode cez 1 periódu}) = V_{00}.\ln(1-p) + V_{01}.\ln p + V_{11}.\ln 1 + V_{10}.\ln 0$$

Maximalizáciou funkcie vzhľadom na parameter $p$ dostávame maximálne vierohodný odhad

$$\hat{p} = \frac{V_{01}}{V_{00} + V_{01}} \tag{2.6}$$

Z Chapmanov-Kolmogorových vieme spočítať iteráciami pravdepodobnosť $p_{ab}(m,n)$ ako pravdepodobnosť prechodu, keď v $m$-tom kroku je systém v stave $a$ a v $n$-tom kroku v stave $b$ nasledovne

$$p_{ab}(m,n) = \sum_c p_{ac}(m,q).p_{cb}(q,n)$$

Z čoho vieme, že matica $[p_{ab}(m, m+n)] = \mathbf{P}^n$ Teda poznáme prechodové pravdepodobnosti po $j$ krokoch, sú nimi prvky $\mathbf{P}(j) = \mathbf{P}^j$. Pomocou takto zaiterovanej pravdepodobnosti prechodu si vyjadríme kumulatívnu distribučnú funkciu prežitia klienta ( klient spláca svoje dlhy podľa zmluvnej dohody) do $j$-tej splátky (vrátane). Je daná pravdepodobnosťou prechodu zo stavu 0 do stavu 0 po $j$ krokoch $p_{00}^{(j)}$ matice $\mathbf{P}^j$.

$$p_{00}^{(j)} \ = \ \prod_{i=1}^{j}(1-\hat{p}) \ = \ (1-\hat{p})^j$$

Potom pravdepodobnostné rozdelenie prežitia do $j$-tej splátky a následného zlyhania na $j$-tej splátke je dané vzťahom:

$$p_{01}^{(j)} \ = \ 1 - p_{00}^{(j)} \ = \ 1 - (1-\hat{p})^j$$

Rizikové náklady na $j$-tej splátke sú determinované pravdepodobnosťou nesplatenia tejto platby.

**Veta 2.3.3** *Rizikové náklady $RC_{t_i}$ na $i$-tej zmluvnej splátke $P_i$ v čase $t_i$ za predpokladu konštantného rizika zlyhania determinovaného vzorcom 2.6 sú dané:*

$$RC_{t_i} = [1 - (1 - \hat{p})^i].P_i \ ; \quad 0 \leq t_i \leq T \ ; \ i = 0, 1, \cdots, n \qquad (2.7)$$

Veľkosť rizikových nákladov je citlivá na presnosť odhadu parametra $\hat{p}$, ktorý bol odhadnutý z historicky dostupných dát tvoriacich tzv. vývojovú vzorku. Predpokladáme, že vývojová vzorka dostatočne reprezentuje správanie sa populácie, ktorá požiada o úver a prejde schvaľovacím procesom. Avšak v praxi je málo pravdepodobné, že požiada o úver znovu tá istá vzorka klientov, ako vzorka, na ktorej bol model vyvinutý. Nabudúce môže požiadať o úver viac klientov s istým špecifickým typom správania sa a iný typ klienta nemusí prísť vôbec. Rizikový menežment v banke potrebuje mať istotu, že príliš nepodhodnotí ani nepredcení riziko klienta a preto je potrebné analyzovať rôzne scenáre portfólia s náhodným zastúpením rôznych typov dlžníkov a tak môcť určiť spoľahlivosť daného odhadu. Uvažujme napríklad dve portfóliá pozostávajúce zo štyroch úverov. Prvé portfólio obsahuje dva úvery, z ktorých jeden bol úplne splatený a druhý nebol splatený vôbec. Očakávané náklady tohoto portfólia sú 50%. Druhé portfólio obsahuje dva úvery, ktoré boli v oboch prípadoch z polovice splatené. Aj pre toto portfólio sú očakávané náklady 50%. Je zrejmé, že pre banku je výhodnejšie zvoliť druhé portfólio, pretože očakáva, že budúci dlžníci zaradení do tohoto portfólia prinesú pre banku nižšie riziko vzhľadom na varianciu očakávaných nákladov. Týmto príkladom vidíme, že samotná informácia o očakávaných nákladoch nedokáže identifikovať rozdiel v rizikovosti týchto dvoch portfólií. Preto je nutné analyzovať varianciu odhadu a interval spoľahlivosti pomocou rôznych simulácii. Najrozšírenejším typom je monte carlo simulácia. Avšak v našom prípade je nepoužiteľná, pretože predpokladá, že pozná presné rozdelenie skúmanej vzorky. Rozdelenie našej vzorky je neznáme, môže byť asymetrické, môže mať ťažké chvosty a pod. V takomto prípade je vhodné použiť Bootstrap metódu. Ako získame bootsrap rozdelenie a ako sa počítajú intervaly spoľahlivosti si vysvetlíme v nasledovnej časti.

