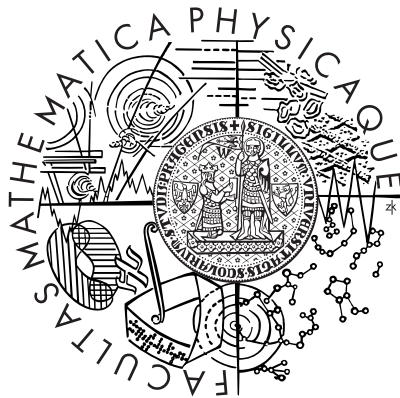


Univerzita Karlova v Praze  
Matematicko-fyzikální fakulta

## DIPLOMOVÁ PRÁCE



Monika Šťástková

### **Kreditní riziko**

Katedra pravděpodobnosti a matematické statistiky

Vedoucí diplomové práce: Doc. RNDr. Jan Hurt, CSc.

Studijní program: Matematika

Studijní obor: Finanční a pojistná matematika

2009

Děkuji doc. Janu Hurtovi za přínosné náměty a připomínky při psaní diplomové práce a své rodině a přátelům za podporu při studiu.

Prohlašuji, že jsem svou diplomovou práci napsala samostatně a výhradně s použitím citovaných pramenů. Souhlasím se zapůjčováním práce.

V Praze dne 17.4.2009

Monika Štátková

# Obsah

Úvod	5
<b>1 Rizika ve financích</b>	<b>7</b>
1.1 Úvěrové riziko . . . . .	10
1.2 Regulace finančních rizik . . . . .	11
<b>2 Modelování úvěrového rizika</b>	<b>13</b>
2.1 Modely pro řízení úvěrového rizika . . . . .	16
2.1.1 Modely založené na úvěrové migraci . . . . .	16
<b>3 Časová struktura pravděpodobnosti selhání</b>	<b>19</b>
3.1 Funkce spolehlivosti a intenzita poruch . . . . .	21
3.2 Přímé odvození víceleté pravděpodobnosti selhání . . . . .	22
3.3 Úvěrová migrace . . . . .	24
3.3.1 Statistická analýza v Markovových řetězcích . . . . .	28
3.4 Časová struktura založená na úvěrovém rozpětí . . . . .	33
3.4.1 Úvěrová křivka . . . . .	35
<b>4 Modelování úvěrového rizika založené na analýze dat o přežití</b>	<b>38</b>
4.1 Náhodné cenzorování a intervaly spolehlivosti . . . . .	39
4.1.1 Metoda maximální věrohodnosti . . . . .	39
4.1.2 Vlastnosti maximálně věrohodných odhadů . . . . .	42
4.2 Model pro dobu do selhání . . . . .	43
4.3 Model s vysvětlujícími proměnnými . . . . .	51
<b>5 Aplikace na reálná data</b>	<b>54</b>
5.1 Statistická analýza v Markovových řetězcích . . . . .	54
5.1.1 $\chi^2$ -test dobré shody . . . . .	55
5.1.2 Test poměrem věrohodností . . . . .	56
5.2 Modelování úvěrového rizika pomocí analýzy přežití . . . . .	62
<b>Závěr</b>	<b>67</b>
<b>Příloha</b>	<b>68</b>
<b>Literatura</b>	<b>77</b>

Název práce: Kreditní riziko  
Autor: Monika Šťástková  
Katedra pravděpodobnosti a matematické statistiky  
Vedoucí diplomové práce: Doc. RNDr. Jan Hurt, CSc.  
e-mail vedoucího: jan.hurt@mff.cuni.cz

Abstrakt: V předložené práci se zabýváme modelováním úvěrového rizika, konkrétně pravděpodobnosti selhání a doby do selhání. Nejdříve prezentujeme dvě v praxi často používané metody pro určení časové struktury pravděpodobnosti selhání. První z nich vychází z úvěrové migrace. Předpokládá, že přechody mezi kategoriemi ohodnocení lze popsat homogenním Markovovým řetězcem. K testování hypotézy, zda je předpoklad homogenity v praxi splněn, jsou v práci uvedeny dvě podoby testové statistiky. Druhá metoda využívá myšlenku, že určení úvěrové křivky je podobné jako vyhlazování křivky výnosové. Proto k jejímu stanovení využívá tvar Nelson-Siegelovy funkce. Dále se práce věnuje modelu pro dobu do selhání. K určení jejího rozdělení využívá teorii náhodného cenzorování. Některé z uvedených postupů na závěr ilustrujeme na reálných datech.

Klíčová slova: Úvěrové riziko, Úvěrová migrace, Úvěrová křivka, Náhodné cenzorování

Title: Credit risk  
Author: Monika Šťástková  
Department of Probability and Mathematical Statistics  
Supervisor: Doc. RNDr. Jan Hurt, CSc.  
Supervisor's e-mail address: jan.hurt@mff.cuni.cz

Abstract: This thesis is concerned in credit risk modelling, especially the default probability and time to default variable. It deals with two commonly used methods to figure out the term structure of default probability. The first one is based on credit migration. It is assumed that the credit migration process follows a time homogeneous Markov chain. To test the assumption of time homogeneity two test statistics are proposed in the thesis. The later method for establishing the term structure of the default probability uses the fact, that estimation of credit curve is analogous to a construction of the yield curves. So Nelson-Siegel function can be used for this aim. Then the model for the time to default is described. It is based on the theory of random censorship. Some of these methods are illustrated on real data.

Keywords: Credit risk, Credit migration, Credit curve, Random censorship

# Úvod

Úvěrové riziko můžeme jednoduše vyjádřit jako riziko, že dlužník v budoucnosti nevyplácí své dohodnuté závazky. Věřiteli tak vznikne ztráta. Objevuje se jako součást většiny finančních transakcí a v případě, že účastníci trhu chtějí na základě svých obchodů realizovat zisk, nemohou se mu zcela vyhnout. Jako v každém jiném případě účastníci trhu požadují za přebírání rizika určitou kompenzaci, která pro ně v případě, že nedojde k selhání protistrany, znamená zisk. Nabízí se tedy prostor pro teorie a modely, které se snaží co nejpřesněji úvěrové riziko odhadnout, modelovat a tak řídit.

Nejčastěji jsou úvěrovému riziku vystaveny banky. Právě pro ně je modelování úvěrového rizika podstatnou součástí jejich úspěšného působení na trhu. Význam modelování úvěrového rizika potvrdily i nedávné události na finančních trzích, kde jeho podhodnocení a nerozpoznání vedlo k velkým neočekávaným ztrátám. Banky proto kladou velký důraz na vývoj metod, které informace dostupné na trhu a poskytnuté potenciálním klientem převedou do matematické podoby. Umožní jim co nejpřesněji určit úvěrové riziko, které je s tímto klientem spojeno. Právě z důvodu možného vzniku velkých ztrát v bankovním sektoru se v minulém století začaly objevovat myšlenky vedoucí k regulaci bankovního systému. Bylo to hlavně z toho důvodu, aby banky nechtěly přenášet své ztráty na klienty. Nejen na popud regulatorních požadavků vzniklo a dále vzniká množství různých přístupů k modelování jednotlivých veličin, které úvěrové riziko popisují. Nejčastěji jsou to modely pro dobu do selhání a pravděpodobnost selhání.

V této práci nejdříve popíšeme jednotlivé složky finančních rizik, přičemž podrobněji se budeme věnovat riziku úvěrovému a veličinám, které ho nejvíce ovlivňují. V rámci této části se krátce zmíníme i o ratingových agenturách, které hrají pro modelování úvěrového rizika v některých přístupech významnou roli. Dále se pak budeme věnovat časové struktuře pravděpodobnosti selhání, jakožto základní veličiny, pomocí níž lze úvěrové riziko modelovat. Pravděpodobnosti selhání jsou dostupné na trhu, ovšem pouze v určitých časových okamžicích pro určitý časový horizont. Pokud chceme získat jejich hodnoty i pro jiné časové intervaly, musíme využít některý z matematických přístupů. V této práci se věnujeme třem metodám. První z nich je poměrně snadná, avšak výsledky, které poskytuje, neodpovídají příliš reálnému očekávání. Druhý přístup vychází z úvěrové migrace. Tato metoda je často používána a lze pomocí ní získat poměrně snadno celou časovou strukturu pravděpodobnosti selhání.

Avšak aby to bylo možné, je nutné předpokládat, že přechody mezi kategoriemi ohodnocení lze popsat homogenním markovským řetězcem. Tento předpoklad by v reálu nemusel být často splněn. Z tohoto důvodu se budeme věnovat metodám, pomocí nichž je možné platnost předpokladu ověřit. Na závěr práce tento postup demonstrujeme na reálných datech. Třetí metoda pro stanovení časové struktury pravděpodobnosti selhání využívá postup podobný konstrukci výnosových křivek. Úvěrová křivka zobrazující časovou strukturu pravděpodobnosti selhání je určena pomocí Nelson-Siegelovy funkce. Tato metoda se používá nejčastěji pro modelování úvěrového rizika úvěrových derivátů.

V další kapitole se budeme zabývat modelem pro dobu do selhání, který vychází z analýzy přežití. Při modelování úvěrového rizika jsme postaveni před problémem spojený s tím, že úvěrové smlouvy se obvykle váží k dlouhým časovým intervalům. Často tedy potřebujeme provádět analýzu, aniž by úvěrové smlouvy už vypršely nebo klienti podleli nesplácení. Možnost, jak tuto situaci řešit, nabízí náhodné cenzorování. Parametry popisující rozdělení doby do selhání odhadneme pomocí metody maximální věrohodnosti a vyjádříme pro ně intervaly spolehlivosti. Na závěr práce jsme ukázali použití tohoto modelu na reálných datech.

# Kapitola 1

## Rizika ve financích

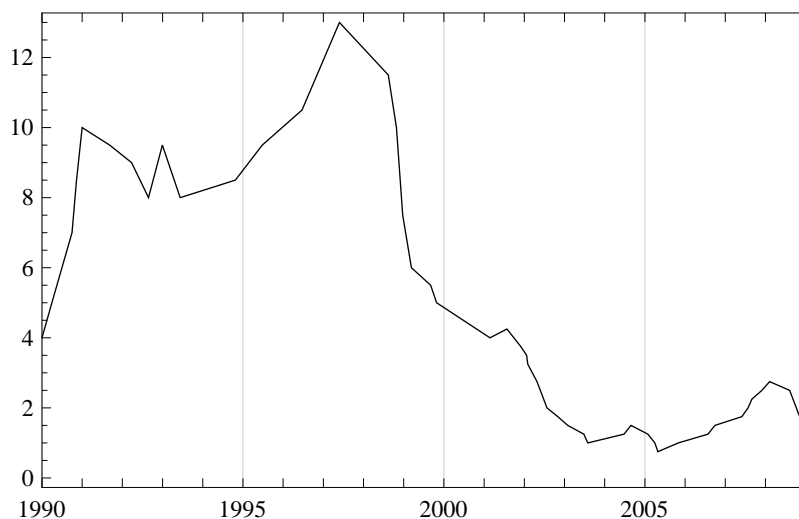
Riziko můžeme chápat jako nebezpečí nebo také vysokou míru pravděpodobnosti nezdaru a zejména ve finanční oblasti jako riziko ztráty spojené s měnící se hodnotou držených aktiv a pasiv. Jedná se přitom o ztrátu neočekávanou, kterou může společnost, nejčastěji banka, utrpět v souvislosti s vývojem situace na finančním trhu. Právě fakt, že se jedná o neočekávanou ztrátu a její možný dopad na klienty bank, byl důvodem zavedení regulátorů a regulačních požadavků, kterým se budeme více věnovat později.

Riziko samozřejmě nemusí být spojeno jen s finančním prostředím. Zde se však budeme věnovat právě rizikům finančním a uvedeme jejich základní dělení. Existuje více možností, jak jednotlivé skupiny finančních rizik dále dělit. V této práci budeme vycházet ze struktury, která je uvedena např. v [4]. Důvodem k tomuto členění finančních rizik je skutečnost, že jednotlivé typy rizik se od sebe v mnohém liší a tak vyžadují i odlišný přístup k řízení. Zvláště patrné jsou rozdíly mezi dvěma pro bankovní sektor nejvýznamnějšími finančními riziky - rizikem tržním a úvěrovým. Z povahy každého z nich je zřejmé, že každé vyžaduje modelování jiných parametrů. Další odlišnosti spočívají v různé délce časových horizontů, pro které se modeluje jejich vývoj. Tržní riziko se váže spíše ke krátkodobým transakcím, a tak si žádá modelace hlavně pro kratší časové intervaly. Riziko úvěrové se zase obvykle váže k dlouhým časovým horizontům v řádu desítek let a vyžaduje tak modelování pro dlouhé časové intervaly. V neposlední řadě hraje významnou roli i to, že v případě modelování tržního rizika máme k dispozici mnohem více údajů o selháních než u rizika úvěrového.

**Tržní riziko** Jak název napovídá, *tržní riziko* souvisí s vývojem finančních trhů a tak i tržních cen aktiv a pasiv. Projevuje se právě díky jejich změnám, které mohou vést k menším či větším ztrátám. Především v oblasti hodnocení tržního rizika se využívá známé *hodnoty v riziku*, neboli *VaR*. Tržní riziko můžeme ještě dále podrobněji rozdělit na podskupiny (viz. [4]), které teď krátce popíšeme.

- (i) *Úrokové riziko* představuje hlavně nebezpečí ztráty plynoucí z instrumentů citlivých na změny úrokových měr v důsledku jejich negativního po-

hybu. Kromě toho úrokové riziko zahrnuje i riziko změny tvaru výnosové křivky, riziko změn vztahů mezi různými úrokovými indexy aj. Vývoj diskontní sazby vyhlášené Českou národní bankou od 1.1.1990 do 6.2.2009 ukazuje následující graf (viz. [5]). Úrokové sazby komerčních bank se odvíjejí od sazeb vyhlášených Českou národní bankou, protože banka bude půjčovat peníze svým klientům jedině v případě, že dosáhne takového zisku, aby dostála svým závazkům vůči České národní bance. Tedy jakýkoli pohyb sazeb u banky státní má za následek reakci u bank komerčních. Z grafu je patrné, že sazby vykazují poměrně velkou volatilitu i v krátkých časových intervalech. Graf diskontní sazby tak podporuje využívání teorií zahrnujících volatilitu úrokových sazeb k eliminaci úrokového rizika.



Obr. 1.1: Vývoj diskontní sazby vyhlášené ČNB v čase (v %)

- (ii) *Měnové riziko* spočívá ve změnách spotových měnových kurzů a jejich volatilitě, což může způsobit ztrátu na instrumentech, které jsou na změny měnových kurzů citlivé.
- (iii) *Akciové riziko* je způsobeno změnami cen akcií a vztahů mezi akciovými indexy. Tyto změny pak mohou vést ke ztrátám na nástrojích na tyto změny citlivé.
- (iv) *Komoditní riziko* je především riziko ztráty na instrumentech, jež ovlivní změny cen komodit jako jsou ropa, zlato, zemní plyn atd.
- (v) *Riziko úvěrového rozpětí* je riziko ztráty způsobené změnou úvěrového rozpětí u cenných papírů různého úvěrového hodnocení. *Úvěrové rozpětí* neboli *kreditní spread* vyjadřuje rozdíl mezi hodnotou méně rizikového



dluhopisu (obvykle se uvažuje státní dluhopis, protože jej lze považovat prakticky za bezrizikový) a rizikovějšího dluhopisu. Podrobněji se o významu a využití úvěrového rozpětí zmíníme v jedné z následujících kapitol.

- (vi) *Korelační riziko* vzniká jako důsledek změny uvažovaných historických korelací mezi rizikovými kategoriemi, které daná společnost považuje pro ni za významné.

**Úvěrové riziko** *Úvěrové riziko* vyjadřuje nebezpečí vzniku ztráty jako důsledek neschopnosti protistrany splatit své závazky. Více se mu budeme věnovat v kapitole 1.1.

**Likvidní riziko** *Likvidnímu riziku* společnost podlehne tehdy, nemá-li dostatek volných hotových peněžních prostředků v okamžiku, kdy je potřebuje, aby dostala svým závazkům. Likvidní riziko může mít dvě podoby (viz. [4]).

- (i) *Riziko tržní likvidity* vzniká jako důsledek podmínek na trhu. Pokud situace na trhu neumožňuje společnosti dostatečně rychle, v dostatečném množství a za příznivou cenu vypořádat finanční pozice, do kterých vstoupila, nemá možnost získat hotové peníze tak, jak předpokládala. Riziko tržní likvidity je pak riziko, že dojde v takové situaci ke ztrátě. Často se totiž stane, že společnost uplatní svá aktiva za nevýhodných podmínek, aby zabránila podlehnutí likvidnímu riziku.
- (ii) *Riziko financování* spočívá v nebezpečí, že společnost bude v momentální platební neschopnosti v důsledku vzniklého nesouladu ve finančních tocích. Jinými slovy společnost není schopná získat hotovost, aby dostala svým závazkům v okamžicích jejich splatnosti s danou úrokovou mírou.

**Operační riziko** *Operační riziko* se objevilo až v poslední době v souvislosti s vývojem a nárůstem využívání interních operačních systémů. V rámci regulatorních požadavků se poprvé o operačním riziku začalo hovořit až v rámci metodiky *Basel II*, o níž se zmíníme v kapitole 1.2. Operační riziko tedy představuje pro společnost nebezpečí vzniku ztráty v důsledku chyb interních operačních systémů a případně chyb lidí, kteří s nimi pracují. Stejně jako předchozí rizika i riziko operační může mít několik podob ([4]).

- (i) *Transakční riziko* je riziko transakčních chyb a kvůli nim vzniklých ztrát. Transakční omyly mohou vzniknout nejčastěji jako chyby způsobené lidským faktorem jak neúmyslně tak i úmyslně, např. podvody, dále jako účetní chyby, chybné vypořádání obchodů aj.
- (ii) *Riziko operačního řízení* vzniká díky chybám v řízení, tedy managementu. Tyto chyby by bylo možné aspoň zčásti odstranit zavedením a dodržováním sofistikovanějšího systému kontroly. Často se stává, že jsou

uzavřeny obchody, které neprošly schvalovacím procesem a překračují povolené limity.

- (iii) *Riziko systému* spočívá v nebezpečí vzniku ztrát zaviněných chybami operačních systémů. Tyto chyby mohou nastat nesprávným sestavením matematických modelů či jejich chybným zavedením do počítačových systémů, špatným odhadem parametrů, nahodilými výpadky počítačové sítě nebo chybami v přenosech dat apod.

## 1.1 Úvěrové riziko

*Úvěrové riziko* je způsobeno reálným nebezpečím vzniku ztráty kvůli náhlé neschopnosti protistrany dostát svým závazkům. Nejčastějším příkladem je situace, kdy dlužník přestane věřiteli splácet úvěr. Dále pak úvěrové riziko nastává u dluhopisového kontraktu v případě, kdy emitent selže a nesplatí tak jeho nominální hodnotu. U opčních smluv je pak spojeno s nebezpečím vzniku situace, kdy nedojde k odkoupení resp. dodání podkladového aktiva. Velký nárůst obchodů s deriváty v poslední době vede k nárůstu potřeby najít cesty, jak se úvěrovému riziku účinně bránit a tedy jak ho řídit. Prvním krokem k tomu je rozlišení jednotlivých druhů úvěrového rizika a následná analýza, které z nich jsou pro společnost reálnou hrozbou. Jednotlivé typy úvěrového rizika nyní shrneme.

- (i) *Přímé úvěrové riziko* je základní kategorií úvěrového rizika. Jedná se o klasické riziko ztráty způsobené selháním protistrany v případě úvěrů, půjček, dluhopisů aj. Z hlediska bankovníctví bylo a stále zůstává nejvýznamnějším typem finančního rizika. Banky tak kladou největší důraz na hledání způsobů jak řídit právě toto riziko.
- (ii) *Riziko úvěrových ekvivalentů* se týká tzv. *podrozvahových položek*, jako jsou poskytované záruky nebo otevřené úvěry na nákup zboží, za které se ručí právě tímto zbožím. Ztráta se v tomto případě realizuje, pokud protistrana selže před splněním svých závazků.
- (iii) *Riziko změny úvěrového hodnocení* je dáno díky častému využívání úvěrového hodnocení externích ratingových agentur. Podrobněji způsob takového hodnocení popíšeme v kapitole 2. Společnosti vstupují do rizikových kontraktů na základě zveřejněného ohodnocení externí ratingové agentury. Pak se může stát, že v důsledku poklesu tohoto ohodnocení pro společnost bude obtížnější za přijatelných podmínek získat financování svých aktivit, tedy uzavřít úvěrový kontrakt.
- (iv) *Vypořádací riziko* je už v dnešní době díky novým vypořádacím mechanismům víceméně eliminované. Týká se ztrát vzniklých během vypořádání transakce, kdy jedna strana svou smluvně dohodnutou hodnotu už

splatila, ale hodnota od protistrany ještě nebyla dodána. Příčinou tohoto prodlení a případně i nesplacení závazků mohou být jak technické problémy, tak neochota protistrany dostát svým závazkům. Nejčastěji provází obchody s cennými papíry a měnové operace.