## Bootstrap metóda

Po prvýkrát túto metódu predstavil v roku 1979 Elfron ako spôsob pre výpočet štandardnej odchýlky ľubovolného odhadu. Bootstrap metóda simuluje, čo by sa stalo, ak by sme pozorovali opakujúce sa vzorky zo skúmanej populácie a to tak, že skúma opakujúce sa náhodné výbery z dostupných dát. Tieto náhodné výbery môžu mať menší rozmer ako je rozmer dostupných dát a môžu byť vytvorené s vracaním alebo bez vracania. Empirický výskum ukázal, že najlepšie výsledky dosahujeme vtedy, ak náhodné výbery budú robené s vracaním o rovnakej veľkosti ako dostupné dáta. V praxi potrebujeme vedieť odhad štandardnej chyby a výchylky od štatistiky, ktorá nás zaujíma. Túto metódu aplikujeme na náš problém odhadu rizika zlyhania. V predchádzajúcej časti sme si vypočítali odhad $\hat{p}$ z pozorovaného výberu pochádzajúceho z neznámeho rozdelenia F. Označme jeho štandardnú chybu ako $se_F(\hat{p})$ a výchylku ako $bias_F(\hat{p})$. Odhad štandardnej chyby pomocou Bootstrap metódy, ktorý navrhol Elfron je daný:

$$se_B = \left\{ \sum_{b=1}^{B} [\hat{p}^*(b) - \hat{p}^*(\cdot)]^2 / (B-1) \right\}^{1/2} \tag{2.8}$$

kde $\hat{p}^*(b)$ je pôvodný odhad $\hat{p}$ vypočítaný z $b$-teho prevzorkovania ($b = 1, 2, \ldots, B$), $\hat{p}^*(\cdot) = \sum_{b=1}^{B} \frac{\hat{p}^*(b)}{B}$ je stredná hodnota parametra $\hat{p}$ bootstrap rozdelenia a $B$ je celkový počet prevzorkovaní. Zo zákona veľkých čísel máme:

$$\lim_{B \to \infty} \hat{se}_B = \hat{se}_F(\hat{p})$$

Výchylka je daná vzťahom

$$bi\hat{a}s_B = \sum_{b=1}^{B} \frac{\hat{p}^*(b)}{B} - \hat{p} \tag{2.9}$$

Pre lepšiu predstavu bude výhodnejšie, ak namiesto bodového odhadu $\hat{p}$ nájdeme interval, ktorý pokryje skutočnú hodnotu parametra s vopred danou pravdepodobnosťou $(1-\alpha)$. Takýto interval nazývame intervalom spoľahlivosti s koeficientom spoľahlivosti $1-\alpha$ alebo tiež $100(1-\alpha)\%$-ným intervalom spoľahlivosti. Horná hranica intervalu spoľahlivosti bude poukazovať na maximálnu stratu pri danom stupni spoľahlivosti. Intervaly spoľahlivosti vychádzajúce z t-štatistiky, či percentuálne intervaly spoľahlivosti sú pri bootsrap metóde v praxi nepoužiteľné, pretože fungujú len za podmienky normality bootstrap rozdelenia $\hat{G}$ a malej výchylky. Tieto predpoklady nie sú v praxi takmer nikdy splnené. Najrozšírenejšou metódou na výpočet intervalu spoľahlivosti pre bootstrap je BCA interval, ktorý je modifikáciou percentilovej metódy. Prispôsobuje percentily na korekciu výchylky a šikmosti. Korekciu výchylky definujeme ako

$$\hat{z}_0 = \Phi^{-1}\Big(\frac{\#\{\hat{p}^*(b) < \hat{p}\}}{B}\Big)$$

kde $\#\{podmienka\}$ je počet prevzorkovaní, ktoré spĺňajú špecifikovanú podmienku a $\Phi^{-1}(x)$ je inverznou funkciou ku štandartnému normálnemu rozdeleniu.
Korekciu šikmosti robíme tak, že skúmame, ako jednotlivé záznamy vo vývojovej vzorke ovplyvňujú odhad. Počítame odhad tzv. akcelerácie použitím vzorca

$$\hat{a} = \frac{\sum_{i=1}^{n}(\hat{p}(\cdot) - \hat{p}(i))^3}{6\{\sum_{i=1}^{n}(\hat{p}(\cdot) - \hat{p}(i))^2\}^{3/2}} \tag{2.10}$$

kde $\hat{p}(i)$ je odhad vypočítaný na pôvodnej vzorke s vynechaním $i$-tym záznamom, $n$ je počet vzoriek vytvorených vynechaním jedného záznamu (zodpovedá počtu záznamov v pôvodnej vzorke), $\hat{p}(\cdot) = \sum_{i=1}^{n} \frac{\hat{p}}{n}$ je strednou hodnotou z $n$ vytvorených vzoriek
Hranice $100(1-\alpha)\%$-ného intervalu spoľahlivosti sú dané

$$(\hat{p}_l, \hat{p}_u) = (G^{-1}(\alpha), G^{-1}(1-\alpha)) = (\hat{p}^*_B(\alpha), \hat{p}^*_B(1-\alpha))$$

, kde

$$\alpha_1 = \Phi\Big(\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-}1(\alpha)}{1 - \hat{a}(\hat{z}_0 + \Phi^{-}1(\alpha))}\Big) \tag{2.11}$$

$$\alpha_2 = \Phi\Big(\hat{z}_0 + \frac{\hat{z}_0 + \Phi^{-}1(1-\alpha)}{1 - \hat{a}(\hat{z}_0 + \Phi^{-}1(1-\alpha))}\Big) \tag{2.12}$$

Teraz už poznáme štandardnú odchýlku odhadu a zároveň odhad parametra, ktorý vznikne simuláciou príchodov rôznych typov klientov. Súčasne vieme určiť maximálnu možnú strata v rámci určitého intervalu spoľahlivosti a tým dokážeme podať presnú informáciu o spoľahlivosti nášho odhadu. Jediné, čo ostáva nevyriešené, je otázka, akú veľkú rizikovú maržu treba uvažovať v cene úveru, aby vykrývala náklady spôsobené v danej skupine klientov.

### 2.3.2 Efektívna úroková miera a riziková marža

Nech $i$ je zmluvná úroková miera a $i_m$ od nej odvodená úroková sadzba pre jednu pozorovanú periódu, povedzme mesiac. Efektívna úroková miera je z hľadiska veriteľa časť zmluvnej úrokovej miery $i_m$ vyjadrujúca efektívnu mieru návratnosti zapožičaných finančných prostriedkov, pri ktorej sa navráti požičaný objem bez toho, aby boli z obchodu ďalšie zisky. Túto mieru udáva vnútorné výnosové percento (IRR), ktoré diskontuje peňažné toky tak, aby súčasná hodnota obchodu bola nulová. Nech $CF_i$ je objem peňažných tokov, ktorý môže byť odchádzajúceho alebo prichádzajúceho charakteru, v čase $t_i$. Odchádzajúci charakter majú peňažné toky odchádzajúce z pohľadu veriteľa, t.j. v čase poskytovania finančných prostriedkov. Potom peňažné toky prichádzajúceho charakteru sú prichádzajúce splátky od dlžníkov. Potom efektívnu úrokovú mieru, označme $i_{ef}$, vypočítame implicitne použitím vzorca

$$0 = \sum_{t=0}^{n} \frac{CF_t}{(1+i_{ef})^t} \tag{2.13}$$

V prípade nulového rizika zo zlyhania, t.j. klient zaplatí všetky splátky podľa zmluvnej dohody, je výnosové percento z obchodu (resp. efektívna úroková miera) jednoznačne dané úrokovou sadzbou, ktorú si veriteľ stanovil. Pri výskyte zlyhania klienta sa toto percento znižuje vzhľadom na pôvodnú úrokovú mieru $i_m$ . Rozdiel medzi očakávanou efektívnou úrokovou mierou a skutočnou vyjadruje percento tokov, ktoré za periódu strácame.