- (v) *Riziko úvěrové angažovanosti* společnost podstupuje, pokud se dostane do úvěrových vztahů jen s úzkou skupinou partnerů, států, ekonomických sektorů apod. Poměrně snadno a účinně lze toto riziko eliminovat stanovením úvěrových limitů pro jednotlivé protistrany či skupiny případných partnerů s podobnými parametry a jejich následným dodržováním.

Parametry specifické pro popis právě úvěrového rizika uvedeme v následující kapitole, ve které se budeme podrobněji věnovat způsobům modelování úvěrového rizika. Nyní se krátce zmíníme o způsobech regulace finančních rizik v bankovníctví (podrobněji např. viz. [10]).

## 1.2 Regulace finančních rizik

V bankovníctví a investování se nelze zcela vyhnout finančním rizikům a s nimi spojenými neočekávanými ztrátami. Aby finanční instituce neinklinovaly k přenesení těchto případných neočekávaných ztrát na své klienty, funguje ve všech vyspělých státech proces regulace finančních rizik. V jeho rámci regulátor určuje způsob měření finančních rizik a limity na tato rizika za účelem ochrany uživatelů finančních služeb. Nejvýznamnější aktivitou regulátora a pravděpodobně také nejzásadnější je předepsání minimálních kapitálových požadavků. Každá instituce, jíž se tato nařízení týkají, musí vytvořit z vlastních zdrojů a držet po celou dobu svého působení předepsaný minimální kapitál. Tento tzv. *regulační kapitál* musí mít určitou předepsanou strukturu, aby finanční společnosti byly schopné v každém okamžiku svého působení pokrýt ztrátu utrpěnou v souvislosti se svými rizikovými aktivitami.

První myšlenky regulace ve finančním sektoru sahají až do osmnáctého století. Tehdy bylo však nutné spoléhat se především na svůj vlastní přístup k regulaci případně na lokální pravidla, která byla v té době v platnosti. A tak se většina myšlenek vedoucích k dnešní podobě regulatorního řízení finančních rizik objevila až ve dvacátém století.

V bankovním sektoru bylo jedním z hlavních mezníků procesu zavádění regulace založení mezinárodního výboru pro dohled v bankovníctví s názvem *Basel Committee of Banking Supervision*. Tento výbor se poprvé sešel ke konci roku 1974 na popud guvernérů centrálních bank států *G-10*. Skupinu *G-10* tvoří paradoxně jedenáct průmyslových států, které spolu konzultují a spolupracují v otázkách ekonomických i finančních. Pravomoci výboru nejsou právního rozsahu, nemají tedy pro nikoho platnost zákonů. Jedná se spíše o návrhy obecných standardů pro dohled a o doporučení, jak nejlépe tyto návrhy implementovat do národních zákonů.

V roce 1988 výbor vydal dohodu nazvanou zkráceně *Basel I*, která znamenala první významný krok ke stanovení minimálních kapitálových požadavků. Tato dohoda se nejvíce zaměřila na riziko úvěrové, neboť právě to představuje významný zdroj potenciálních ztrát v bankovní sféře. Ačkoli byla dohoda *Basel I* průlomem pro regulaci v bankovníctví, v mnoha ohledech byla nedokonalá. Nejvíce je to patrné právě v oblasti úvěrového rizika. Například pro portfolia bank neklade prakticky žádné požadavky na diverzifikaci a stejně tak i požadavky na minimální hladinu kapitálu jsou stanoveny příliš obecně, aniž by zohlednily skutečné riziko, které konkrétní banka podstupuje. A tak si vyžádala rozšíření známé zkráceně jako *Basel II*. Ovšem ještě do okamžiku vydání nových regulatorních doporučení došlo k velice významné události v oblasti řízení finančních rizik. V roce 1993 se poprvé ve zprávě skupiny složené z odborníků bankovního sektoru veřejného i státního a vědeckých pracovníků států skupiny třiceti předních průmyslových států světa známé jako *G-30* objevila systematická zmínka o tzv. *podrozvahových produktech* a finančních derivátech. Současně s touto zprávou se objevila i potřeba řízení rizik spojených s těmito novými produkty. A právě na základě požadavku vytvoření snadného nástroje k vyjádření tržního rizika, které společnost podstupuje, se poprvé objevuje *hodnota v riziku*, nebo také zkrácením anglického *Value-at-Risk* rozšířený název *VaR*.

V roce 2001 byly tedy zahájeny konzultace vedoucí ke vzniku nových doporučení, jejichž hlavním tématem se stala oblast řízení úvěrového rizika. Výsledkem se potom v roce 2004 stalo zveřejnění metodiky *Basel II*. Snaha o prosazení citlivějších metod ke stanovení rizikovosti portfolií bank, takových, které by zohlednily situaci a zkušenosti každé banky, vedla k prosazení možnosti využívat interní modely k ohodnocení protistran. Banky se tak nemusí spoléhat pouze na vnější ratingová ohodnocení od nezávislých agentur, ale mohou také využít vlastní úvěrová ohodnocení. S tím souvisí i podrobné zpracování požadavků na minimální regulační kapitál, tzv. *kapitálová přiměřenost*, a na jeho strukturu. *Basel II* nadnesl ještě další důležité téma. Poprvé se totiž objevila nová kategorie rizika, *riziko operační*, a začalo se hovořit o možnostech, jak se s ním vypořádat. Regulatorní orgány zejména požadují po bankách, aby zvýšily obezřetnost svých kontrolních mechanismů.

Podobně je řešena i regulace v oblasti pojišťovnictví, kde by analogií metodice *Basel II* měla být *Solvency II*, navazující na *Solvency I*, která je v platnosti od roku 2004.

## Kapitola 2

# Modelování úvěrového rizika

Jak jsme již nastínili v předchozí kapitole, úvěrové riziko je spojené s portfoliem každé finanční společnosti. To jen potvrzuje, že modelování úvěrového rizika je velice důležitou součástí úspěšných obchodních strategií bank a dalších finančních institucí, a že se oprávněně stalo předmětem regulatorních aktivit. Společnosti se však musí potýkat s řadou překážek při modelování úvěrového rizika. Největší potíže jsou dané tím, že na trhu je k dispozici jen malé množství dat, ze kterých by banky mohly vycházet. Vede to pak k nerovnováze v informovanosti - potenciální klient má k dispozici mnohem přesnější a podrobnější informace o své situaci a očekávaném budoucím vývoji než banka. Navíc nejkratší časový horizont, který se v problematice modelování úvěrového rizika obvykle objevuje, je jeden rok. Proto je velice obtížné sladit modely s aktuální situací na trhu.

Další vlastnost, která modelování nijak neusnadňuje, je šikmost rozdělení ztrát. Typické úvěrové portfolio bude v dlouhém časovém intervalu produkovat velké ztráty vzhledem k relativně malým výnosům. Z toho důvodu jsou požadavky kladené na kapitál držený k pokrytí případných ztrát poměrně velké.

Navíc je nutné uvažovat, jak může selhání jedné protistrany ovlivnit selhání jiných dlužníků. Předně je rozumné vzít v úvahu jev, kdy v případě že podlehnou selhání jedna společnost na základě nepříznivého ekonomického vývoje, se mohou do stavu nesplácení dostat i jiné společnosti citlivé na stejné faktory. Navíc závislost mezi selháními dostaneme i díky četným vazbám typu věřitel - dlužník mezi různými společnostmi. A tak selhání jedné může následně způsobit platební neschopnost i další atd. Vraťme se ale nyní k základním parametrům vyžadujícím modelování při řízení úvěrového rizika, které jsou uvedeny např. v [4].

**Pravděpodobnost úvěrového selhání** Název této veličiny zcela vystihuje informaci, která je v ní obsažena. *Pravděpodobnost úvěrového selhání* je pravděpodobnost, že nastane situace, kdy protistrana v úvěrovém kontraktu nebude schopna splatit své závazky. Jedná se o základní veličinu, kterou musí

banky určit, aby mohly smysluplně řídit úvěrové riziko. U každého klienta jsou tak postaveny před úkol, jak co nejpřesněji určit jeho pravděpodobnost selhání. Nejčastěji vychází z dat, která mají k dispozici na trhu a která jim klient sám poskytne. Vychází tak z finančních výkazů, analyzují, zda u klienta převažují dlouhodobé nebo krátkodobé závazky, snaží se co nejpřesněji odhadnout budoucí vývoj peněžních toků, kapitálové struktury, likvidity klienta a také sledují kvalitu řízení společnosti, historické působení společnosti na finančních trzích apod. V neposlední řadě využívají i externí úvěrové hodnocení, je-li k dispozici. Nakonec je klientovi na základě vnitřního i vnějšího hodnocení v bance přiřazen určitý *rating* neboli ohodnocení vyjadřující jeho schopnost dostát svým závazkům.

Nyní se krátce zmíníme o externím úvěrovém hodnocení. *Externí úvěrové hodnocení* poskytují nezávislé agentury. Mezi nejznámější patří celosvětové Standard & Poor's, Moody's, Fitch aj. Jednoduše řečeno ratingové agentury usnadňují v mnohém bankám práci, neboť zpracovávají analýzy založené na finančních výkazech, chování a struktuře vybraných společností. Když mají banky k dispozici vysoké externí ratingové ohodnocení od některé z těchto renomovaných agentur, nepodstupují často tak obsáhlý interní proces vedoucí k ohodnocení klienta. Z historie je totiž patrné, že externí ratingové ohodnocení velice spolehlivě vyjadřuje odhad budoucího vývoje dané společnosti, co se týče schopnosti splácet své závazky. Navíc pro společnosti s vysokým ratingovým ohodnocením od externích agentur je pravděpodobnost selhání malá, takže takové společnosti podléhají nesplácení jen ve zlomku případů. Často tak je pro banku v průměru méně nákladné uhradit tuto prakticky nepravděpodobnou ztrátu než vynaložit prostředky na své vlastní ohodnocení společnosti s vysokým externím ratingem. Každá z těchto agentur tedy publikuje pro hodnocené společnosti mj. jejich zařazení do určité kategorie hodnocení, neboli ratingu. Takové kategorie například u agentury Standard & Poor's mají následující podobu a význam:

AAA	nejlepší úvěrové hodnocení
	schopnost takového dlužníka dostát svým závazkům je velice silná
AA	liší se pouze nepatrně od hodnocení AAA
	schopnost takového dlužníka dostát svým závazkům je také silná
A	větší citlivost na nepříznivý vývoj vnějších okolností a ekonomických podmínek
	schopnost dostát svým závazkům je i tak poměrně silná
BBB	vykazuje odpovídající parametry ochrany proti nepříznivému vývoji
	nepříznivý vývoj ale může způsobit, že takový dlužník nedostojí svým závazkům

<i>BB</i>	nejistota ve splnění závazků
	nepříznivý vývoj s velkou pravděpodobností způsobí neschopnost splatit závazky
<i>B</i>	velice zranitelná společnost
	může dostát svým závazkům, ale nepatrně nepříznivý vývoj způsobí neschopnost splácet dluhy
<i>CCC</i>	společnost objektivně náchylná k nesplácení
	schopnost splácet závisí na příznivém vývoji vnějších podmínek, při nepříznivém vývoji není pravděpodobné, že splatí své závazky
<i>D</i>	neschopnost splácet
	vypořádání závazku neproběhlo v čase dle smluvního ujednání a není pravděpodobné, že k němu dojde později
<i>NR</i>	rating nebyl požadován, nespolehlivé informace, nehodnocení z důvodu zamítavého postoje společnosti

Agentura Standard & Poor's uvádí, že investice do společností s ohodnocením *BB* a nižším patří do kategorie spekulací. Pro každou ratingovou kategorii pak může být určena pravděpodobnost selhání, což lze matematicky zapsat na příkladě ratingové škály agentury Standard & Poor's jako zobrazení

$$\{AAA, AA, \dots, D\} \longrightarrow \langle 0, 1 \rangle.$$

Události v nedávné době bohužel ale také poukázaly na nedostatky při využívání externích ratingových ohodnocení. Ukázalo se, že ratingové agentury vykazují jisté zpoždění ve zveřejňování svých ohodnocení oproti aktuálnímu vývoji. Nebylo totiž neobvyklé, že společnost s vysokým ratingovým ohodnocením byla v situaci, kdy by její ohodnocení mělo patřit do kategorie spekulativních investic. Ukázalo se, že je nezbytné spoléhat se kromě externích ohodnocení i na analýzu potenciálních dlužníků založenou na aktuální situaci.

Nyní se vrátíme k popisu parametrů používaných k modelování úvěrového rizika.

**Expozice při selhání** Tato veličina vyjadřuje, jak velká je aktuální výše možných nesplacených pohledávek v případě selhání protistrany.

**Ztráta při selhání** *Ztráta při selhání* vyjadřuje část pohledávek, kterou v případě selhání dlužníka věřitel nezíská zpět. Přibližně ji můžeme vyjádřit v relativní podobě vzhledem k expozici při selhání pomocí *stupně pokrytí*, který označíme jako *REC*. Ztráta při selhání pak může být vyjádřena jako

$$(1 - REC) \cdot \text{expozice při selhání}.$$

Problém při stanovení ztráty při selhání spočívá v tom, že do jejího vyjádření nelze přesně zahrnout, jak moc je ovlivněná kvalitou záruk úvěru, jak je konkrétní dluh nadřizen či podřizen jiným dluhům atd.

**Očekávaná ztráta** Ztrátu můžeme pomocí dosud definovaných veličin určit jako součin pravděpodobnosti selhání, ztráty při selhání a expozice při selhání. Jedná se o náhodnou veličinu, takže když vyjádříme její střední hodnotu, získáme očekávanou hodnotu, tedy *očekávanou ztrátu*.

## 2.1 Modely pro řízení úvěrového rizika

Silný důraz na řízení úvěrového rizika v metodice Basel II a rozvoj trhů s úvěrovými deriváty vedly k vývoji mnohých kvantitativních modelů pro řízení úvěrového rizika. Můžeme mezi nimi vysledovat dvě hlavní skupiny (podrobněji viz. např. [10]) - analýza rizikových úvěrových cenných papírů a řízení úvěrového rizika v ostatních případech. Modely používané při řízení úvěrového rizika se zaměřují především na stanovení rozdělení ztrát portfolia půjček nebo dluhopisů během pevně stanoveného časového období, nejčastěji minimálně jednoho roku. Z tohoto důvodu jsou tyto modely obvykle statické ve významu, že hlavním úkolem je modelování rozdělení ztrát v pevně daném časovém intervalu místo modelování stochastického procesu popisujícího vývoj rizika v čase. Na druhé straně pro modelování rizika úvěrových cenných papírů se využívají právě modely dynamické, protože výplata mnohých závisí na okamžiku selhání. V této práci se zaměříme na statické modely pro řízení úvěrového rizika.

Modely pro úvěrové riziko mohou být také rozděleny do dvou skupin podle toho, jak jsou definovány. Jsou to modely *strukturální* a modely *s redukovanou formou*. Toto rozdělení jde napříč modely statickými a dynamickými. Většina modelů strukturálních vychází z modelu Mertonova (1974), který popisuje riziko selhání společnosti pomocí modelování vývoje jejích aktiv a pasiv, přičemž okamžik selhání se objevuje v momentě, kdy náhodná veličina představující hodnotu aktiv klesne pod určitou hranici představující závazky. V modelech s redukovanou formou je příčina způsobující selhání ponechána stranou a selhání se modeluje pomocí nezáporné náhodné veličiny reprezentující dobu do selhání. Její rozdělení pak obvykle závisí na ekonomických faktorech. Podrobněji jsou myšlenky jednotlivých modelů zpracovány např. v [10]. V následující části se budeme věnovat jednomu z přístupů, který je využíván modely s redukovanou formou.

### 2.1.1 Modely založené na úvěrové migraci

V této části se budeme věnovat modelům, ve kterých se pravděpodobnost selhání určuje na základě *úvěrové migrace*. Tento přístup spadá do kategorie modelů s redukovanou formou. Podstatou úvěrové migrace je zařazení každé společnosti do určité kategorie ohodnocení neboli ratingu v daném čase. Každá z kategorií vyjadřuje kvalitu společnosti z hlediska schopnosti dostát svým závazkům od nejlepší k nejhorší, nechybí ani kategorie odpovídající selhání. Takových ratingových kategorií je zřejmě konečný počet. Předmětem dalších analýz jsou pak pohyby mezi kategoriemi, tzv. úvěrová migrace, resp.



pravděpodobnost, že dojde ke změně ratingové kategorie. *Pravděpodobnosti přechodů* mezi ratingovými kategoriemi jsou zveřejňovány ratingovými agenturami (viz. výše) obvykle v podobě *matic pravděpodobností přechodu*, které jsou odvozeny na základě historických dat. Modely vycházející z úvěrové migrace se ale mohou uplatnit i v případě interního ohodnocení.

Matice pravděpodobností přechodu jsou standardně zveřejňovány pro jednoletý časový interval. Obsahují tak pravděpodobnosti, s kterými daná společnost dosáhne na konci roku některého z ratingových ohodnocení, případně selhání, pokud na začátku roku měla nějaké konkrétní ratingové ohodnocení. Příkladem jednoleté matice pravděpodobností přechodu může být následující tabulka zveřejněná agenturou Standard & Poor's v roce 1996 (viz. [10]), přičemž poslední řádek pro stav *D* se obvykle vynechává.

rating na začátku roku	rating na konci roku (%)							
	<i>AAA</i>	<i>AA</i>	<i>A</i>	<i>BBB</i>	<i>BB</i>	<i>B</i>	<i>CCC</i>	<i>D</i>
<i>AAA</i>	90,81	8,33	0,68	0,06	0,12	0,00	0,00	0,00
<i>AA</i>	0,70	90,65	7,79	0,64	0,06	0,14	0,02	0,00
<i>A</i>	0,09	2,27	91,05	5,52	0,74	0,26	0,01	0,06
<i>BBB</i>	0,02	0,33	5,95	86,93	5,30	1,17	1,12	0,18
<i>BB</i>	0,03	0,14	0,67	7,73	80,53	8,84	1,00	1,06
<i>B</i>	0,00	0,11	0,24	0,43	6,48	83,46	4,07	5,20
<i>CCC</i>	0,22	0,00	0,22	1,30	2,38	11,24	64,86	19,79
<i>D</i>	0	0	0	0	0	0	0	1

V přístupu vycházejícím z úvěrové migrace se předpokládá, že veškeré informace mající vliv na pravděpodobnosti selhání jsou obsaženy v aktuálním ratingovém ohodnocení, takže pravděpodobnosti selhání mohou být určeny přímo z matice pravděpodobností přechodu. Jak je vidět na příkladě tabulky agentury Standard & Poor's, jednoleté pravděpodobnosti selhání pro společnosti s daným ohodnocením jsou prvky posledního sloupce matice pravděpodobností přechodu. Například pro společnost s ratingem *BB* na začátku roku je pravděpodobnost selhání během roku 1,06%. Agentury publikují matice pravděpodobností přechodu i pro víceleté časové intervaly, v nich jsou pak obsaženy kumulované pravděpodobnosti přechodu mezi kategoriemi ohodnocení během delšího než jednoletého období včetně posledního sloupce pro pravděpodobnosti selhání během daného víceletého období. Tyto matice jsou obvykle určeny přímo z historických dat. Víceleté pravděpodobnosti přechodu můžeme ale stanovit na základě matice jednoleté. Budeme-li předpokládat, že proces úvěrové migrace se řídí homogenním Markovovým řetězcem, *n*-letá matice pravděpodobností přechodu je pak jednoduše *n*-tou mocninou matice jednoleté s *n*-letými pravděpodobnostmi selhání v posledním sloupci. Podrobněji se tomuto přístupu budeme věnovat v kapitole 3.3.

Už jsme zmínili, že zvláště v poslední době se ukázala nevýhoda využívání ratingového ohodnocení od externích agentur. Ohodnocení jimi zveřejněná totiž vykazují jisté zpoždění oproti aktuálnímu vývoji na trhu. Tento jev je důsledkem metodiky, kterou agentury používají. Pravděpodobnosti přechodu a tak i pravděpodobnosti selhání jsou totiž odvozovány na základě historických dat. Může se tedy stát, že publikované hodnoty pravděpodobností selhání bývají vyšší než aktuální v době ekonomického růstu a naopak nižší než reálné v době ekonomické recese. Na druhou stranu z dlouhodobého hlediska je odvození pravděpodobností selhání na základě historické informace výhodou. Během dlouhého období, které tvoří základnu pro stanovení pravděpodobností selhání, došlo k vystřídání několika různých ekonomických scénářů, tedy jak růstu tak recese, které jsou pak v hodnotách pravděpodobností selhání zahrnuty. Z tohoto důvodu jsou pravděpodobnosti selhání získané pomocí úvěrové migrace nezávislé na aktuálním ekonomickém prostředí. Chceme-li pak pomocí úvěrové migrace hodnotit krátkodobé úvěrové kontrakty, je vhodné upravit průměrné pravděpodobnosti selhání podle aktuální situace na trhu.

## Kapitola 3

# Časová struktura pravděpodobnosti selhání

Při modelování úvěrového rizika jsme postaveni před úkol najít parametry pro dlouhé časové intervaly, neboť úvěrové kontrakty jsou obvykle dlouhodobého charakteru. Na trhu jsou však k dispozici převážně údaje pro kratší časový horizont, například jednoleté pravděpodobnosti selhání. Musíme tedy najít způsob, jak z dostupných údajů získat celou časovou strukturu pravděpodobností selhání. V této kapitole popíšeme tři možné přístupy. První z nich ukazuje poměrně jednoduchý způsob, jak z jednoletých pravděpodobností selhání získat kumulované pravděpodobnosti selhání pro delší časové intervaly. Výsledky získané tímto postupem ale nemají vlastnosti, které od časové struktury pravděpodobností selhání požadujeme. Jedná se především o nesplnění *monotonie*. V takovém případě můžeme pro horší ratingové ohodnocení dostat menší pravděpodobnost selhání než pro lepší. Navíc pravděpodobnosti selhání mohou vyjít nulové i pro jiné než nejlepší ohodnocení. Ani jedna z těchto situací za předpokladu, že ratingové ohodnocení je stanoveno správně, není příliš realistická. A tak se využívá spíše druhá metoda popsaná v kapitole 3.3, která vychází z *úvěrové migrace*. Ta umožňuje tyto nepříjemnosti odstranit. Třetí přístup vysvětlený v kapitole 3.4 se používá hlavně pro úvěrové deriváty. Využívá analogie mezi odvozením úvěrové křivky a výnosové křivky a vychází z cen dluhopisů nebo pevných plateb swapů, jejichž hodnoty jsou k dispozici na trhu.

Jeden z prvních úkolů, který musíme při modelování úvěrového rizika vyřešit, je vybrat si odpovídající pravděpodobnostní míru. Buď to budeme využívat rizikově neutrální nebo reálné pravděpodobnosti, které riziko zohledňují. Volba mezi nimi není zcela jednoznačná, závisí na konkrétní problematice, kterou se chceme zabývat. Chceme-li stanovit minimální kapitálové požadavky a rizikové poplatky, využijeme přístup vycházející z historických pravděpodobností selhání. Někdy se také díky podobnosti s aktuárskými metodami nazývá *aktuárský*. V tomto případě předpokládáme, že množství kapitálu, které pokrýje ztráty, je spolehlivě stanoveno aktuálními pravděpodobnostmi selhání,

ačkoli ty jsme určili pomocí historických dat. Uvažujeme přitom i takové ztráty, které nastanou v důsledku nejhorších z uvažovaných scénářů. Při oceňování úvěrových derivátů je situace jiná. V tomto případě vycházíme nejčastěji z teorie pro oceňování opcí. Ta využívá bezrizikové pravděpodobnosti selhání. V rizikově neutrálním prostředí jsou všichni účastníci trhu indiferentní k riziku a nepožadují žádnou kompenzaci za jeho přebírání. U úvěrového rizika se rizikově neutrální pravděpodobnosti selhání porovnají s pravděpodobnostmi získanými z cen dluhopisů nebo pevných plateb swapů a navýší o rozdíl, který mezi nimi je. Tento rozdíl odpovídá prémii za riziko, kterou v rizikovém prostředí požadují účastníci trhu za přebírání rizika.

K ohodnocení rizika lze kromě pravděpodobností selhání využít i *kreditní spready* neboli *úvěrové rozpětí*. Úvěrové rozpětí se často používá při porovnávání dvou dluhopisů. Mějme dva dluhopisy, jeden státní a druhý firemní, přičemž státní dluhopis představuje nástroj s prakticky bezrizikovým výnosem do splatnosti. Úvěrové rozpětí pak představuje rozdíl mezi výnosem do splatnosti firemního a bezrizikového dluhopisu. Cenu bezkupónového bezrizikového dluhopisu v čase  $t$  pro dobu do splatnosti  $T$ , kde  $t \leq T$ , s úrokovou mírou  $r$  definujeme jako  $p_0(t, T) = \exp(-r(T - t))$ . Analogicky, cenu rizikového dluhopisu budeme značit jako  $p_1(t, T)$ . Označíme-li úvěrové rozpětí v čase  $t$  pro časovou dobu splatnosti  $T$  jako  $c(t, T)$ , můžeme jej vyjádřit (viz. [10]) jako

$$c(t, T) = \frac{-1}{T - t} (\ln p_1(t, T) - \ln p_0(t, T)) = \frac{-1}{T - t} \ln \frac{p_1(t, T)}{p_0(t, T)}.$$

Při modelování úvěrového rizika pomocí úvěrového rozpětí se pak místo pravděpodobností selhání analyzují křivky úvěrového rozpětí z hlediska jejich tvaru a pohybu v čase. Oba postupy by měly dát stejné výsledky a to díky spojitosti mezi úvěrovým rozpětím a pravděpodobnostmi selhání. Vztah mezi úvěrovým rozpětím a pravděpodobnostmi selhání je analogický vztahu mezi úrokovou mírou a diskontním faktorem pro pevně daný výnos. Tento vztah nyní nastíníme.

Mějme *stupeň pokrytí*  $REC$ ,  $REC \in (0, 1)$ , dobu do splatnosti  $T$  rizikového dluhopisu s jednotkovou nominální hodnotou a dobu do selhání  $\tau \leq T$ . V případě selhání vlastník dluhopisu obdrží přesně v okamžik selhání platbu velikosti  $(1 - REC_\tau)$ , takže hodnota rizikového dluhopisu v okamžiku splatnosti je

$$p_1(T, T) = \chi\{\tau > T\} + \frac{1 - REC_\tau}{p_0(\tau, T)} \cdot \chi\{\tau \leq T\}.$$

Pro jednoduchost budeme předpokládat, že stupeň pokrytí je konstantní

$$REC_\tau = REC.$$

Multiplikativní úvěrové rozpětí  $s$  pak splňuje

$$DP = \frac{1 - \frac{1}{1+s}}{1 - REC} \approx \frac{s}{1 - REC},$$

kde  $DP$  je pravděpodobnost selhání se zahrnutou rizikovou přírážkou (viz. [3]). Tato aproximace platí i v případě aditivních úvěrových rozpětí.

Dříve než podrobněji popíšeme jednotlivé metody pro určení časové struktury pravděpodobností selhání, zavedeme veličiny, které budeme dále využívat.

### 3.1 Funkce spolehlivosti a intenzita poruch

V této části budeme používat následující značení:

- $S(t)$  ... funkce spolehlivosti,
- $\tau$  ... doba do selhání,
- $F(t)$  ... distribuční funkce doby do selhání,  $F(t) = 1 - S(t)$ ,
- $h(t)$  ... intenzita poruch,
- $DP_t$  ... pravděpodobnost selhání do času  $t$  získaná z tržních dat.

Jedním ze základních parametrů využívaných při modelování úvěrového rizika je spojitá náhodná veličina  $\tau$ , která vyjadřuje *dobu do selhání*. Pomocí doby do selhání definujeme *funkci spolehlivosti* v čase  $t$  označenou jako  $S(t)$  vztahem

$$S(t) = P(\tau > t), \quad t \geq 0.$$

Za předpokladu, že nemáme k dispozici žádnou další informaci ohledně budoucího vývoje selhání,  $S(t)$  vyjadřuje pro počátek v  $t = 0$  kumulovanou pravděpodobnost, že protistrana neselže do času  $t$ . Zřejmě platí  $S(0) = 1$ . Pravděpodobnost selhání během časového intervalu  $(s, t)$  pro  $t \geq s$  může pak být vyjádřena jako rozdíl  $S(s) - S(t)$ . Speciálně pro  $s = 0$  dostáváme kumulovanou *pravděpodobnost selhání do  $t$  označenou  $F(t)$*  jako

$$F(t) = 1 - S(t) = P(\tau \leq t), \quad t \geq 0. \quad (3.1)$$

$F(t)$  je distribuční funkcí náhodného času selhání  $\tau$ . Odpovídající hustota je

$$f(t) = F'(t) = -S'(t).$$

Dále zavedeme *podmíněnou pravděpodobnost selhání*

$$q(t|s) = P(\tau \leq t \mid \tau > s), \quad t \geq s \geq 0,$$

kteřá vyjadřuje pravděpodobnost selhání určitého dlužníka mezi časy  $t$  a  $s$  za podmínky, že neselhal do času  $s$ . Analogicky dostaneme *podmíněnou pravděpodobnost přežití*

$$p(t|s) = 1 - q(t|s) = P(\tau > t \mid \tau > s) = S(t)/S(s), \quad t \geq s \geq 0.$$

Rozdělení času selhání  $\tau$  lze vyjádřit také pomocí *intenzity poruch*, která vyjadřuje okamžitou pravděpodobnost selhání v čase  $t$  za podmínky, že k selhání nedošlo do času  $t$ . Intenzita poruch je motivována vztahem

$$P(t < \tau \leq t + \Delta t \mid \tau > t) = \frac{F(t + \Delta t) - F(t)}{1 - F(t)} \approx \frac{f(t)\Delta t}{1 - F(t)}$$

a definována jako

$$h(t) = \frac{f(t)}{1 - F(t)}.$$

Z (3.1) plyne, že

$$h(t) = \frac{f(t)}{1 - F(t)} = -\frac{S'(t)}{S(t)},$$

příčemž řešení této diferenciální rovnice vede k vyjádření funkce spolehlivosti  $S(t)$  jako

$$S(t) = e^{-\int_0^t h(s)ds}. \quad (3.2)$$

Podmíněné pravděpodobnosti  $p(t|s)$  a  $q(t|s)$  pak mají tvar

$$p(t|s) = e^{-\int_s^t h(u)du}, \quad (3.3)$$

$$q(t|s) = 1 - e^{-\int_s^t h(u)du}. \quad (3.4)$$

Navíc jako důsledek dostáváme vyjádření kumulované pravděpodobnosti selhání do času  $t$  neboli distribuční funkce času do selhání  $\tau$

$$F(t) = 1 - S(t) = 1 - e^{-\int_0^t h(s)ds},$$

a příslušnou hustotu

$$f(t) = S(t)h(t).$$

## 3.2 Přímé odvození víceleté pravděpodobnosti selhání

Přímé odvození časové struktury pravděpodobnosti selhání vychází z představy, že úvěrový vývoj můžeme posuzovat jako homogenní Markovův řetězec se dvěma stavy, jedním z nich je přežití a tím druhým selhání (podrobněji např. v [3]). Na trhu máme k dispozici pravděpodobnosti selhání  $DP_T$  do času  $T$ . Pomocí nich odvodíme pravděpodobnosti přežití pro kratší diskrétní časové okamžiky  $t_i$ ,  $0 \leq t_1 < \dots < t_n = T$ , využitím následujícího vztahu pro nějaké  $k \in \mathbb{N}$  a pro  $\Delta = T/k$

$$P(\tau > T + \Delta | \tau \geq T) = (1 - DP_T)^{\Delta/T}.$$

Pravděpodobnost selhání pro diskrétní časové okamžiky  $t_i$ , které si můžeme představit jako násobky  $\Delta$  je doplňkem do jedničky k pravděpodobnosti přežití a má tvar

$$DP_{t_i} = 1 - (1 - DP_T)^{t_i/T}. \quad (3.5)$$

Využijeme-li funkci spolehlivosti, kterou jsme definovali v předchozí části, můžeme podmíněnou pravděpodobnost přežití  $p(T|0)$  do  $T$  vyjádřit jako

$$1 - DP_T = P(\tau > T | \tau > 0) = S(T)/S(0) = e^{-\int_0^T h(u)du} = e^{-hT}.$$

Poslední rovností definujeme průměrnou intenzitu poruch  $h$ , která nezávisí na čase. Jednoduchou úpravou ji můžeme vyjádřit jako

$$h = \frac{-\log(1 - DP_T)}{T}.$$

Za předpokladu konstantní intenzity poruch během celé doby splácení dluhu, rovnost (3.5) dává vyjádření pro pravděpodobnosti selhání v diskretních časových okamžicích

$$DP_{t_i} = 1 - e^{-ht_i}.$$

Máme tedy k dispozici kumulované pravděpodobnosti selhání v diskretních časových okamžicích  $DP_{t_i} = F(t_i)$ , takže můžeme přejít k situaci se spojitým časem. Popíšeme postup založený na konstantní intenzitě poruch  $h$ , který je podrobněji popsán např. v [3]. Forwardová pravděpodobnost selhání mezi  $t_i$  a  $t_{i+1}$  v tomto případě odpovídá

$$p(t_{i+1}|t_i) = \frac{DP_{t_{i+1}} - DP_{t_i}}{1 - DP_{t_i}} = 1 - \exp\left(-\int_{t_i}^{t_{i+1}} h(u)du\right).$$

Definujme nyní pro časový interval  $\langle t_i, t_{i+1} \rangle$  průměrnou intenzitu poruch pomocí intenzity poruch na čase závislé jako

$$h_i = \frac{1}{t_{i+1} - t_i} \int_{t_i}^{t_{i+1}} h(u)du, \quad \text{pro } i = 0, \dots, n.$$

Dosazením pravděpodobností selhání v diskretních časových okamžicích dostaneme intenzitu poruch pro čas  $t_i$

$$h_i = -\frac{1}{t_{i+1} - t_i} \log\left(\frac{1 - DP_{t_{i+1}}}{1 - DP_{t_i}}\right).$$

Kumulovaná pravděpodobnost selhání do času  $t$ ,  $t_i \leq t < t_{i+1}$  pak má tvar

$$\begin{aligned} DP_t &= 1 - q(t_i|0) \cdot q(t|t_i) \\ &= 1 - (1 - DP_{t_i}) \left(\frac{1 - DP_{t_{i+1}}}{1 - DP_{t_i}}\right)^{\frac{(t-t_i)}{(t_{i+1}-t_i)}}. \end{aligned}$$

Pro  $0 < t < 1$  probíhá odvození stejným způsobem. Pro  $t > t_n$  časovou strukturu získáme extrapolací za předpokladu konstantní intenzity poruch  $h_{n-1}$  od času  $t_{n-1}$  (viz. [3]),

$$DP_t = 1 - (1 - DP_{t_{n-1}}) \left(\frac{1 - DP_{t_n}}{1 - DP_{t_{n-1}}}\right)^{\frac{(t-t_n)}{(t_{n+1}-t_n)}}.$$

### 3.3 Úvěrová migrace

Spolehlivost postupu odvození časové struktury pravděpodobnosti selhání, který jsme vyložili v předchozí části, značně závisí na množství a kvalitě dat, která máme na trhu k dispozici. Z tohoto důvodu se nabízí využít i jiný přístup, který vychází z *úvěrové migrace*. O myšlence tohoto postupu jsme se již krátce zmínili v kapitole 2, ale nyní se jí budeme věnovat podrobněji.

Největší výhoda postupu založeném na úvěrové migraci spočívá ve snadném odvození víceletých pravděpodobností selhání z jednoletých. Ty jsou běžně dostupné na trhu. Aby ale mohly být tyto víceleté pravděpodobnosti selhání pro delší než jednoleté časové intervaly odvozeny, vyžaduje teorie úvěrové migrace zavést předpoklad týkající se nezávislosti pravděpodobností selhání na okamžiku, ve kterém se stanovují. Jedná se o tzv. *časovou homogenitu*. Protože v praxi se ukazuje, že okamžiky selhání závisí na různých ekonomických faktorech vyvíjejících se v čase, předpoklad nezávislosti pravděpodobností selhání na čase nebude často splněn. Z tohoto důvodu uvádíme v závěru této kapitoly dva statistické testy. První z nich nám umožní porovnat matice přechodových četností s teoretickou maticí pravděpodobností přechodu a druhý poskytne testové kritérium pro rozhodnutí o časové homogenitě příslušného Markovova řetězce.

V úvěrové migraci se k popisu přechodů mezi jednotlivými kategoriemi ohodnocení využívají *Markovovy řetězce* (viz. [12]). Mějme pravděpodobnostní prostor  $(\Omega, \mathcal{A}, P)$  a na něm posloupnost  $X = \{X_n, n \in \mathbb{N}_0\}$  celočíselných náhodných veličin. Nechť  $S$  představuje množinu celých čísel  $i$  takových, že  $i \in S$  právě tehdy, když existuje index  $n \in \mathbb{N}_0$  takový, že  $P(X_n = i) > 0$ . Posloupnost náhodných veličin  $X$  se pak nazývá *Markovův řetězec s diskrétním časem a množinou stavů  $S$* , je-li

$$P(X_{n+1} = j | X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) = P(X_{n+1} = j | X_n = i) \quad (3.6)$$

pro všechna  $n \in \mathbb{N}_0$  a  $j, i, i_{n-1}, \dots, i_0 \in S$  taková, že

$$P(X_n = i, X_{n-1} = i_{n-1}, \dots, X_0 = i_0) > 0.$$

Vztah pro pravděpodobnosti (3.6) se nazývá *markovská vlastnost*. Markovův řetězec můžeme analogicky definovat i pro případ se spojitým časem. V takovém případě máme systém celočíselných náhodných veličin  $\{X_t, t \geq 0\}$  s množinou stavů  $S$ , který splňuje markovskou vlastnost v následující podobě

$$P(X_t = j | X_s = i, X_{t_n} = i_n, \dots, X_{t_1} = i_1) = P(X_t = j | X_s = i) \quad (3.7)$$

pro všechna  $j, i, i_n, \dots, i_1 \in S$  a  $0 \leq t_1 < \dots < t_n < s < t$ , pro která je opět

$$P(X_s = i, X_{t_n} = i_n, \dots, X_{t_1} = i_1) > 0.$$

Další podrobnosti týkající se markovských řetězců lze nalézt např. v [12].



V následující části se budeme zabývat Markovovými řetězci s diskrétním časem. *Pravděpodobnosti přechodu 1. řádu* ze stavu  $i$  v čase  $n$  do stavu  $j$  v čase  $n + 1$  jsou definovány rovností (3.6). Dále pro ně budeme používat zjednodušené označení

$$p_{ij}(n, n + 1) = P(X_{n+1} = j | X_n = i).$$

Obecně pro  $k \geq 0$  budeme *pravděpodobnosti přechodu  $k$ -tého řádu* značit

$$p_{ij}(n, n + k) = P(X_{n+k} = j | X_n = i).$$

Pokud pravděpodobnosti přechodu nezávisí na časových okamžicích  $n$  a  $n + k$  ale jen na jejich rozdílu  $k$ , je příslušný Markovův řetězec *homogenní*. Dále se budeme zabývat jen homogenními Markovovými řetězci, pro které pravděpodobnosti přechodu ze stavu  $i$  do stavu  $j$  za časové období délky  $k$  označíme  $p_{ij}(k)$  a pro  $k = 1$  budeme časový údaj vynechávat a zjednodušeně psát  $p_{ij}(1) = p_{ij}$ . Navíc pokud nezdůrazníme, k jakému časovému intervalu se pravděpodobnosti přechodu vztahují, budeme mluvit o jednoletých pravděpodobnostech přechodu  $p_{ij}$ .

Jedním ze základních parametrů, které popisují Markovův řetězec, jsou právě pravděpodobnosti přechodu. Ty můžeme sestavit do čtvercové *matice pravděpodobností přechodu*  $\mathbf{P} = (p_{ij})_{i,j \in S}$ . Matice pravděpodobností přechodu splňuje následující vlastnosti:

- (i) prvky matice  $\mathbf{P}$  jsou nezáporné:  $p_{ij} \geq 0$  pro  $i, j \in S$ ;
- (ii) součet prvků v řádku je roven jedné:  $\sum_{j \in S} p_{ij} = 1$  pro  $i \in S$ .