**Definícia 2.3.4** *Riziková marža (RM) je daná rozdielom medzi zmluvnou úrokou sadzbou $i_m$ a efektívnou úrokovou mierou $i_{ef}$*

$$Rm = i_m - i_{ef} \tag{2.14}$$

Vychádzajúc z druhého predpokladu máme: $CF_0 = -V$. Veriteľ v čase $t_0$ poskytol úver vo výške V, ktorý by mal dlžník splácať v nasledovných časových okamihoch $t_i, i \in \{1, \cdots, n\}$v zmluvne stanovených výškach splátok $P_i$. Vzľadom na mieru rizika a príslušné rizikové náklady na $i$-tej splátke je splátka $P_i$ znížená o tieto náklady podľa vety 2.3.3.
Potom $CF_t = (1-\hat{p})^t.P_t \qquad$ pre $t \in \{1, \cdots, n\}$. Následným dosadením do vzťahu 2.13 dostávame

$$V = \sum_{t=1}^{n} \frac{(1-\hat{p})^t.P_t}{(1+i_{ef})^t} = P.\frac{\lambda_1(\lambda_1^n - 1)}{\lambda_1 - 1} \qquad kde \lambda_1 = \frac{1-\hat{p}}{1+i_{ef}} \tag{2.15}$$

Pre fixný amortizačný plán platí

$$V = \sum_{t=1}^{n} \frac{P_t}{(1+i_m)^t} = P.\frac{\lambda_2(\lambda_2^n - 1)}{\lambda_2 - 1} \qquad kde \lambda_2 = \frac{1}{1+i_m} \tag{2.16}$$

Spojením výrazov 2.15 a 2.16 dostávame

$$\frac{1-\hat{p}}{1+i_{ef}} = \frac{1}{1+i_m}$$

Po úprave a dosadení do vzťahu 2.14 máme

$$RM_m = \hat{p}(1+i_m) \tag{2.17}$$

Týmto sa dostávame k konečnému výpočtu rizikovej marže, ktorý je sformulovaný v nasledujúcej vete.

**Veta 2.3.5** *Za predpokladu konštantnej miery zlyhania a fixného amortizačného plánu pre ročnú rizikovú maržu platí:*

$$RM \approx \hat{p}(12+i) \tag{2.18}$$

*kde $\hat{p}$ je determinovaná vzťahom 2.6*

**Dôkaz:**
Vzťah medzi mesačnou úrokovou sadzbou a ročnou si môžme vyjadriť nasledovne

$$i = (1+i_m)^{12} - 1$$

Využitím tayloroveho vzorca rad $(1+x)^{12}$ rozvinieme do nasledovného tvaru vzhľadom na premennú x v bode 0:

$$(1+x)^{12} = 1 + \binom{12}{1}x + \binom{12}{2}x^2 + \binom{12}{3}x^3 + \cdots$$

Lineárnou aproximáciou dostávame:

$$(1+x)^{12} - 1 \approx 12x$$

Riziková marža je časť z úrokovej miery a preto aj pre ročnú rizikovú maržu platí

$$RC = (1+RC_m)^{12} - 1$$

Využitím Taylorovej aproximácie v lineárnom člene aj pre tento prípad dostávame

$$(1 + RC_m)^{12} - 1 \approx 12.RC_m$$

Dosadením dostaneme vzťah pre výpočet ročnej rizikovej marže

$$RC = 12.RC_m \approx 12(\hat{p}(1 + i_m)) \approx 12(\hat{p}(1 + \frac{i}{12}) = \hat{p}(12 + i)$$

Aplikovaním uvedeného postupu na výpočet rizikovej marže pre jednotlivé rizikovo homogénne skupiny klientov dostávame rizikovo založený stupňovitý spôsob cenotvorby - risk based pricing.