Matice splňující tyto vlastnosti, tedy i  $\mathbf{P}$ , se nazývá *stochastická matice*. Omezíme se nyní na případ, kdy Markovův řetězec popisuje přechody mezi ratingovými kategoriemi. Budeme tak v situaci, kdy množina stavů  $S$  je konečná,  $S = \{0, 1, \dots, m\}$ . Pokud bychom se vrátili k matici pravděpodobností přechodu mezi ratingovými kategoriemi agentury Standard & Poor's, kterou jsme uvedli v kapitole 2, měla by množina stavů  $S$  prvky

$$S = \{AAA, AA, A, BBB, BB, B, CCC, D\}.$$

Podobu matic pravděpodobností přechodu mezi ratingovými kategoriemi jsme podrobněji popsali v předchozí kapitole 2. Na základě těchto poznatků můžeme zformulovat další vlastnosti speciálně pro matici pravděpodobností přechodu mezi ratingovými kategoriemi  $\mathbf{P}$ :

- (iii)  $m$ -tý sloupec obsahuje jednoleté pravděpodobnosti selhání pro rating  $i$ , označme  $p_{i,m} = DP(1, i)$  pro  $i = 1, \dots, m - 1$ ;
- (iv) stav nesplácení je absorpční:  $p_{m,j} = 0$  pro  $j = 1, \dots, m - 1$  a  $p_{m,m} = 1$ .

Pro homogenní Markovův řetězec s pravděpodobnostmi přechodu  $p_{ij}$  a pro  $n \geq 1$  platí

$$p_{ij}^{(n+1)} = \sum_{k \in S} p_{ik}^{(n)} p_{kj}.$$

Lze dokázat, že matice pravděpodobností přechodu  $n$ -tého řádu homogenního Markovova řetězce  $\mathbf{P}(n) = (p_{ij}(n))_{i,j \in S}$  jsou stochastické matice. Navíc pro všechna celá  $m \geq 0$ ,  $n \geq 0$  a  $P(X_m = i) > 0$  pak pravděpodobnosti přechodu  $n$ -tého řádu splňují

$$p_{ij}(n) = P(X_{m+n} = j | X_m = i) = p_{ij}^n, \quad i, j \in S.$$

Platnost této vlastnosti je dokázána například v [12]. Pravděpodobnosti  $n$ -tého řádu jsou tedy  $n$ -tou mocninou pravděpodobností prvního řádu. Budeme-li předpokládat, že Markovův řetězec popisující přechody mezi ratingovými kategoriemi je homogenní, můžeme  $n$ -leté pravděpodobnosti přechodu mezi ratingy sestavené v  $n$ -leté matici pravděpodobností přechodu  $\mathbf{P}(n)$  určit jednoduše jako  $n$ -tou mocninu matice jednoleté

$$\mathbf{P}(n) = \mathbf{P}^n,$$

přičemž zejména vlastnost (iii) jednoleté matice zůstane zachována i pro matici  $n$ -letou. Díky ní získáme pro stav  $i$  pravděpodobnosti selhání během  $n$  let jako

$$DP(n, i) = p_{i,m}(n).$$

Od matice pravděpodobností přechodu pro úvěrová ohodnocení klientů můžeme dále očekávat, že bude splňovat:

- (v) méně rizikové stavy by neměly vykazovat větší pravděpodobnost selhání než více rizikové, např.:

$$p_{i,m} \leq p_{i+1,m}, \quad i = 1, \dots, m-1;$$

- (vi) *řádková monotonie*: pravděpodobnost přechodu do bližších stavů by měla být větší než do vzdálenějších

$$\begin{aligned} p_{i,i+1} &\geq p_{i,i+2} \geq p_{i,i+3} \dots \\ p_{i,i-1} &\geq p_{i,i-2} \geq p_{i,i-3} \dots \end{aligned}$$

- (vii) *sloupcová monotonie*: pravděpodobnost přechodu do daného stavu by měla být větší pro bližší ratingové kategorie

$$\begin{aligned} p_{i+1,i} &\geq p_{i+2,i} \geq p_{i+3,i} \dots \\ p_{i-1,i} &\geq p_{i-2,i} \geq p_{i-3,i} \dots \end{aligned}$$

Ačkoliv je oprávněné očekávat, že bude matice pravděpodobností přechodu mezi ohodnoceními tyto vlastnosti splňovat, často tomu tak u matic odvozených z dat na trhu není. Navíc na základě teorie, kterou jsme teď vyložili, získáme pravděpodobnosti selhání jen pro diskrétní časové okamžiky. Proto se využívají postupy k odvození matic pravděpodobností přechodu jiným způsobem, který zajistí, že budou tyto vlastnosti splněny. Jedna z možných variant jak tuto problematiku řešit, která navíc umožní stanovit matice pravděpodobností přechodu a tedy i pravděpodobnosti selhání pro spojitý čas, spočívá ve využití konstrukce matice přechodu pomocí matice intenzit (viz. např. [3]).

**Matice intenzit** Matice přechodu mezi jednotlivými ratingovými stavy jsou publikovány obvykle pro jeden rok. Pro kratší časový interval chybí totiž dostatečné množství dat, které by zajistilo spolehlivé stanovení matice pravděpodobností přechodu. Postupovat v tomto případě podobně jako při odvozování víceletých matic pravděpodobností přechodu, tedy pomocí odmocnin, není vhodné. Je to z toho důvodu, že odmocniny matice přechodu nemusí být stochastické matice a tak nespĺňují všechny podmínky (i) - (vii). Navíc v případě, kdy existuje více řešení pro odmocninu, vyvstává problém, které řešení považovat za správné. Je tedy nutné převést dosud diskrétní přístup na variantu se spojitým časem a uvažovat Markovův řetězec se spojitým časem. Pomocí něj definujeme matici intenzit přechodu pomocí derivací pravděpodobností přechodu v bodě 0 a pak následně využijeme souvislost mezi intenzitami přechodu a derivacemi pravděpodobností přechodu v obecném bodě.

V Markovově řetězci se spojitým časem s vlastností (3.7) označme pravděpodobnost přechodu ze stavu  $i$  v čase  $s$  do stavu  $j$  v čase  $t$

$$P(X_t = j | X_s = i) = p_{ij}(s, t), \quad s \geq 0 \text{ a } t > 0.$$

Pro homogenní Markovův řetězec se spojitým časem pak analogicky  $p_{ij}(t)$  jsou pravděpodobnosti přechodu ze stavu  $i$  do stavu  $j$  za čas  $t > 0$ .

**Definice.** Matice  $Q = (q_{ij})$ , kde

$$\begin{aligned} q_{ii} &= -q_i = -\lim_{h \rightarrow 0^+} \frac{1 - p_{ii}(h)}{h} \leq \infty \\ q_{ij} &= \lim_{h \rightarrow 0^+} \frac{p_{ij}(h)}{h} < \infty \end{aligned}$$

se nazývá *matice intenzit* přechodu. Nezáporná čísla  $q_{ij}$  se nazývají *intenzity přechodu* ze stavu  $i$  do stavu  $j$ , nezáporné číslo  $q_i$  se nazývá *celková intenzita*.

Je-li množina stavů  $S$  konečná, což v našem případě s ratingovými kategoriemi vždy bude, platí

$$q_i = \sum_{j \neq i} q_{ij} \quad \forall i \in S. \quad (3.8)$$

**Věta 3.3.1** (Kolmogorovy diferenciální rovnice). Předpokládejme, že  $q_i < \infty$  pro všechna  $i \in S$  a platí (3.8). Potom pro  $t > 0$  platí

$$\mathbf{P}'(t) = \mathbf{Q}\mathbf{P}(t) \quad (3.9)$$

(retrospektivní rovnice).

*Důkaz.* Důkaz nalezneme v [12]. □

**Věta 3.3.2.** Necht'  $\mathbf{Q} = (q_{ij})_{i,j \in S}$  je matice, pro jejíž prvky platí

$$q_{ij} \geq 0, \quad i \neq j \quad q_{ii} = - \sum_{i \neq j} q_{ij}.$$

Potom existuje jediné řešení soustavy (3.9), které vyhovuje počáteční podmínce  $\mathbf{P}(0) = \mathbf{I}$ , kde  $\mathbf{I}$  je jednotková matice příslušných rozměrů. Toto řešení představuje soustavu pravděpodobností přechodu Markovova řetězce se spojitým časem a konečnou množinou stavů. Toto řešení lze zapsat ve tvaru

$$\mathbf{P}(t) = e^{\mathbf{Q}t} = \sum_{k=0}^{\infty} \frac{(\mathbf{Q}t)^k}{k!}. \quad (3.10)$$

*Důkaz.* Důkaz uveden v [12]. □

Bohužel pro empirické matice pravděpodobností přechodu není existence řešení pro matici intenzit zaručena. Navíc situaci komplikuje fakt, že máme obvykle k dispozici pouze matici pravděpodobností přechodu pro jedno období. Avšak rovnice (3.10) nám poskytuje určitý návod, jaký mají mít tvar výsledné intenzity pro danou matici pravděpodobností přechodu. Problematika nalezení matice intenzit pro empirické matice přechodu byla velice podrobně zkoumána, shrnutí některých poznatků nabízí [3].

### 3.3.1 Statistická analýza v Markovových řetězcích

V této části se budeme opět zabývat výhradně homogenními Markovovými řetězci. Mějme homogenní Markovův řetězec  $X = \{X_n, n \in \mathbb{N}_0\}$  s konečnou množinou stavů  $S = \{0, 1, \dots, m\}$ . Pravděpodobnosti přechodu homogenního Markovova řetězce  $p_{ij}(n, n+k)$  nezávisí na časových okamžicích  $n$  a  $n+k$ , ale jen na jejich rozdílu  $k$ . Pravděpodobnostní rozdělení  $\mathbf{p} = (p_1, \dots, p_m)$ , kde

$$p_i = P(X_0 = i) \quad i \in S,$$

splňující

$$p_i \geq 0 \quad i \in S; \quad \sum_{i \in S} p_i = 1$$

se nazývá *počáteční rozdělení* Markovova řetězce. Matici pravděpodobností přechodu označíme stejně jako dosud  $\mathbf{P} = (p_{ij})_{i,j \in S}$ . Budeme se nyní zabývat

situací, kdy matici pravděpodobností přechodu  $\mathbf{P}$  neznáme a chceme ji odhadnout z realizace řetězce délky  $n + 1$

$$X_0 = i_0, \dots, X_n = i_n. \quad (3.11)$$

Máme-li homogenní Markovův řetězec s počátečním rozdělením  $\mathbf{p}$  a maticí pravděpodobností přechodu  $\mathbf{P}$ , pak v [12] je dokázáno, že všechna konečně-rozměrná rozdělení tohoto řetězce jsou tvaru

$$P(X_0 = i_0, X_1 = i_1, \dots, X_n = i_n) = p_{i_0} p_{i_0 i_1} \cdots p_{i_{n-1} i_n} = p_{i_0} \prod_{ij} p_{ij}^{n_{ij}}, \quad (3.12)$$

kde  $n_{ij}$  označuje četnost přechodů v (3.11) ze stavu  $i$  do stavu  $j$ . Četnosti přechodů mezi stavy mohou být sestaveny do matice četností  $\mathbf{N} = (n_{ij})_{i,j \in S}$ , přičemž tato matice je realizací náhodné matice  $\widetilde{\mathbf{N}}$ . V dalším uvedeme několik tvrzení, na jejichž základě pak zkonstruujeme testovou statistiku pro test hypotézy o shodě dané teoretické matice pravděpodobností přechodu Markovova řetězce s pozorovanými četnostmi přechodů. Podrobněji je tento postup zpracován v [9].

**Tvrzení 3.3.1.** *Počáteční stav a matice četností přechodu  $(X_0, \widetilde{\mathbf{N}})$  je postačující statistikou pozorování řetězce s neznámou maticí pravděpodobností přechodu  $\mathbf{P}$  a počátečním rozdělením  $\mathbf{p}$ , tj. dvojice  $(X_0, \widetilde{\mathbf{N}})$  obsahuje o neznámých parametrech tolik informace jako celé pozorování.*

*Důkaz.* Důkaz nalezneme v [9]. □

Další tvrzení popisuje vlastnost postačujících statistik.

**Tvrzení 3.3.2.** *Nechť  $g(\mathbf{p}, \mathbf{P})$  je funkce počátečního rozdělení  $\mathbf{p}$  a matice pravděpodobností přechodu  $\mathbf{P}$ . Dále nechť  $g^* = g^*(X_0, \dots, X_n)$  je nestranným odhadem  $g(\mathbf{p}, \mathbf{P})$  založeným na pozorování (3.11). Je možné najít nestranný odhad  $\tilde{g} = \tilde{g}(X_0, \widetilde{\mathbf{N}})$  funkce  $g(\mathbf{p}, \mathbf{P})$ , který závisí pouze na postačující statistice  $(X_0, \widetilde{\mathbf{N}})$  a zároveň nemá větší rozptyl než  $g^*$ .*

*Důkaz.* Důkaz je uveden v [9]. □

Nyní odvodíme maximálně věrohodné odhady pravděpodobností přechodu. Logaritmickou věrohodnostní funkci určíme na základě (3.12). Máme

$$\begin{aligned} \ell &= \log P(X_0 = i_0, \dots, X_n = i_n) \\ &= \log p_{i_0} + \sum_{i=1}^m \left( \sum_{j=1}^{m-1} n_{ij} \log p_{ij} + n_{im} \log \left( 1 - \sum_{j=1}^{m-1} p_{ij} \right) \right). \end{aligned}$$

Pro  $i \in S$  označme  $n_{i\cdot} = \sum_j n_{ij}$  počet přechodů ze stavu  $i$  a  $n_{\cdot i} = \sum_j n_{ji}$  počet přechodů do stavu  $i$ . Je-li  $n_{i\cdot} > 0$ , dostaneme pro pravděpodobnosti přechodu  $p_{i1}, p_{i2}, \dots, p_{im-1}, p_{im}$ , kde

$$p_{im} = 1 - \sum_{j=1}^{m-1} p_{ij},$$

rovnice

$$0 = \frac{\partial}{\partial p_{ik}} \ell = \frac{n_{ij}}{\hat{p}_{ij}} - \frac{n_{im}}{1 - \sum_{k=1}^{m-1} \hat{p}_{ik}}, \quad k = 1, 2, \dots, m-1.$$

Jejich řešením získáme maximálně věrohodné odhady pravděpodobností přechodu

$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i\cdot}}, \quad i = 1, \dots, m, \quad j = 1, \dots, m.$$

Pokud  $n_{i\cdot} = 0$ , tedy v situaci, kdy pozorování neposkytují podklady pro odhad pravděpodobností přechodu ze stavu  $i$ , položíme  $\hat{p}_{ik} = 0$ ,  $k = 1, \dots, m$ .

Odhadem matice pravděpodobností přechodu  $\mathbf{P}$  je potom matice  $\hat{\mathbf{P}}$  *relativních četností přechodu*, kde

$$\hat{\mathbf{P}} = \left( \frac{n_{ij}}{n_{i\cdot}} \right)_{i,j \in S}.$$

$\hat{\mathbf{P}}$  je funkcí postačující statistiky  $(X_0, \tilde{\mathbf{N}})$ .

**$\chi^2$ -test dobré shody** Statistika  $\chi^2$  se využívá v případě, kdy chceme zjistit, zda výsledky nezávislých pokusů odpovídají předpokládaným pravděpodobnostem. Jsme v situaci, kdy máme  $n$  nezávislých realizací náhodných veličin

$$X_1, \dots, X_n, \tag{3.13}$$

kteří nabývají hodnoty  $1, 2, \dots, m$  s pravděpodobnostmi

$$P(X_k = i) = p_i > 0, \quad i = 1, \dots, m; \quad k = 1, \dots, n.$$

Označíme-li četnost výskytu stavu  $i$  v (3.13) jako  $\nu_i$ , veličina

$$\chi^2 = \sum_{i=1}^m \frac{(\nu_i - np_i)^2}{np_i} = \sum_{i=1}^m \frac{\nu_i}{np_i} - n \tag{3.14}$$

má při  $n \rightarrow \infty$  asymptoticky  $\chi^2$ -rozdělení o  $m-1$  stupních volnosti (viz. [9]).

V našem případě využijeme modifikovaný tvar statistiky (3.14), který umožňuje testovat hypotézu o shodě dané teoretické matice pravděpodobností přechodu Markovova řetězce s napozorovanými četnostmi přechodů. Nechť

$X = \{X_n, n \in \mathbb{N}_0\}$  je homogenní Markovův řetězec s konečnou množinou stavů

$$S = \{0, 1, \dots, m\},$$

maticí pravděpodobností přechodu  $\mathbf{P} = (p_{ij})_{i,j \in S}$  a libovolným počátečním rozdělením  $\mathbf{p} = (p_1, \dots, p_m)$ . Mějme realizaci řetězce  $X$  délky  $n + 1$

$$X_0, X_1, \dots, X_n. \quad (3.15)$$

Stejně jako dosud označíme  $n_{ij}$  četnost přechodů ze stavu  $i$  do stavu  $j$  v realizaci (3.15) a  $n_{i\cdot} = \sum_j \nu_{ij}$  četnost výskytu stavu  $i$  v posloupnosti  $X_0, \dots, X_{n-1}$ . Platí

$$n_{i\cdot} = \sum_{k=1}^{n-1} \chi\{X_k = i\}.$$

**Tvrzení 3.3.3. Statistika**

$$\chi^2 = \sum_{ij} \frac{(n_{ij} - n_{i\cdot} p_{ij})^2}{n_{i\cdot} p_{ij}}$$

má pro  $n = \sum_{ij} n_{ij} \rightarrow \infty$  asymptoticky  $\chi^2$ -rozdělení o  $d - m$  stupních volnosti, kde  $d$  vyjadřuje počet kladných pravděpodobností přechodu a  $m + 1$  je počet stavů,  $m = |S| - 1$ .

*Důkaz.* Stručné vysvětlení důkazu je uvedeno v [9]. □

**Test nulové hypotézy poměrem věrohodností** V této části ukážeme test časové homogenity Markovových řetězců tak, jak je navržen v [14], který nám umožní zjistit, pro jaké časové intervaly je modelování pravděpodobností přechodu mezi ratingovými kategoriemi pomocí homogenních Markovových řetězců vyhovující. Využijeme test hypotézy časové homogenity poměrem věrohodností. Nejdříve stručně popíšeme princip testování nulové hypotézy poměrem věrohodností.

Nechť je  $L(\Theta)$  věrohodnostní funkce. K testování nulové hypotézy poměrem věrohodností  $H_0 : \Theta = \Theta_0$  proti alternativě definujeme poměr věrohodnostních funkcí jako

$$\Lambda = \frac{L(\Theta_0)}{\sup_{\Theta} L(\Theta)},$$

přičemž extrém věrohodnostní funkce ve jmenovateli nastává pro  $\Theta = \hat{\Theta}$ , kde  $\hat{\Theta}$  je maximálně věrohodným odhadem parametru  $\Theta$  (podrobněji např. v [10]). Zřejmě platí, že  $\Lambda \leq 1$ . Pokud je  $\Lambda$  blízko jedné, pak nulová hypotéza pravděpodobně platí. K přesnému určení platnosti hypotézy se využívá testové kritérium

$$-2 \ln \Lambda = 2 \left( \ln L(\hat{\Theta}) - \ln L(\Theta_0) \right) \doteq 2 \left( \ell(\hat{\Theta}) - \ell(\Theta_0) \right),$$

kteřé má za platnosti hypotézy přibližně  $\chi^2(q)$  rozdělení. Počet stupňů volnosti  $q$  odpovídá počtu odhadovaných parametrů, tedy  $q = \dim(\Theta)$ .

Vraťme se nyní k našemu problému, testu hypotézy, zda napozorované četnosti přísluší k homogennímu Markovovu řetězci. Na základě realizace Markovova řetězce  $X$  (ne nutně homogenního) s množinou stavů  $S = \{0, 1, \dots, m\}$  délky  $n + 1$

$$X_0, X_1, \dots, X_n$$

určíme četnosti přechodů  $n_{ij}$  ze stavu  $i$  do stavu  $j$  a četnosti výskytu stavu  $i$   $n_{i\cdot} = \sum_j n_{ij}$ ,  $i, j \in S$ . Maximálně věrohodné odhady pravděpodobností přechodu pak dostaneme jako

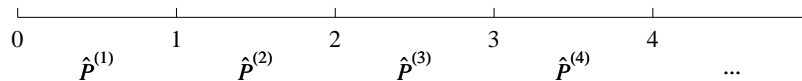
$$\hat{p}_{ij} = \frac{n_{ij}}{n_{i\cdot}}, \quad i, j = 1, \dots, m \quad (3.16)$$

a sestavíme je do matice relativních četností  $\hat{\mathbf{P}}$ .

Mějme nyní  $T$  takových realizací Markovova řetězce v různých po sobě jdoucích časech, tedy dvě po sobě následující realizace jsou od sebe vzdáleny vždy o jedno časové období. K dosud zavedenému značení přidáme index  $t$ ,  $t = 1, \dots, T$ , vyjadřující pořadí dané realizace Markovova řetězce v čase. Následkem toho dostaneme  $T$  matic relativních četností, které označíme  $\hat{\mathbf{P}}^{(t)}$  pro  $t = 1, \dots, T$ , pro které platí

$$\hat{\mathbf{P}}^{(t)} = (\hat{p}_{ij}^{(t)})_{i,j \in S} = \left( \frac{n_{ij}^{(t)}}{n_{i\cdot}^{(t)}} \right)_{i,j \in S}.$$

Graficky je matice relativních četností interpretována na obrázku 3.1.



Obr. 3.1: Grafické znázornění matic relativních četností



Dále mějme matice  $n$ -tého řádu  $\mathbf{P}(n) = (p_{ij}(n))_{i,j \in S}$ , kde  $n = 1, \dots, T$ . Pokud bude Markovův řetězec, jehož realizace teď uvažujeme, homogenní, bude pro nějaké  $n = 1, \dots, T$  platit (viz. [12])

$$\mathbf{P}(n) = \prod_{\Delta \leq t \leq \Delta+n-1} \widehat{\mathbf{P}}^{(t)}, \quad \Delta = 1, \dots, T-n.$$

Právě tato vlastnost bude předmětem testu poměrem věrohodností. Budeme tedy testovat nulovou hypotézu, že pro dané  $n$  matice relativních četností  $\widehat{\mathbf{P}}^{(t)}$ ,  $\Delta \leq t \leq \Delta+n-1$  pro  $\Delta = 1, \dots, T-n$ , nejsou statisticky rozdílné od matice pravděpodobností přechodu  $n$ -tého řádu. Testová statistika k tomuto účelu má pro  $i, j \in S$ ,  $\Delta = 1, \dots, T-n$  následující tvar

$$-2 \ln \Lambda = 2 \sum_{\Delta \leq t \leq \Delta+n-1} \sum_{i,j} n_{ij}^{(t)} \left[ \ln \hat{p}_{ij}^{(t)} - \ln p_{ij}(n) \right]. \quad (3.17)$$

Testová statistika má asymptoticky  $\chi^2$ -rozdělení o  $n \cdot (m+1) \cdot m$  stupních volnosti, kde  $m = |S| - 1$  představuje počet stavů Markovova řetězce (viz. [14]). Je-li nulová hypotéza tohoto testu zamítnuta na určité hladině spolehlivosti, nemůže být Markovův řetězec na této hladině spolehlivosti považován za homogenní.

### 3.4 Časová struktura založená na úvěrovém rozpětí

V této části budeme používat následující značení:

$B(0, t)$	... bezrizikový diskontní faktor na období délky $t$ , neboli současná hodnota bezkupónového dluhopisu s jednotkovou nominální hodnotou a dobou splatnosti $t$ ,
$S(t)$	... funkce spolehlivosti,
$F(t)$	... distribuční funkce doby do selhání, $F(t) = 1 - S(t)$ ,
$X$	... budoucí hodnota rizikové platby v čase $t$ ,
$V$	... nominální hodnota dluhopisu,
$T$	... doba splatnosti dluhopisu,
$C$	... kupón,
$s$	... hodnota pevné platby swapu,
$REC$	... stupeň pokrytí,
$h(t)$	... intenzita poruch.

Časová struktura pravděpodobnosti selhání může být odvozena také na základě informací, které jsou k dispozici na trhu, jako ceny dluhopisů nebo pevně stanovené platby swapů. Tento přístup je nejvíce rozšířen při oceňování

úvěrových derivátů, ale uplatní se i v případě obvyklejších úvěrových obchodů jako jsou úvěry (podrobněji např. v [3]).

Mějme počátek pozorování,  $t = 0$ , a časový horizont  $T$ . Předpokládejme, že okamžik selhání a bezrizikový diskontní faktor jsou na sobě navzájem nezávislé. Potom současná hodnota rizikové platby  $X$ , která se má uskutečnit v čase  $t$ , je rovna

$$B(0, t)S(t)X,$$

kde  $B(0, t)$  je bezrizikový diskontní faktor pro období délky  $t$  a  $S(t)$  funkce přežití do času  $t$ . Uvažujme nyní dluhopis s nominální hodnotou  $V$ , pevným kupónem  $C$  a dobou splatnosti  $T$ . Dále mějme předem pevně stanovené okamžiky plateb  $0 \leq T_1 < \dots < T_n = T$ . Pro jednoduchost se nebudeme zabývat situací, kdy je okamžik platby stanoven na jiný než pracovní den. V takovém případě se kupónová platba uskuteční v nejbližší následující pracovní den a kupónovou platbu je nutné o příslušný počet dnů upravit. V této práci ale budeme nadále předpokládat, že kupónová platba se uskuteční přesně v okamžiku  $T_i$  v případě, že do času  $T_i$  nedošlo k selhání, a její hodnota je  $C$ . Je-li stupeň pokrytí  $REC$  nenulový, je nutné zahrnout jej do výpočtu a zohlednit tak požadavky držitelů dluhopisu v případě selhání. Stupeň pokrytí budeme uvažovat jako pevně daný, založený na kvalitě dluhopisu, pořadí, v jakém bude v případě selhání daný dluh vypořádán atd. V případě selhání vlastníci dluhopisu nezískají žádné budoucí kupóny. Čistá současná hodnota plateb rizikového dluhopisu, kterou označíme  $NPV$ , je pak dána

$$\begin{aligned} NPV = & \sum_{T_i > 0} B(0, T_i) \cdot C \cdot S(T_i) \\ & + V \cdot \left[ B(0, T_n) \cdot S(T_n) + REC \cdot \int_0^{T_n} B(0, t) dF(t) \right]. \end{aligned} \quad (3.18)$$

Integrál v tomto vztahu by mohl být interpretován jako stupeň pokrytí násobený diskontním faktorem pro čas  $t$  a pravděpodobností selhání, kde  $t$  probíhá celý časový interval od počátku až po dobu splatnosti.

Podobně budeme postupovat i pro tzv. *credit default swap* neboli úvěrový derivát, kdy jedna strana platí pravidelně druhé straně pevně stanovené platby za příslib, že druhá strana swapu začne první straně vyplácet platby v případě úvěrového selhání třetí strany (viz. [4]). Pro takový swap máme pevné platby  $s$  v časech  $T_i$ ,  $T_1 < \dots < T_n$ , pokud selhání nenastalo do času  $T_i$ . V případě spravedlivého stanovení pevných swapových plateb na trhu se současná hodnota těchto plateb spolu s hodnotou získanou v případě nesplácení rovná nominální hodnotě swapu, což vyjádříme jako

$$0 = \sum_{i=1}^n B(0, T_i) \cdot s \cdot S(T_i) - V \cdot (1 - REC) \int_0^{T_n} B(0, t) dF(t). \quad (3.19)$$

Budeme-li mít množinu takovýchto spravedlivých swapových plateb nebo cen dluhopisů s rozdílnou dobou do splatnosti a daným stupněm pokrytí,

můžeme určit časovou strukturu pravděpodobnosti selhání. V případě dluhopisů je jen nutné, aby celá množina byla stejného úvěrového hodnocení. Jednu z možných metod stanovení úvěrové křivky vyložíme v následující části.

### 3.4.1 Úvěrová křivka

Na začátku této kapitoly jsme rozdělení doby do selhání  $\tau$  popsali pomocí intenzity poruch  $h(t)$ . Ta jako základ časové struktury pravděpodobnosti selhání je analogií forwardové úrokové míry pro dluhopis s nulovým kupónem. Předpokládejme nyní, že intenzita poruch  $h(t)$  je spojitá a nezávislá na čase, tedy  $h(t) = h$ . Potom  $h\Delta t$  vyjadřuje pravděpodobnost selhání mezi  $t$  a  $t + \Delta$  za podmínky, že nedošlo k selhání do času  $t$ . Budeme-li modelovat dobu do selhání  $\tau$  jako dobu do první události homogenního Poissonova procesu, intenzitou tohoto procesu je pak intenzita poruch  $h$  (definice a vlastnosti viz. [12]). Protože přírůstky Poissonova procesu jsou nezávislé a mají opět Poissonovo rozdělení, má doba do první události exponenciální rozdělení s distribuční funkcí

$$P(\tau \leq t) = 1 - e^{-ht}.$$

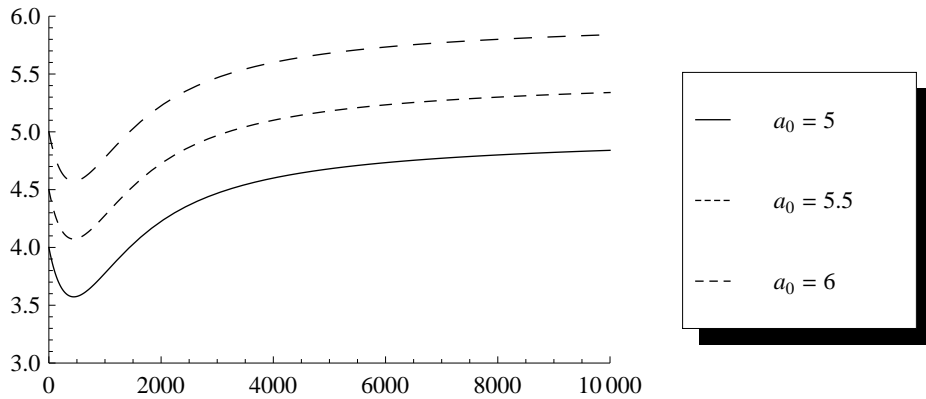
V našem případě se jedná o rozdělení doby do selhání. Rovnice (3.18) a (3.19) spolu s (3.2), kde pro intenzitu poruch nezávislou na čase dostaneme rovnost  $S(t) = e^{-ht}$ , umožňují určit intenzitu poruch z tržně pozorovaných cen dluhopisů nebo konstantních swapových plateb. Využívá se k tomu analogie s výnosovou křivkou.

Výnosová křivka zobrazuje závislost úrokových měr na čase do splatnosti. K její konstrukci z tržních dat se využívají různé statistické a numerické postupy. Jedním z nejjednodušších postupů je interpolace mezi daty, které máme k dispozici. Avšak průběh takto získané křivky mezi známými hodnotami může být často divoký a neodpovídat situaci na trhu. Proto se doporučuje využít spíše než interpolaci postupy vyhlazující průběh křivky (viz. [8]). Jeden z postupů umožňující zachycení všech obvyklých průběhů výnosové křivky, tedy monotónního, hrbatý neboli tzv. *humped*, který je rostoucí pro kratší doby do splatnosti a klesající pro delší, a i tvary odpovídající písmenu S, využívá funkci navrženou Nelsonem a Siegelem (viz. [11]). Ta je definovaná jako

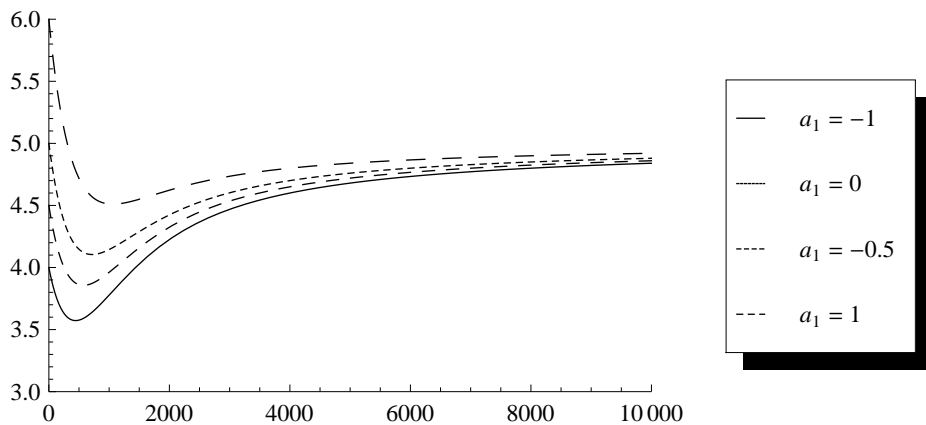
$$f_{\text{NS}}(t; a_0, a_1, a_2, c) = a_0 + (a_1 + a_2) \left( \frac{1 - \exp(-t/c)}{t/c} \right) - a_2 \exp(-t/c), \quad (3.20)$$

kde  $a_0$ ,  $a_1$ ,  $a_2$ ,  $c$  jsou parametry. Jednotlivé koeficienty vyjadřují sílu vlivu dlouhodobých, střednědobých a krátkodobých komponent křivky. Pro  $t = 0$  je hodnota Nelson-Siegelovy funkce rovna  $a_0 + a_1$ , jedná se tedy o počátek křivky. Pro  $t \rightarrow \infty$  konverguje k hodnotě  $a_0$ , který tak má význam dlouhodobého průměru. Parametr  $a_1$  vyjadřuje krátkodobou odchylku od průměru dlouhodobého. Koeficient  $a_2$  reprezentuje střednědobou komponentu, jako jediný má vliv na tvar křivky, tedy bude-li monotónní, *humped* nebo bude-li mít

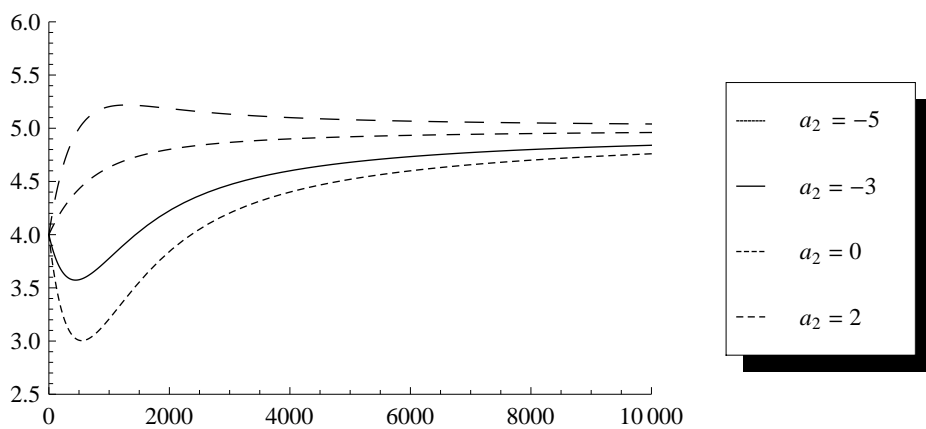
tvar písmene S. Parametr  $c$  má význam lokálního extrému. Následující grafy ukazují vliv změny jednotlivých parametrů, přičemž vycházíme z číselného vyjádření  $a_0 = 5$ ,  $a_1 = -1$ ,  $a_2 = -3$  a  $c = 400$ .



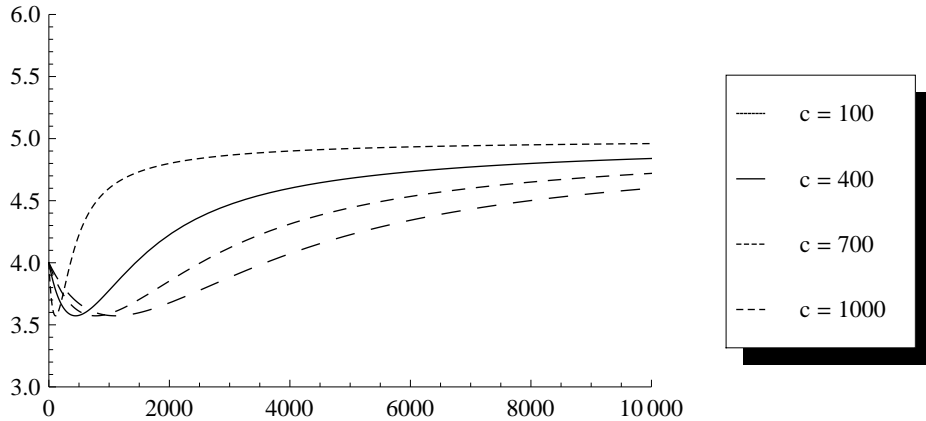
Obr. 3.2: Parametr  $a_0$



Obr. 3.3: Parametr  $a_1$



Obr. 3.4: Parametr  $a_2$



Obr. 3.5: Parametr  $c$

Vrátíme-li se nyní k našemu problému, určení časové struktury pravděpodobnosti selhání, nabízí se možnost využít tuto podobnost s křivkou výnosovou a stanovit tak *úvěrovou křivku*, která bude zachycovat závislost intenzity poruch  $h(t)$  na čase. Pro jednoduchost bychom mohli uvažovat intenzitu poruch konstantní v čase, ale tento postup v praxi obvykle vede k chybným výsledkům. Využívá se tak sice stále deterministického tvaru intenzity poruch, ale navíc se zahrnuje časová závislost, tedy například  $\int_0^t h(s)ds = \Psi(t) \cdot t$ , kde funkce  $\Psi(t)$  zachycuje efekt časové struktury (viz. [3]). Existuje více možností jak funkci  $\Psi(t)$  definovat, jednou z nich je například kubický spline, ale budeme-li vycházet z poznatků o konstrukci výnosových křivek, využijeme stejně jako v jejich případě poznatků Nelsona a Siegela. Ukazuje se totiž, že definice funkce  $\Psi(t)$  pomocí (3.20) je dostatečně flexibilní k odlehlým pozorováním a chybám, které data na trhu často obsahují. Máme tedy

$$\Psi(t; a_0, a_1, a_2, c) = a_0 + (a_1 + a_2) \left( \frac{1 - \exp(-t/c)}{t/c} \right) - a_2 \exp(-t/c). \quad (3.21)$$

Pro vyjádření funkce přežití budeme předpokládat, že parametr  $a_2$  je roven nule. Dostaneme tak tvar křivky, který nejčastěji odpovídá realitě. Rovnice (3.21) spolu s (3.2) určují tvar funkce přežití jako

$$S(t) = \exp \left[ - \left( a_0 + a_1 \left( \frac{1 - \exp(-t/c)}{t/c} \right) \right) \cdot t \right]. \quad (3.22)$$

Při konstrukci úvěrové křivky vycházíme ze skupiny  $N$  dluhopisů nebo swapů nebo kombinace obou. Abychom získali hodnotu parametrů  $\{a_0, a_1, c\}$ , s pomocí (3.22) dosadíme do rovnic (3.18) a (3.19) ceny pozorované na trhu a využijeme nelineární optimalizační algoritmus s počátečními podmínkami  $c > 0$ ,  $S(0) = 1$  a  $S(t) - S(t+1) \geq 0$ . Postupovat regresí využívající absolutní hodnotu odchylky od průměru se v [3] uvádí jako vhodnější postup než přes metodu nejmenších čtverců, neboť ta první reaguje méně citlivě na odlehlá pozorování.

## Kapitola 4

# Modelování úvěrového rizika založené na analýze dat o přežití

V této kapitole se budeme zabývat modelováním úvěrového rizika pro drobnější klienty. V takovém případě totiž máme obvykle nedostatek informací o protistraně a tak klasické modely jako Mertonův nejsou použitelné. Navíc údaje o nesplácení drobnějších klientů vykazují obvykle jen malé procento selhání. Proto využijeme statistický model založený na analýze přežití s cenzorováním pro čas do selhání.

Nejdříve převedeme situaci, kdy dlužník není schopen splnit své závazky, do matematické podoby. Označme  $\chi_i(t)$  indikátor jevu, že dlužník  $i$  přestane splácet v čase  $t$ , tedy

$$\chi_i(t) = \begin{cases} 1 & \text{dojde-li k selhání v čase } t \\ 0 & \text{jinak.} \end{cases}$$

Riziko selhání dlužníka  $i$  v čase  $t$  tak představuje pravděpodobnost

$$P(\chi_i(t) = 1)$$

a čas do selhání  $\tau$  můžeme vyjádřit jako

$$\tau = \inf\{t : \chi_i(t) = 1\}.$$

Čas do selhání  $\tau$  představuje v případě úvěrového rizika náhodnou veličinu, kterou je potřeba modelovat. Jedná se o nezápornou náhodnou veličinu, přičemž v případě, kdy k selhání nedojde, nabývá nekonečné hodnoty. Z historie víme, že k nesplácení dochází jen zřídka, tedy čas do selhání bývá v poměru k době analyzování dlouhý. Nabízí se tedy využít *neúplných výběrů* a poznatků týkajících se náhodného cenzorování. Podrobně je tato problematika popsána např. v [7].

## 4.1 Náhodné cenzorování a intervaly spolehlivosti

V této části budeme používat následující značení:

- $X_i$  ... doba do selhání  $i$ -tého subjektu,
- $F$  ... distribuční funkce doby do selhání a k ní příslušná hustota  $f$ ,
- $T_i$  ... časový cenzor  $i$ -té doby do selhání,
- $G$  ... distribuční funkce časového cenzoru a k ní příslušná hustota  $g$ ,
- $W_i$  ... minimum z doby do selhání a časového cenzoru  $i$ -tého subjektu,  
 $W_i = \min(X_i, T_i)$ ,
- $I_i$  ... náhodná veličina vyjadřující, pro kterou náhodnou veličinu nabývá  $W_i$  minima,  $I_i = \chi\{X_i \leq T_i\}$ .