# Kapitola 3

# Záver

V práci sme stručne popísali možnosti konštrukcie štatistického modelu na počítanie časti ceny úveru - rizikovej marže. Jej správne určenie je pre banku ako poskytovateľa úveru kľúčové, pretože nadhodnotenie, ale aj podhodnotenie má negatívny dopad na celkový zisk z úveru. Výpočtu rizikovej marže predchádza vytvorenie skóringovej funkcie, ktorá prideľuje klientom žiadajúcim o úver hodnotu skóre, ktorá by mala vyjadrovať ich spôsobilosť a ochotu poskytnutý úver splatiť. Na základe skóringovej funkcie sú klienti ďalej rozdelení do rizikovo homogénnych skupín (tried) a pre tieto triedy sú následne vypočítané rizikové náklady. Z rizikových nákladov je potom vypočítaná samotná riziková marža.

Navrhovaná metóda je bežne využívaná v praxi kvôli viacerým vlastnostiam. Model priradenia hodnoty skóre jednotlivým klientom pomocou logistickej regresie zároveň vypočítava aj tzv. PD - pravdepodobnosti zlyhania, ktoré sú ďalej použiteľné pri ďalších výpočtoch. Izotónna regresia, ktorá je použitá na zatrieďovanie klientov do rizikovo homogénnych skupín nedáva jednoznačný výsledok ako majú tieto skupiny vyzerať, ale necháva priestor pre banku na vytvorenie interných prístupov, ktorá zadefinovaním týchto skupín sama rozdelí klientov do vhodných tried podľa výsledku izotónnej regresie. Ďalšou a veľmi dôležitou výhodou modelu je jeho relatívne jednoduchá aplikovateľnosť v praxi. Nemá veľkú výpočtovú náročnosť a v pomerne krátkom čase je možné v rámci interných zdrojov banky vyvinúť metódu s dostatočnou diskriminačnou silou, ktorá pomáha rozlíšiť rizikových klientov od menej rizikových a následne podľa toho stanoviť rizikové marže. Tieto budú rozdielne pre jednotlivé rizikovo homogénne skupiny.

Popísaný model má však aj svoje nedostatky. Je založený na predpokladoch, ktoré sú síce zlučiteľné s praxou, ale v niektorých prípadoch môžu byť príliš zjednodušujúce. Na výpočet rizikovej marže je použitý predpoklad o konštantnosti rizika zlyhania na každej splátke, jeho prípadné nedodržanie môže viesť ku nepresným odhadom. Navyše klasifikačný problém je založený na ďalších silných predpokladoch o vstupných dátach. Tieto predpoklady by mohli byť oslabené využitím v druhej kapitole spomínaných nelineárnych metód. V rozšírení tejto práce by sme sa mali zamerať na porovnanie týchto sofistikovanejších metód s navrhovaným modelom, či už na reálnych alebo vygenerovaných dátach a ukázať, či je možné tieto metódy rovnako efektívne aplikovať v bankovej praxi.

# Literatúra


[1] Thielbar M, Patetta M, Marovich P. (2005): *Introduction to ANOVA, Regression, and Logistic Regression*

[2] Hosmer W. D, Lemeshow S. - $2^{nd}$ ed. (2005): *Applied Logistic Regression*

[3] Falkenstein, E. (2002): *Credit scoring for corporate debt, in Credit Ratings: Methodologies, Rationale and Default Risk*, chapter 8, pp. 169-188

[4] Basel Comittee on Banking Supervision: *Studies on the Validation of Internal Rating Systems*, Working Paper No. 14 (May 2005)

[5] Moore D. S, McCabe G. P - $5^{th}$ ed.: *Introduction to the Practice of Statistics*, Chapter 14 - *http : //bcs.whfreeman.com/ips5e/content/cat_080/pdf/moore14.pdf*

[6] Lessons, : *Application of the bootstrap*, Available at *http : //www.meb.ki.se/ ~ aleplo/Compstat/CS05_7.pdf*

[7] Schebes K. B, Stecking R: *Support Vector Machines for Classifying and Describing Credit Applicants: Detecting Typical and Critical Regions*

[8] Kiss F (2003): *Credit Scoring Processes from a Knowledge Management Perspective*

# Kapitola 4

# Príloha