Předpokládejme, že v čase  $t = 0$  začneme pozorovat  $n$  subjektů stejného typu a že jejich doba do selhání má distribuční funkci  $F$ . Pokud bychom pozorování mohli provádět do té doby, než všechny subjekty podlehnou nesplácení, dostali bychom úplný náhodný výběr a na jeho základě prováděli statistickou indukci o  $F$ . V našem případě ovšem provádíme analýzu dříve, než k takové situaci dojde. Máme tedy k dispozici pouze *neúplný výběr*,  $X_1, \dots, X_n$ , přičemž ne všechna  $X_i$  jsou skutečně pozorovatelná. Budeme-li uvažovat případ, kdy pozorování každého prvku je ukončeno v náhodném okamžiku  $T$ , hovoříme o *náhodném cenzorování* (podrobněji v [7]). Spolu s náhodnou veličinou  $X$  reprezentující dobu do selhání uvažujeme nezápornou náhodnou veličinou  $T$  reprezentující *časový cenzor*. U každého subjektu pak pozorujeme buďto  $X$  nebo  $T$  podle toho, co nastalo dříve, selhání nebo ukončení pozorování, tj.  $\min(X, T)$ . Navíc víme, pro kterou z těchto hodnot minimum nastalo. Výsledek experimentu pak je  $n$  dvojic

$$(W_1, I_1), \dots, (W_n, I_n),$$

kde

$$\begin{aligned} W_j &= \min(X_j, T_j), \\ I_j &= \chi\{X_j \leq T_j\}. \end{aligned}$$

Jedná se tedy o úplný náhodný výběr z dvourozměrného rozdělení, přičemž  $W$  má spojité rozdělení a  $I$  diskrétní, konkrétně alternativní.

### 4.1.1 Metoda maximální věrohodnosti

Předpokládejme, že rozdělení doby do selhání  $X$  je absolutně spojité s distribuční funkcí  $F$  a hustotou  $f$  závislými na neznámém  $k$ -rozměrném parametru  $\Theta$ . Náhodné veličiny  $X_1, \dots, X_n$  jsou stejně rozdělené a nezávislé. Nyní odvodíme věrohodnostní funkci v případě náhodného cenzorování. K tomu musíme dále předpokládat, že časový cenzor  $T$  je náhodná veličina s absolutně spojitým rozdělením s distribuční funkcí  $G$  a hustotou  $g$  a že náhodné

veličiny  $T_1, \dots, T_n$  jsou nezávislé. Navíc jsou i doby do selhání nezávislé na časových cenzorech. Nejdříve odvodíme rozdělení náhodného vektoru  $(W, I)$ . Protože  $W$  vyjadřuje minimum ze dvou absolutně spojitých náhodných veličin, má také absolutně spojitě rozdělení. Pro zkrácení zápisu nebudeme většinou zdůrazňovat závislost na parametru  $\Theta$ . Indikátor  $I$  má alternativní rozdělení. Pro  $w > 0$  dostáváme

$$\begin{aligned}
P(W < w, I = 1) &= P(W < w, X \leq T) = P(X < w, X \leq T) \\
&= \iint_{x < w, x < t} dF(x)dG(t) = \int_0^w \int_x^\infty dF(x)dG(t) \\
&= \int_0^w \left( \int_x^\infty dG(t) \right) dF(x) = \int_0^w (1 - G(x)) dF(x) \\
&= \int_0^w (1 - G(x)) f(x) dx. \tag{4.1}
\end{aligned}$$

Analogicky máme

$$P = (W < w, I = 0) = \int_0^w (1 - F(t)) g(t) dt. \tag{4.2}$$

Těmito pravděpodobnostmi je rozdělení náhodného vektoru  $(W, I)$  plně popsáno. Derivováním (4.1) resp. (4.2) podle  $w$  získáme

$$\begin{aligned}
\frac{\partial P(W < w, I = 1)}{\partial w} &= f(w) [1 - G(w)] \\
\frac{\partial P(W < w, I = 0)}{\partial w} &= g(w) [1 - F(w)].
\end{aligned}$$

Položme nyní

$$h(w, i) = \{f(w) [1 - G(w)]\}^i \{g(w) [1 - F(w)]\}^{1-i} \tag{4.3}$$

pro  $w > 0$  a  $i = 0, 1$ . Funkce  $h(w, i)$  je hustota náhodného vektoru  $(W, I)$  vzhledem k součinové míře Lebesgueova  $\times$  čítací, takže

$$\int_0^\infty \int_{i=0,1} h(w, i) d\lambda(i) dw = \int_0^\infty [h(w, 0) + h(w, 1)] dw = 1,$$

kde  $\lambda(i)$  je čítací míra,  $\lambda(0) = \lambda(1) = 1$ .

Věrohodnostní funkce na základě výsledku experimentu

$$(W_1, I_1), \dots, (W_n, I_n)$$

pak má následující tvar

$$L(\Theta | \underbrace{(W_1, I_1), \dots, (W_n, I_n)}_{(*)}) = \prod_{j=1}^n h(W_j, I_j),$$



přičemž dále budeme používat zkrácené označení  $L(\Theta|(\star))$ .

Označme  $U = \{j; I_j = 1\}$  množinu indexů necenzorovaných pozorování a  $C = \{j; I_j = 0\}$  množinu indexů cenzorovaných pozorování. Věrohodnostní funkci pak můžeme zapsat jako součin přes tyto dvě množiny

$$L(\Theta|(\star)) = \prod_{j \in U} h(W_j, I_j) \prod_{j \in C} h(W_j, I_j).$$

Mohou tedy nastat dvě možnosti, buď

$$j \in U \Rightarrow W_j = X_j \text{ a } I_j = 1$$

nebo

$$j \in C \Rightarrow W_j = T_j \text{ a } I_j = 0.$$

Využitím (4.3) upravíme vyjádření věrohodnostní funkce na tvar využívající pouze distribuční funkce a hustoty doby do selhání  $X$  a časového cenzoru  $T$

$$\begin{aligned} L(\Theta|(\star)) &= \prod_{j \in U} f(W_j) [1 - G(W_j)] \prod_{j \in C} g(W_j) [1 - F(W_j)] \\ &= \prod_{j \in U} f(X_j) [1 - G(X_j)] \prod_{j \in C} g(T_j) [1 - F(T_j)] \end{aligned} \quad (4.4)$$

Parametr  $\Theta$  představuje  $k$ -rozměrný vektor parametrů obou rozdělení  $F$  a  $G$ ,  $\Theta = (\Theta_1, \Theta_2)$ , kde  $\Theta_1$  odpovídá neznámým parametrům rozdělení  $F$  a  $\Theta_2$  vektor neznámých parametrů rozdělení  $G$ . Předpokládejme, že rozdělení  $X$  a  $T$  neobsahují společné parametry a že neexistuje funkční závislost mezi  $\Theta_1$  a  $\Theta_2$ . Protože nás zajímá pouze rozdělení doby do selhání, zajímají nás jen složky vektoru  $\Theta_1$  a vektor  $\Theta_2$  pro nás představuje rušivé parametry. Proto upravíme věrohodnostní funkci do následující podoby

$$L(\Theta|(\star)) = \underbrace{\prod_{j \in U} f(W_j) \prod_{j \in C} [1 - F(W_j)]}_{L_1(\Theta_1|(\star))} \underbrace{\prod_{j \in C} g(W_j) \prod_{j \in U} [1 - G(W_j)]}_{L_2(\Theta_2|(\star))}, \quad (4.5)$$

tedy

$$L(\Theta|(\star)) = L_1(\Theta_1|(\star)) \cdot L_2(\Theta_2|(\star)).$$

Hledání maximálně věrohodného odhadu pro nás zajímavého parametru  $\Theta_1$  se tak zjednoduší, protože

$$\max_{\Theta_1} L(\Theta|(\star)) \Leftrightarrow \max_{\Theta_1} L_1(\Theta_1|(\star)),$$

kde

$$L_1(\Theta_1|(\star)) = \prod_{j \in U} f(X_j) \prod_{j \in C} [1 - F(T_j)]. \quad (4.6)$$

Tedy numerický tvar maximálně věrohodného odhadu  $\hat{\Theta}_1$  nezávisí na  $G$ , ale rozdělení  $\hat{\Theta}_1$  na něm už závisí (podrobněji viz. [7]).

### 4.1.2 Vlastnosti maximálně věrohodných odhadů

Postup hledání maximálně věrohodných odhadů spočívá v přechodu od věrohodnostní funkce k *logaritmické věrohodnostní funkci*  $\ell(\Theta|(\star))$ . Dále se sestaví *systém věrohodnostních rovnic*, ve kterém se parciální derivace podle jednotlivých složek neznámého parametru  $\Theta$  položí rovny nule, tedy

$$\frac{\partial}{\partial \Theta_i} \ell(\Theta|(\star)) = 0, \quad i = 1, \dots, k.$$

V [1] je podrobně odvozeno, že za platnosti podmínek regularity existuje řešení věrohodnostních rovnic  $\widehat{\Theta}$ , pro které platí

$$(i) \quad \widehat{\Theta} \rightarrow \Theta_0 \text{ s. j.,}$$

$$(ii) \quad \sqrt{n}(\widehat{\Theta} - \Theta_0) \xrightarrow{\mathcal{D}} N(0, \mathbf{J}^{-1}(\Theta_0)) \quad n \rightarrow \infty,$$

kde  $\mathbf{J}(\Theta_0)$  je *Fisherova informační matice* a  $\Theta_0$  skutečná hodnota parametru. Výraz v bodě (ii) můžeme pro Fisherovu informační matici a do ní dosazené maximálně věrohodné odhady parametru  $\Theta$ , upravit na

$$\sqrt{n} \mathbf{J}^{1/2}(\widehat{\Theta})(\widehat{\Theta} - \Theta_0) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}_k) \quad n \rightarrow \infty,$$

kde  $\mathbf{I}_k$  označuje  $k$ -rozměrnou jednotkovou matici. Další možností je využít přímo maximálně věrohodný odhad Fisherovy informační matice, tj.

$$\sqrt{n} (\widehat{\mathbf{J}(\Theta)})^{1/2}(\widehat{\Theta} - \Theta_0) \xrightarrow{\mathcal{D}} N(0, \mathbf{I}_k) \quad n \rightarrow \infty.$$

Potřebujeme tedy znát prvky Fisherovy informační matice, resp. její odhad. Za platnosti podmínek regularity máme

$$\begin{aligned} \mathbf{J}_{ij} = & - \text{E} \left[ \frac{\partial^2 \ln h(W, I)}{\partial \Theta_i \partial \Theta_j} \right] = \\ & - \int_0^\infty \int_{\{0,1\}} \frac{\partial^2 \ln h(w, i)}{\partial \Theta_i \partial \Theta_j} h(w, i) \, dw \, d\lambda(i). \end{aligned} \quad (4.7)$$

Dosadíme-li

$$\begin{aligned} \ln h(w, 1) &= \ln f(w) + \ln[1 - G(w)], \\ \ln h(w, 0) &= \ln g(w) + \ln[1 - F(w)] \end{aligned}$$

do (4.7), prvky Fisherovy informační matice mají tvar

$$\begin{aligned} \mathbf{J}_{ij} = & - \int_0^\infty \frac{\partial^2 [\ln f(w) + \ln[1 - G(w)]]}{\partial \Theta_i \partial \Theta_j} f(w)[1 - G(w)] \, dw \\ & - \int_0^\infty \frac{\partial^2 [\ln g(w) + \ln[1 - F(w)]]}{\partial \Theta_i \partial \Theta_j} g(w)[1 - F(w)] \, dw. \end{aligned} \quad (4.8)$$

Prvky Fisherovy informační matice můžeme také přímo odhadnout výběrovým průměrem (viz. [7]), který má tuto podobu

$$\widehat{J}_{ij} = -\frac{1}{n} \sum_{m=1}^n \frac{\partial^2 \ln h(W_m, I_m)}{\partial \Theta_i \partial \Theta_j}. \quad (4.9)$$

## 4.2 Model pro dobu do selhání

V této části budeme používat následující značení:

- $X_i$  ... doba do selhání  $i$ -tého subjektu,
- $F$  ... distribuční funkce doby do selhání a k ní příslušná hustota  $f$ ,
- $T_i$  ... časový cenzor  $i$ -té doby do selhání,
- $W_i$  ... minimum z doby do selhání a časového cenzoru  $i$ -tého dlužníka,  
 $W_i = \min(X_i, T_i)$ ,
- $I_i$  ... náhodná veličina vyjadřující, pro kterou náhodnou veličinu  
nabývá  $W_i$  minima,  $I_i = \chi\{X_i \leq T_i\}$ .

V této části vytvoříme model pro náhodnou veličinu  $X$  představující čas do selhání protistrany, který bude vycházet z teorie náhodného cenzorování. Podrobně je tento postup popsán v [2]. Nevlastní náhodná veličina  $X$  tak může nabývat pouze nekonečné hodnoty nebo hodnoty časového cenzoru. Přičemž v případě, kdy známe maximální možnou délku trvání kontraktu, jako je tomu například u půjček, budeme nahrazovat nekonečno známým  $X_{\max}$ . Náhodnou veličinu  $X$  si tedy můžeme představit jako směs těchto dvou hodnot

$$\begin{aligned} X &= \xi \cdot \infty + (1 - \xi)Z, & \text{resp.} \\ X &= \xi \cdot X_{\max} + (1 - \xi)Z, \end{aligned} \quad (4.10)$$

kde  $P(\xi = 1) = p = 1 - P(\xi = 0)$ . Nezáporná náhodná veličina  $Z$  nezávisí na  $\xi$  a má distribuční funkci  $F_Z$  a k ní příslušnou hustotu  $f_Z$ . V kontextu s půjčkami můžeme  $p$  interpretovat jako část klientů, jejichž pravděpodobnost selhání do  $X_{\max}$  je rovna nule. Distribuční funkci nevlastní náhodné veličiny  $X$  budeme značit jako  $F$ . Díky tomu, jak je doba do selhání  $X$  definována a že rozdělení  $Z$  nezávisí na rozdělení  $\xi$ , můžeme její distribuční funkci v bodě  $x$  vyjádřit pomocí  $p$  a distribuční funkce  $F_Z(x)$  v podobě

$$\begin{aligned} F(x) &= P(X < x) = P(X < x, \xi = 1) + P(X < x, \xi = 0) \\ &= P(X_{\max} < x, \xi = 1) + P(Z < x, \xi = 0) \\ &= \begin{cases} P(\xi = 1) + P(Z < x, \xi = 0) & X_{\max} < x \\ P(Z < x, \xi = 0) & X_{\max} \geq x \end{cases} \\ &= \begin{cases} p + (1 - p)F_Z(x) & X_{\max} < x \\ (1 - p)F_Z(x) & X_{\max} \geq x. \end{cases} \end{aligned} \quad (4.11)$$

Jak jsme uvedli v předchozí teoretické části, budeme pozorovat dvojice  $(W, I)$ , kde  $W = \min(X, T)$  pro náhodný časový cenzor  $T$ . Informaci o tom, zda dané pozorování je či není cenzorované, dává náhodná veličina  $I$  definovaná jako  $I = \chi\{X \leq T\}$ . Předpokládejme, že náhodné veličiny  $X$  a  $T$  jsou vzájemně nezávislé, díky čemuž jsou nezávislé i veličiny  $W$  a  $I$  a zřejmě i dvojice  $(W, I)$ . Dále předpokládejme, že náhodný časový cenzor  $T$  má absolutně spojitou distribuční funkci označenou jako  $G$ . Rozdělení náhodné veličiny  $I$  je určeno pravděpodobností  $p$ , maximální hodnotou  $X_{\max}$ , distribuční funkcí  $F_Z$  a hodnotou časového cenzoru  $T$  jako

$$\begin{aligned}
P(I = 0) &= P(X > T) = \iint_{x>t} dF(x)dG(t) \\
&= \int_0^\infty \left( \int_0^x dG(t) \right) dF(x) = \int_0^\infty G(x)dF(x) \\
&= \int_0^\infty \left( \int_t^\infty dF(x) \right) dG(t) = \int_0^\infty [1 - F(t)]dG(t).
\end{aligned} \tag{4.12}$$

Analogickým postupem získáme i druhou pravděpodobnost jako

$$P(I = 1) = P(X \leq T) = \int_0^\infty [1 - G(x)]dF(x). \tag{4.13}$$

Sdružené rozdělení náhodných veličin  $W$  a  $I$  je potom pomocí (4.12) resp. (4.13) tvaru

$$\begin{aligned}
P(W < w, I = 1) &= P(X < w, X < T) = \iint_{x<w, x<t} dG(t)dF(x) \\
&= \int_0^w \int_x^\infty dG(t)dF(x) = \int_0^w [1 - G(x)]dF(x) \\
&= \begin{cases} \int_0^w [1 - G(x)]d[(1 - p)F_Z(x)] & w < X_{\max} \\ \int_0^w [1 - G(x)]d[p + (1 - p)F_Z(x)] & w > X_{\max}. \end{cases}
\end{aligned} \tag{4.14}$$

Podobně v případě, kdy  $I = 0$  máme

$$\begin{aligned}
P(W < w, I = 0) &= \int_0^w [1 - F(t)]dG(t) \\
&= \begin{cases} \int_0^w [1 - (1 - p)F_Z(t)]dG(t) & w < X_{\max} \\ \int_0^w [1 - p - (1 - p)F_Z(t)]dG(t) & w > X_{\max}. \end{cases}
\end{aligned} \tag{4.15}$$

Derivováním (4.14) podle  $w$  dostaneme

$$\frac{\partial P(W < w, I = 1)}{\partial w} = \begin{cases} (1-p)f_Z(w)[1-G(w)] & w < X_{\max} \\ (p+(1-p))f_Z(w)[1-G(w)] & w > X_{\max} \end{cases} \quad (4.16)$$

a podobně derivací (4.15) podle  $w$

$$\frac{\partial P(W < w, I = 0)}{\partial w} = \begin{cases} g(w)[1-(1-p)f_Z(w)] & w < X_{\max} \\ g(w)[1-p-(1-p)f_Z(w)] & w > X_{\max}. \end{cases} \quad (4.17)$$

Budeme pozorovat skupinu  $n$  dlužníků a ke každému z nich dvojici náhodných veličin  $(W, I)$ , kterou označíme  $(W_i, I_i)$  pro  $i = 1, \dots, n$ . Veličiny  $W_i$  a  $I_i$  jsou definovány stejně jako dosud, tj.  $W_i = \min(X_i, T_i)$  a  $I_i = \chi\{X_i \leq T_i\}$ . Předpokládejme, že náhodné veličiny  $X_1, \dots, X_n$  a  $T_1, \dots, T_n$  jsou vzájemně nezávislé. Pak jsou vzájemně nezávislé i  $W_1, \dots, W_n$  a  $I_1, \dots, I_n$  a přirozeně i dvojice  $(W_i, I_i)$  pro  $i = 1, \dots, n$ . Nadále budeme předpokládat, že časové cenzory mají známou distribuční funkci, kterou jsme už označili  $G$ .

Dále budeme postupovat podobně jako v předchozí části, takže sestavíme logaritmickou věrohodnostní funkci k odhadnutí neznámého parametru popisujícího rozdělení doby do selhání. Předpokládejme tedy, že distribuční funkce  $F_Z$  závisí na  $m$ -rozměrném parametru  $\boldsymbol{\lambda}$ . Pomocí metody maximální věrohodnosti budeme odhadovat  $(m+1)$ -rozměrný parametr  $\boldsymbol{\Theta} = (p, \boldsymbol{\lambda})$ . Tvar logaritmické věrohodnostní funkce udává následující tvrzení.

**Tvrzení 4.2.1.** *Nechť  $U$  označuje množinu indexů příslušejících těm pozorováním, která nejsou cenzorovaná, pro která  $I_i = 1$ , a  $C$  množinu indexů cenzorovaných pozorování, pro která  $I_i = 0$ . Potom má pro  $W_i < X_{\max}$ ,  $i = 1, \dots, n$ , logaritmická věrohodnostní funkce tvar*

$$\ell(\boldsymbol{\Theta}) = |U| \cdot \ln(1-p) + \sum_{i \in U} \ln f_Z(w_i; \boldsymbol{\lambda}) + \sum_{i \in C} \ln(1 - (1-p)F_Z(t_i; \boldsymbol{\lambda})). \quad (4.18)$$

*Důkaz.* Při sestavování věrohodnostní funkce vyjdeme z předchozí části, kde jsme neznámý parametr  $\boldsymbol{\Theta}$  odhadovali pomocí rovnice (4.6). V tomto případě neuvažujeme, že by rozdělení náhodných cenzorů záviselo na nějakém neznámém parametru, tedy věrohodnostní rovnice je rovna přímo (4.6) a má tedy tvar

$$L(\boldsymbol{\Theta}) = \prod_{i \in U} f(X_i) \prod_{i \in C} [1 - F(T_i)].$$

Dosazením za hustotu a distribuční funkci doby do selhání pro  $W_i < X_{\max}$ ,  $i = 1, \dots, n$ , pak dostaneme tvar

$$L(\Theta) = \prod_{i \in U} (1-p) f_Z(w_i; \lambda) \prod_{i \in C} (1 - (1-p) F_Z(t_i; \lambda)). \quad (4.19)$$

Logaritmováním (4.19) a použitím vlastností logaritmu dojdeme k logaritmické věrohodnostní funkci (4.18) uvedené ve znění tvrzení.  $\square$

Derivací logaritmické věrohodnostní funkce (4.18) podle jednotlivých složek parametru  $\Theta = (p, \lambda)$  a položením těchto derivací rovno nule, získáme maximálně věrohodný odhad parametru  $\Theta$ . Pro jednoduchost nebudeme dále většinou značit závislost logaritmické věrohodnostní funkce na parametru  $\Theta$ . Pro  $j = 1, \dots, m$ , jsou parciální derivace (4.18) tvaru

$$\begin{aligned} \frac{\partial}{\partial p} \ell &= -\frac{|U|}{1-p} + \sum_{i \in C} \frac{F_Z(t_i; \lambda)}{1 - (1-p) F_Z(t_i; \lambda)}, \\ \frac{\partial}{\partial \lambda_j} \ell &= \sum_{i \in U} \frac{\frac{\partial f_Z}{\partial \lambda_j}(w_i; \lambda)}{f_Z(w_i; \lambda)} - (1-p) \sum_{i \in C} \frac{\frac{\partial F_Z}{\partial \lambda_j}(t_i; \lambda)}{1 - (1-p) F_Z(t_i; \lambda)}. \end{aligned}$$

Maximálně věrohodný odhad parametru  $\Theta$  označíme  $\hat{\Theta} = (\hat{p}, \hat{\lambda})$ . Označme skutečnou hodnotu parametru  $\Theta$  jako  $\Theta_0 = (p_0, \lambda_0)$ . V [2] je na základě formulovaných podmínek odvozena konvergence v distribuci maximálně věrohodného odhadu  $\hat{\Theta}$  analogická té, kterou jsme uvedli v předchozí teoretické části,

$$\sqrt{n}(\hat{\Theta} - \Theta_0) \xrightarrow{D} N(0, J_n^{-1}(\Theta_0)) \quad n \rightarrow \infty.$$

Fisherova informační matice je za platnosti podmínek regularity (podrobněji viz. [1]) definována jako

$$J_n(\Theta) = E \left[ -\frac{\partial^2}{\partial \Theta^2} \ell \right].$$

Nyní odvodíme vyjádření prvků matice  $J_n(\Theta)$  tak, jak je uvedeno v [2]. Nejdříve určíme prvky pomocné matice  $D_n(\Theta)$  definované jako

$$D_n(\Theta) = -\frac{\partial^2}{\partial \Theta^2} \ell = \sum_{i=1}^n D^{(i)}(\Theta).$$

Fisherovu informační matici pak dostaneme jako střední hodnotu pomocné matice  $D_n(\Theta)$ . Vyjádření prvků matic  $D^{(i)}(\Theta)$  pro  $i = 1, \dots, n$  udává následující tvrzení. Pro zjednodušení zápisu nebudeme závislost distribuční funkce a hustoty náhodné veličiny  $Z$  na vektoru parametrů  $\lambda$  většinou dále vyznačovat. V tvrzení využijeme následující označení

$$\begin{aligned} \alpha_{jk}(w) &= f_Z^{-2}(w) \frac{\partial}{\partial \lambda_j} f_Z(w) \frac{\partial}{\partial \lambda_k} f_Z(w) - f_Z^{-1}(w) \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} f_Z(w), \\ \beta_{jk}(w) &= \frac{\frac{\partial^2}{\partial \lambda_j \partial \lambda_k} F_Z(w)}{1 - (1-p) F_Z(w)} + (1-p) \frac{\frac{\partial}{\partial \lambda_j} F_Z(w) \frac{\partial}{\partial \lambda_k} F_Z(w)}{(1 - (1-p) F_Z(w))^2}. \end{aligned}$$

**Tvrzení 4.2.2.** Pro prvky matic  $\mathbf{D}^{(i)}(\Theta) = (\mathbf{D}_{jk}^{(i)})_{j,k=1,\dots,m+1}$ ,  $i = 1, \dots, n$  platí

$$\mathbf{D}_{11}^{(i)} = \frac{1}{(1-p)^2} \cdot \chi\{I_i = 1\} + \left( \frac{F_Z(t_i)}{1 - (1-p)F_Z(t_i)} \right)^2 \cdot \chi\{I_i = 0\},$$

a pro  $j, k = 2, \dots, m+1$

$$\mathbf{D}_{1k}^{(i)} = -\frac{\partial}{\partial \lambda_{k-1}} F_Z(t_i) (1 - (1-p)F_Z(t_i))^2 \cdot \chi\{I_i = 0\},$$

$$\mathbf{D}_{jk}^{(i)} = \alpha_{j-1,k-1}(w_i) \cdot \chi\{I_i = 1\} + (1-p)\beta_{j-1,k-1}(t_i) \cdot \chi\{I_i = 0\}.$$

*Důkaz.* Parciální derivace prvního řádu logaritmické věrohodnostní funkce tak můžeme ještě po drobné úpravě uvést do následující podoby

$$\begin{aligned} \frac{\partial}{\partial p} \ell &= \sum_{i=1}^n \left( -\frac{\chi\{I_i = 1\}}{1-p} + \frac{F_Z(t_i)}{1 - (1-p)F_Z(t_i)} \cdot \chi\{I_i = 0\} \right), \\ \frac{\partial}{\partial \lambda_j} \ell &= \sum_{i=1}^n \left( \frac{\frac{\partial}{\partial \lambda_j} f_Z(w_i)}{f_Z(w_i)} \cdot \chi\{I_i = 1\} - \right. \\ &\quad \left. - (1-p) \frac{\frac{\partial}{\partial \lambda_j} F_Z(t_i)}{1 - (1-p)F_Z(t_i)} \cdot \chi\{I_i = 0\} \right). \end{aligned} \quad (4.20)$$

Abychom získali prvky matice  $\mathbf{D}(\Theta)$ , musíme vyjádřit druhé derivace (4.20) podle jednotlivých parametrů, čímž dostaneme

$$\begin{aligned} \frac{\partial^2}{\partial p^2} \ell &= \sum_{i=1}^n \left( -\frac{\chi\{I_i = 1\}}{(1-p)^2} - \left( \frac{F_Z(t_i)}{1 - (1-p)F_Z(t_i)} \right)^2 \cdot \chi\{I_i = 0\} \right), \\ \frac{\partial^2}{\partial p \partial \lambda_j} \ell &= \sum_{i=1}^n \frac{\frac{\partial}{\partial \lambda_j} F_Z(t_i)}{(1 - (1-p)F_Z(t_i))^2} \cdot \chi\{I_i = 0\}, \\ \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} \ell &= \sum_{i=1}^n \left[ \left( f_Z^{-1}(w_i) \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} f_Z(w_i) - \right. \right. \\ &\quad \left. \left. - f_Z^{-2}(w_i) \frac{\partial}{\partial \lambda_j} f_Z(w_i) \frac{\partial}{\partial \lambda_k} f_Z(w_i) \right) \cdot \chi\{I_i = 1\} - \right. \\ &\quad \left. - (1-p) \left( \frac{\frac{\partial^2}{\partial \lambda_j \partial \lambda_k} F_Z(t_i)}{1 - (1-p)F_Z(t_i)} + \right. \right. \\ &\quad \left. \left. + (1-p) \frac{\frac{\partial}{\partial \lambda_j} F_Z(t_i) \frac{\partial}{\partial \lambda_k} F_Z(t_i)}{(1 - (1-p)F_Z(t_i))^2} \right) \cdot \chi\{I_i = 0\} \right] \\ &= \sum_{i=1}^n [-\alpha_{jk}(w_i) \cdot \chi\{I_i = 1\} - (1-p)\beta_{jk}(t_i) \cdot \chi\{I_i = 0\}]. \end{aligned} \quad (4.21)$$

Z toho, jak jsme definovali matici  $\mathbf{D}_n(\Theta)$

$$\mathbf{D}_n(\Theta) = -\frac{\partial^2}{\partial \Theta^2} \ell(\Theta) = \sum_{i=1}^n \mathbf{D}^{(i)}(\Theta),$$

je podoba prvků matic  $\mathbf{D}^{(i)}(\Theta)$  na základě (4.21) pro  $j, k = 1, \dots, m$  zřejmě následující

$$\begin{aligned} \mathbf{D}_{11}^{(i)} &= \frac{\chi\{I_i = 1\}}{(1-p)^2} + \left( \frac{F_Z(t_i)}{1 - (1-p)F_Z(t_i)} \right)^2 \cdot \chi\{I_i = 0\}, \\ \mathbf{D}_{1,j+1}^{(i)} &= \frac{\frac{\partial}{\partial \lambda_j} F_Z(t_i)}{(1 - (1-p)F_Z(t_i))^2} \cdot \chi\{I_i = 0\}, \\ \mathbf{D}_{j+1,k+1}^{(i)} &= \alpha_{jk}(w_i) \cdot \chi\{I_i = 1\} + (1-p)\beta_{jk}(t_i) \cdot \chi\{I_i = 0\}. \end{aligned}$$

Rovnost

$$\mathbf{D}_{jk}^{(i)} = \mathbf{D}_{kj}^{(i)} \quad \forall j, k$$

plyne z Cauchy-Schwarzovy nerovnosti díky tomu, že logaritmická věrohodnostní funkce je dvakrát spojitě diferencovatelná (viz. [2]).  $\square$

Budeme-li mít skutečnou hodnotu parametru  $\Theta_0$  a pokud budeme uvažovat střední hodnotu matice  $\mathbf{D}_n(\Theta)$  pro známé  $\Theta_0$ , dostaneme Fisherovu informační matici jako  $\mathbf{J}_n(\Theta) = \sum_{i=1}^n \mathbf{J}^{(i)}(\Theta)$ , kde  $\mathbf{J}_n(\Theta) = (\mathbf{J}_{jk})_{j,k=1,\dots,m+1}$  a  $\mathbf{J}^{(i)}(\Theta) = (\mathbf{J}_{jk}^{(i)})_{j,k=1,\dots,m+1}$ . Vyjádření prvků Fisherovy informační matice popisuje další tvrzení.

**Tvrzení 4.2.3.** *Prvky matice  $\mathbf{J}^{(i)}(\Theta)$  mají tvar*

$$\begin{aligned} \mathbf{J}_{11}^{(i)} &= \frac{1}{1-p_0} \mathbb{E}_{\Theta_0} \left[ \frac{F_Z(T)}{1 - (1-p)F_Z(T)} \right], \\ \mathbf{J}_{1j}^{(i)} &= \mathbf{J}_{j1}^{(i)} = -\mathbb{E}_{\Theta_0} \left[ \frac{\frac{\partial}{\partial \lambda_{j-1}} F_Z(T)}{1 - (1-p)F_Z(T)} \right], \quad j > 1, \end{aligned}$$

a smíšené prvky splňují

$$\begin{aligned} \mathbf{J}_{jk}^{(i)} = \mathbf{J}_{kj}^{(i)} &= -(1-p_0) \mathbb{E}_{\Theta_0} \left[ \int_0^T \alpha_{j-1,k-1}(y) f_Z(y) dy + \right. \\ &\quad \left. + \beta_{j-1,k-1}(T) (1 - (1-p)F_Z(T)) \right], \quad j, k > 1. \end{aligned}$$

Pokud navíc hustota  $f_Z$  splňuje pro všechna  $j, k$  a nějakou  $\sigma$ -konečnou míru  $\mu$

$$\frac{\partial^2}{\partial \Theta_j \partial \Theta_k} \int f_Z(\omega, \Theta) d\mu(\omega) = \int \frac{\partial^2}{\partial \Theta_j \partial \Theta_k} f_Z(\omega, \Theta) d\mu(\omega) \quad (4.22)$$



pak  $\mathbf{J}_{jk}^{(i)}$  pro  $j, k > 1$  lze zjednoduřit na

$$\begin{aligned} \mathbf{J}_{jk}^{(i)} &= -(1-p_0)\mathbf{E}_{\Theta_0} \left[ \int_0^T f_Z^{-1}(y) \frac{\partial}{\partial \lambda_{j-1}} f_Z(y) \frac{\partial}{\partial \lambda_{k-1}} f_Z(y) d(y) + \right. \\ &\quad \left. + (1-p) \frac{\frac{\partial}{\partial \lambda_{j-1}} F_Z(T) \frac{\partial}{\partial \lambda_{k-1}} F_Z(T)}{1 - (1-p)F_Z(T)} \right]. \end{aligned}$$

Navíc prvky matic  $\mathbf{J}^{(i)}(\Theta)$  nezávisí na  $i$ , takže platí

$$\mathbf{J}_n(\Theta) = \sum_{i=1}^n \mathbf{J}^{(i)}(\Theta) = n \cdot \mathbf{J}(\Theta).$$

Důkaz. Prvky matice  $\mathbf{J}_{jk}^{(i)}(\Theta)$  jsme definovali jako  $\mathbf{J}_{jk}^{(i)} = \mathbf{E}_{\Theta_0}[\mathbf{D}_{jk}^{(i)}]$ . Z této rovnosti dostáváme

$$\begin{aligned} \mathbf{J}_{11}^{(i)} &= \mathbf{E}_{\Theta_0} \left[ \frac{\chi\{I_i = 1\}}{(1-p)^2} + \left( \frac{F_Z(T)}{1 - (1-p)F_Z(T)} \right)^2 \cdot \chi\{I_i = 0\} \right] \\ &= \frac{1}{(1-p_0)^2} \mathbf{E}_{\Theta_0} [\chi\{X \leq T\}] + \\ &\quad + \mathbf{E}_{\Theta_0} \left[ \left( \frac{F_Z(T)}{1 - (1-p)F_Z(T)} \right)^2 \cdot \chi\{X > T\} \right] \\ &= \frac{1}{(1-p_0)^2} \mathbf{E}_{\Theta_0} [F(T)] + \mathbf{E}_{\Theta_0} \left[ \left( \frac{F_Z(T)}{1 - (1-p)F_Z(T)} \right)^2 (1 - F(T)) \right] \\ &= \frac{1}{(1-p_0)} \mathbf{E}_{\Theta_0} [F_Z(T)] + \mathbf{E}_{\Theta_0} \left[ \frac{(F_Z(T))^2}{1 - (1-p)F_Z(T)} \right] \tag{4.23} \\ &= \frac{1}{(1-p_0)} \mathbf{E}_{\Theta_0} \left[ \frac{F_Z(T)}{1 - (1-p)F_Z(T)} \right]. \end{aligned}$$

K vyjádření (4.23) jsme předpokládali, že  $X_{\max} \geq T$ , a využili tvar distribuční funkce doby do selhání  $X$  odvozený v (4.11) pro tento případ. Dále s využitím tohoto vztahu pro všechna  $j = 1, \dots, m$  je

$$\begin{aligned} \mathbf{J}_{1,j+1}^{(i)} &= -\mathbf{E}_{\Theta_0} \left[ \frac{\frac{\partial}{\partial \lambda_j} F_Z(T)}{(1 - (1-p)F_Z(T))^2} \cdot \chi\{I_i = 0\} \right] \\ &= -\mathbf{E}_{\Theta_0} \left[ \frac{\frac{\partial}{\partial \lambda_j} F_Z(T)}{(1 - (1-p)F_Z(T))^2} \cdot \chi\{X > T\} \right] \end{aligned}$$

$$\begin{aligned}
&= -\mathbf{E}_{\Theta_0} \left[ \frac{\frac{\partial}{\partial \lambda_j} F_Z(T)}{(1 - (1-p)F_Z(T))^2} (1 - F(T)) \right] \\
&= -\mathbf{E}_{\Theta_0} \left[ \frac{\frac{\partial}{\partial \lambda_j} F_Z(T)}{(1 - (1-p)F_Z(T))} \right].
\end{aligned}$$

A podobně pro všechna  $j, k = 1, \dots, m$

$$\begin{aligned}
\mathbf{J}_{j+1, k+1}^{(i)} &= -\mathbf{E}_{\Theta_0} [\alpha_{jk}(w_i) \cdot \chi\{I_i = 1\} + (1-p)\beta_{jk}(t_i) \cdot \chi\{I_i = 0\}] \\
&= -\mathbf{E}_{\Theta_0} [\alpha_{jk}(X) \cdot \chi\{X \leq T\} + (1-p)\beta_{jk}(T) \cdot \chi\{X > T\}] \\
&= -\mathbf{E}_{\Theta_0} \left[ \int_0^T \alpha_{jk}(y) f(y) d(y) + (1-p)\beta_{jk}(T)(1 - F(T)) \right] \\
&= -(1-p_0)\mathbf{E}_{\Theta_0} \left[ \int_0^T \alpha_{jk}(y) f_Z(y) d(y) + \right. \\
&\quad \left. + (1-p)\beta_{jk}(T)(1 - (1-p)F_Z(T)) \right].
\end{aligned}$$

Za platnosti předpokladu (4.22) uvedeného v tomto tvrzení máme

$$\int_0^T \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} f_Z(y) d(y) = \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} F_Z(T).$$

Tuto rovnost můžeme využít a zjednodušit tak vyjádření  $\mathbf{J}_{j+1, k+1}^{(i)}$ . Dosazením za  $\alpha_{jk}(w)$  a  $\beta_{jk}(w)$  potom dostaneme

$$\begin{aligned}
\mathbf{J}_{j+1, k+1}^{(i)} &= -(1-p_0)\mathbf{E}_{\Theta_0} \left[ \int_0^T \left( f_Z^{-1}(y) \frac{\partial}{\partial \lambda_j} f_Z(y) \frac{\partial}{\partial \lambda_k} f_Z(y) - \right. \right. \\
&\quad \left. \left. - \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} f_Z(y) \right) d(y) + \frac{\partial^2}{\partial \lambda_j \partial \lambda_k} F_Z(T) + \right. \\
&\quad \left. + (1-p) \frac{\frac{\partial}{\partial \lambda_j} F_Z(T) \frac{\partial}{\partial \lambda_k} F_Z(T)}{(1 - (1-p)F_Z(T))} \right] \\
&= -(1-p_0)\mathbf{E}_{\Theta_0} \left[ \int_0^T \left( f_Z^{-1}(y) \frac{\partial}{\partial \lambda_j} f_Z(y) \frac{\partial}{\partial \lambda_k} f_Z(y) \right) d(y) + \right. \\
&\quad \left. + (1-p) \frac{\frac{\partial}{\partial \lambda_j} F_Z(T) \frac{\partial}{\partial \lambda_k} F_Z(T)}{(1 - (1-p)F_Z(T))} \right].
\end{aligned}$$

Z vyjádření  $\mathbf{J}_{jk}^{(i)}$  pro všechna  $j$  a  $k$  je patrné, že prvky Fisherovy informační matice nezávisí na  $i$ , tj.  $\mathbf{J}_{jk}^{(i)} = \mathbf{J}_{jk}$ . Z tohoto důvodu může být Fisherova

informační matice zapsána jako

$$\mathbf{J}_n(\Theta) = \sum_{i=1}^n \mathbf{J}^{(i)}(\Theta) = n \cdot \mathbf{J}(\Theta),$$

kde matice  $\mathbf{J}(\Theta)$  má prvky  $\mathbf{J}_{jk}$ . □

Prvky Fisherovy informační matice jsou konečné. Podmínky, za kterých tomu tak je, jsou uvedeny v [2].

### 4.3 Model s vysvětlujícími proměnnými

Velice důležitým úkolem při modelování rizika nesplacení je najít způsob, jak zahrnout informace o dlužníkovi do modelu pro dobu do selhání. Z tohoto důvodu rozšíříme nyní model vyložený výše o vysvětlující proměnné, ve kterých budou uloženy informace o konkrétním dlužníkovi mající vliv na dobu do selhání. Postup uvedený v této práci vychází z [2].

Nechť  $\mathbf{u}^{(i)} = (u_1^{(i)}, \dots, u_{k_1}^{(i)})^\top$  a  $\mathbf{v}^{(i)} = (v_1^{(i)}, \dots, v_{k_2}^{(i)})^\top$  představují pro  $i$ -té pozorování dvě skupiny proměnných, které ovlivňují parametry  $p$  a  $\lambda$ . Budeme předpokládat, že  $\mathbf{u}^{(i)}$  i  $\mathbf{v}^{(i)}$  jsou nezávislé na čase, ale naopak zřejmě závisí na  $i$ -tém dlužníkovi. Díky tomu může parametr  $p$  i rozdělení náhodné veličiny  $Z$  reprezentované distribuční funkcí  $F_Z$  závislou na parametru  $\lambda$  také na  $i$  záviset. Tuto závislost budeme dále značit  $p^{(i)}$ , resp.  $\lambda^{(i)}$ . Vztah mezi parametry  $p^{(i)}$ ,  $\lambda^{(i)}$  a proměnnými zavedeme pomocí funkcí  $\psi_u$  a  $\psi_v$  a to tak, že

$$\begin{aligned} p^{(i)} &= \psi_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)}) \\ \lambda^{(i)} &= \psi_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}), \end{aligned}$$

kde  $\boldsymbol{\gamma}$  je  $k_1$ -rozměrný a  $\boldsymbol{\delta}$  je  $k_2$ -rozměrný parametr. Parametry  $\boldsymbol{\gamma}$  a  $\boldsymbol{\delta}$  vyjadřují velikost vlivu jednotlivých proměnných na rozdělení  $F_Z$  a tak i na rozdělení doby do selhání. Budeme tedy nyní řešit úlohu, jak odhadnout parametry  $\boldsymbol{\gamma}$  a  $\boldsymbol{\delta}$ , takže v jsme v situaci, kdy  $\Theta = (\boldsymbol{\gamma}, \boldsymbol{\delta})$ . K tomu budeme předpokládat, že funkce  $\psi_u : \mathbb{R} \rightarrow \langle 0, 1 \rangle$  a  $\psi_v : \mathbb{R} \rightarrow \mathbb{R}^+$  jsou dvakrát spojitě diferencovatelné.

Nejprve sestavíme z vektorů vysvětlujících proměnných pro  $i = 1, \dots, n$  matice  $\mathbf{M}^{(i)}$  s rozměry  $k \times 2$ ,  $k = k_1 + k_2$ , které mají tvar

$$\mathbf{M}^{(i)} = \begin{pmatrix} \mathbf{u}^{(i)} & 0 \\ 0 & \mathbf{v}^{(i)} \end{pmatrix}.$$

Zavedení vysvětlujících proměnných do modelu si žádá pozorovat data v rozšířené podobě a to konkrétně  $(W_i, I_i, \mathbf{M}^{(i)})$ . Derivování logaritmické věrohodnostní funkce za účelem odhadu parametru  $\Theta$  vede k následující funkci

$$S = \sum_{i=1}^n \mathbf{M}^{(i)} \mathbf{s}^{(i)}(\Theta), \quad (4.24)$$

kde  $\mathbf{s}^{(i)}(\Theta)$  označuje vektor se složkami  $(s_1^{(i)}, s_2^{(i)})^\top$ . Kvůli zjednodušení značení dále většinou využíváme vztah  $p^{(i)} = \psi_u(\boldsymbol{\gamma} \cdot \mathbf{u}^{(i)})$ . Hodnoty složek vektoru  $\mathbf{s}^{(i)}(\Theta)$  určují následující rovnice

$$\begin{aligned} s_1^{(i)} &= -\frac{\psi'_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)})}{1 - p^{(i)}} \cdot \chi\{I_i = 1\} + \\ &\quad + \frac{\psi'_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)})F_Z(t_i)}{1 - (1 - \psi'_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)}))F_Z(t_i)} \cdot \chi\{I_i = 0\} \\ s_2^{(i)} &= \frac{\frac{\partial}{\partial \boldsymbol{\lambda}} f_Z(w_i)}{f_Z(w_i)} \psi'_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}) \cdot \chi\{I_i = 1\} - \\ &\quad - \frac{(1 - p^{(i)}) \frac{\partial}{\partial \boldsymbol{\lambda}} F_Z(t_i)}{1 - (1 - p^{(i)})F_Z(t_i)} \psi'_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}) \cdot \chi\{I_i = 0\} \end{aligned}$$

Matici druhých derivací logaritmické věrohodnostní funkce označíme  $\mathbf{D}_n(\Theta)$ ,

$$\mathbf{D}_n(\Theta) = -\frac{\partial^2}{\partial \Theta^2} \ell = \sum_{i=1}^n \left( \mathbf{M}^{(i)} \mathbf{D}^{(i)}(\Theta) \right)^\top \mathbf{M}^{(i)},$$

kde  $\mathbf{D}^{(i)}(\Theta)$  je matice o rozměrech  $2 \times 2$  s prvky

$$\begin{aligned} \mathbf{D}_{11}^{(i)} &= \frac{(1 - p^{(i)})\psi''_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)}) + (\psi'_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)}))^2}{(1 - p^{(i)})^2} \cdot \chi\{I_i = 1\} - \\ &\quad - \frac{1}{(1 - (1 - p^{(i)})F_Z(t_i))^2} \left( \psi''_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)})F_Z(t_i) - \right. \\ &\quad \left. - [(1 - p^{(i)})\psi''_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)}) + (\psi'_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)}))^2]F_Z^2(t_i) \right) \cdot \chi\{I_i = 0\} \end{aligned} \quad (4.25)$$

$$\mathbf{D}_{12}^{(i)} = \mathbf{D}_{21}^{(i)} = -\frac{\psi'_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)})\psi'_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)})\frac{\partial}{\partial \boldsymbol{\lambda}} F_Z(t_i)}{(1 - (1 - p^{(i)})F_Z(t_i))^2} \cdot \chi\{I_i = 0\} \quad (4.26)$$

$$\begin{aligned} \mathbf{D}_{22}^{(i)} &= -\left[ \frac{\frac{\partial^2}{\partial \boldsymbol{\lambda}^2} f_Z(t_i) \cdot f_Z(t_i) - \frac{\partial}{\partial \boldsymbol{\lambda}} f_Z^2(t_i)}{f_Z^2(t_i)} (\psi'_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}))^2 + \right. \\ &\quad \left. + \frac{\frac{\partial}{\partial \boldsymbol{\lambda}} f_Z(t_i)}{f_Z(t_i)} \psi''_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}) \right] \cdot \chi\{I_i = 1\} + \\ &\quad + (1 - p^{(i)}) \left[ \frac{\frac{\partial^2}{\partial \boldsymbol{\lambda}^2} F_Z(t_i) (\psi'_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}))^2 + \frac{\partial}{\partial \boldsymbol{\lambda}} F_Z(t_i) \psi''_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)})}{1 - (1 - p^{(i)})F_Z(t_i)} + \right. \\ &\quad \left. + \frac{(1 - p^{(i)}) \frac{\partial}{\partial \boldsymbol{\lambda}} F_Z^2(t_i) (\psi'_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}))^2}{(1 - (1 - p^{(i)})F_Z(t_i))^2} \right] \cdot \chi\{I_i = 0\}. \end{aligned} \quad (4.27)$$

Fisherova informační matice je pak opět střední hodnotou matice  $\mathbf{D}_n(\boldsymbol{\Theta})$  (viz. [2])

$$\mathbf{J}_n(\boldsymbol{\Theta}) = \mathbb{E}[\mathbf{D}_n(\boldsymbol{\Theta})].$$

Za předpokladu konečných středních hodnot prvků matice  $\mathbf{D}_n(\boldsymbol{\Theta})$  lze určit hodnoty prvků Fisherovy informační matice pro známé  $\boldsymbol{\Theta}_0$  z následujících rovnic (viz. [2])

$$\mathbf{J}_{11}^{(i)} = \frac{(\psi'_u)^2(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)})}{(1 - p^{(i)})} \mathbb{E}_{\boldsymbol{\Theta}_0} \left[ \frac{F_Z(T)}{1 - (1 - p^{(i)})F_Z(T)} \right], \quad (4.28)$$

$$\mathbf{J}_{12}^{(i)} = \mathbf{J}_{21}^{(i)} = -\psi'_u(\boldsymbol{\gamma}^\top \cdot \mathbf{u}^{(i)})\psi'_v(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}) \mathbb{E}_{\boldsymbol{\Theta}_0} \left[ \frac{\frac{\partial}{\partial \boldsymbol{\lambda}} F_Z(T)}{1 - (1 - p^{(i)})F_Z(T)} \right], \quad (4.29)$$

$$\begin{aligned} \mathbf{J}_{22}^{(i)} = & (1 - p^{(i)})(\psi'_v)^2(\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}) \mathbb{E}_{\boldsymbol{\Theta}_0} \left[ \int_0^T \frac{\frac{\partial}{\partial \boldsymbol{\lambda}} f_Z^2(y)}{f_Z(y)} d(y) + \right. \\ & \left. + \frac{(1 - p^{(i)}) \frac{\partial}{\partial \boldsymbol{\lambda}} F_Z(T)}{1 - (1 - p^{(i)})F_Z(T)} \right]. \end{aligned} \quad (4.30)$$

V případě, že jsou splněny podmínky regularity, maximálně věrohodný odhad  $\widehat{\boldsymbol{\Theta}}$  parametru  $\boldsymbol{\Theta}_0$  existuje. Navíc pro  $n \rightarrow \infty$

$$\widehat{\boldsymbol{\Theta}} \rightarrow \boldsymbol{\Theta}_0 \quad \text{s.j.}$$

a

$$\mathbf{J}_n^{1/2}(\widehat{\boldsymbol{\Theta}})(\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0) \xrightarrow{\mathcal{D}} \mathbf{N}(0, \mathbf{I}_k),$$

kde  $\mathbf{I}_k$  označuje  $k$ -rozměrnou jednotkovou matici,  $k = k_1 + k_2$  je sloupcový rozměr matice vysvětlujících proměnných  $\mathbf{M}^{(i)}$ ,  $\forall i$ . Důkazy existence maximálně věrohodného odhadu a obou konvergencí jsou uvedeny v [2].

# Kapitola 5

## Aplikace na reálná data

### 5.1 Statistická analýza v Markovových řetězcích

Nyní aplikujeme  $\chi^2$ -test dobré shody a test poměrem věrohodností na reálná data. Využijeme hodnoty publikované ratingovou agenturou Standard & Poor's (viz. [13]) z analýzy zveřejněné v roce 2004 pro obecní dluhopisy. O ratingových agenturách a formě, v jaké data publikují, jsme se podrobněji zmínili v kapitole 2. V námi použité studii jsou konkrétně uvedeny dva typy matic vztahující se k období od roku 1986 do roku 2003 včetně. Agentura Standard & Poor's zařazuje dluhopisy do devíti různých kategorií od nejlepší AAA přes AA, A, BBB, BB, B, CCC až ke stavu selhání D a včetně stavu NR pro ratingy, které nebyly ke konci roku k dispozici. Podrobněji jsme význam jednotlivých kategorií popsali v kapitole 2. Každý typ matic nyní popíšeme podrobněji.

Prvním typem jsou *jednoleté matice četností přechodu*. Prvky těchto matic jsme dříve označili jako  $n_{ij}$ , kde  $i = AAA, \dots, CCC$  a  $j = AAA, \dots, NR$ . Množinu stavů, kterou probíhá index  $i$ , označíme  $S_i$  a podobně množinu stavů indexu  $j$  jako  $S_j$ , takže máme

$$\begin{aligned} S_i &= \{AAA, \dots, CCC\} \\ S_j &= \{AAA, \dots, NR\}. \end{aligned}$$

Prvky těchto matic vyjadřují počet přechodů z kategorie  $i$  na začátku roku do kategorie  $j$  na konci. V analýze agentury Standard & Poor's je uvedeno osmnáct takových jednoletých matic pro každý rok ze sledovaného období. Máme tak

$$\mathbf{N}^{(t)} = (n_{ij}^{(t)}) \quad i \in S_i, j \in S_j, t = 1, \dots, T,$$

kde v tomto konkrétním případě  $T = 18$ . Maximálně věrohodné odhady pravděpodobností přechodu  $\hat{p}_{ij}^{(t)}$  získáme jako

$$\hat{\mathbf{P}}^{(t)} = (\hat{p}_{ij}^{(t)}) = \left( \frac{n_{ij}^{(t)}}{n_{i\cdot}^{(t)}} \right) \quad i \in S_i, j \in S_j, t = 1, \dots, T,$$

kde  $n_{i\cdot}^{(t)} = \sum_j n_{ij}^{(t)}$ .

Druhý typ matic představují *matice pravděpodobností přechodu  $n$ -tého řádu*, přičemž v tomto konkrétním případě  $n = 1, \dots, 18$ . Tyto matice a jejich prvky značíme jako  $\mathbf{P}(n) = (p_{ij}(n))$ , kde indexy  $i$  a  $j$  probíhají stejné množiny jako u matic četností.

Oba typy matic neuvažují pohyby mezi ratingovými kategoriemi v průběhu období, ke kterému se vztahují. Zachycují pouze stav na počátku a na konci časového intervalu. Tato data jsme postupně použili pro  $\chi^2$ -test dobré shody a pak pro test poměrem věrohodností.

### 5.1.1 $\chi^2$ -test dobré shody

Pomocí  $\chi^2$ -testu dobré shody testujeme hypotézu, zda se daná matice pravděpodobností přechodu shoduje s maticí pozorovaných četností. Na základě výše popsaných dat jsme tak testovali nulovou hypotézu o shodnosti matice pravděpodobností přechodu prvního řádu  $\mathbf{P}(1)$  s maticí četností  $\mathbf{N}^{(t)}$  postupně pro všechna  $t = 1, \dots, T$ , kde  $T = 18$ . V případě těchto konkrétních dat se matice četností pro  $t = 1$  vztahuje k roku 1986, pro  $t = 2$  k roku 1987 atd. Hodnoty testové statistiky spolu s  $p$ -hodnotou jsou uvedeny v následující tabulce.

$t = 1$	$\chi^2(27) = 93,2141$	$p$ -hodnota =	$3.25548 \times 10^{-9}$
$t = 2$	$\chi^2(27) = 72,0649$	$p$ -hodnota =	$5.66499 \times 10^{-6}$
$t = 3$	$\chi^2(27) = 59,8792$	$p$ -hodnota =	0.000273581
$t = 4$	$\chi^2(27) = 217,079$	$p$ -hodnota =	0.
$t = 5$	$\chi^2(27) = 186,769$	$p$ -hodnota =	0.
$t = 6$	$\chi^2(27) = 277,94$	$p$ -hodnota =	0.
$t = 7$	$\chi^2(27) = 117,287$	$p$ -hodnota =	$3.1819 \times 10^{-13}$
$t = 8$	$\chi^2(27) = 99,4404$	$p$ -hodnota =	$3.17997 \times 10^{-10}$
$t = 9$	$\chi^2(27) = 77,7829$	$p$ -hodnota =	$8.1487 \times 10^{-7}$
$t = 10$	$\chi^2(27) = 140,476$	$p$ -hodnota =	0.
$t = 11$	$\chi^2(27) = 39,4727$	$p$ -hodnota =	0.0573515
$t = 12$	$\chi^2(27) = 63,3444$	$p$ -hodnota =	0.0000945529
$t = 13$	$\chi^2(27) = 99,026$	$p$ -hodnota =	$3.71793 \times 10^{-10}$
$t = 14$	$\chi^2(27) = 164,483$	$p$ -hodnota =	0.
$t = 15$	$\chi^2(27) = 427,035$	$p$ -hodnota =	0.
$t = 16$	$\chi^2(27) = 861,272$	$p$ -hodnota =	0.
$t = 17$	$\chi^2(27) = 62,9982$	$p$ -hodnota =	0.000105305
$t = 18$	$\chi^2(27) = 87,703$	$p$ -hodnota =	$2.44269 \times 10^{-8}$

Na základě  $p$ -hodnoty dojdeme k závěru, že pro všechny matice četností s výjimkou jedné pro  $t = 11$  nulovou hypotézu na hladině 5% pro tato data zamítáme. Matice četností se tedy až na jeden případ neshodují s maticí pravděpodobností přechodu prvního řádu. V osmnáctiletém sledovaném období tak

matice ve většině případů vykazují statisticky významné odchylky od matice prvního řádu. Tento výsledek poukazuje na to, že markovský řetězec, kterým přechody mezi ratingovými kategoriemi popisujeme, nebude v reálu často homogenní. Pro homogenní Markovův řetězec platí, že pravděpodobnosti přechodu pro časové intervaly shodné délky jsou stejné pro různé okamžiky počátku časového intervalu. Tedy v našem případě jednoletých matic četností by pro maximálně věrohodné odhady pravděpodobností přechodu  $\hat{p}_{ij}^{(t)}$  postupně pro všechna  $t$  a pravděpodobnosti přechodu prvního řádu  $p_{ij}(1)$ ,  $i \in S_i$  a  $j \in S_j$ , mělo platit

$$\begin{aligned}\hat{p}_{ij}^{(1)} &\approx p_{ij}(1) \\ \hat{p}_{ij}^{(2)} &\approx p_{ij}(1) \\ &\vdots \\ \hat{p}_{ij}^{(T)} &\approx p_{ij}(1)\end{aligned}$$

Na základě  $\chi^2$ -testu dobré shody však tuto aproximaci nemůžeme potvrdit. Předpokládat, že Markovův řetězec odpovídající použitým hodnotám bude časově homogenní a bude splněno

$$\hat{p}_{ij}^{(1)} \doteq \hat{p}_{ij}^{(2)} \doteq \dots \doteq \hat{p}_{ij}^{(T)}, \quad i \in S_i, j \in S_j,$$

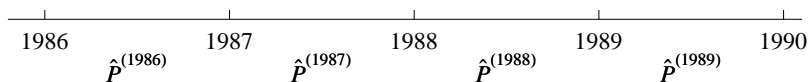
je pro námi použitá data nereálné. Vzhledem k délce časového intervalu a množství dat, které jsme měli k dispozici od agentury Standard & Poor's, můžeme usuzovat, že předpoklad homogenního Markovova řetězce popisujícího přechody mezi ratingovými kategoriemi nebude splněn ani v jiných letech.

### 5.1.2 Test poměrem věrohodností

Test poměrem věrohodností jsme popsali v kapitole 3.3.1. Umožňuje nám testovat hypotézu, zda jsou matice maximálně věrohodných odhadů pravděpodobností přechodu  $\hat{\mathbf{P}}^{(t)}$  statisticky stejné jako matice  $\mathbf{P}(n)$  postupně pro každé  $n = 1, \dots, T$  a k němu  $\Delta \leq t \leq \Delta + n - 1$ , kde  $\Delta = 1, \dots, T - n$ . V případě těchto konkrétních dat máme  $T = 18$  a pro  $t = 1$  máme matici četností vztahující se k roku 1986, pro  $t = 2$  matici četností pro rok 1987 atd. Na obrázku 5.1 je znázorněn vztah mezi maticemi  $\hat{\mathbf{P}}^{(t)}$  a  $\mathbf{P}(n)$  pro  $n = 4$  a  $\Delta = 1986$ , takže  $t = 1986 \dots, 1989$ .



$P(4)$



Obr. 5.1: Příklad vztahu  $\mathbf{P}(n)$  a  $\widehat{\mathbf{P}}^{(t)}$  pro  $n = 4$  a  $t = 1986, \dots, 1989$

Testujeme nulovou hypotézu, že pro dané  $n$  jsou matice maximálně věrohodných odhadů pravděpodobností přechodu statisticky stejné jako matice  $\mathbf{P}(n)$  proti alternativě, že nejsou. Výsledky testu jednoznačně vedou k zamítnutí nulové hypotézy na základě dat agentury Standard & Poor's. Zamítáme tedy hypotézu, že by Markovův řetězec popisující přechody mezi ratingovými kategoriemi v čase byl homogenní. Konkrétní hodnoty testové statistiky, stupně volnosti a  $p$ -hodnota jsou uvedeny v následujících tabulkách pro každé  $n$  a k němu příslušná  $\Delta$ .

$\mathbf{n} = 1$			
$\chi^2(56)$	=	584,93	$p$ -hodnota = 0.
$\chi^2(56)$	=	589,005	$p$ -hodnota = 0.
$\chi^2(56)$	=	603,927	$p$ -hodnota = 0.
$\chi^2(56)$	=	656,385	$p$ -hodnota = 0.
$\chi^2(56)$	=	738,333	$p$ -hodnota = 0.
$\chi^2(56)$	=	818,973	$p$ -hodnota = 0.
$\chi^2(56)$	=	759,164	$p$ -hodnota = 0.
$\chi^2(56)$	=	771,499	$p$ -hodnota = 0.
$\chi^2(56)$	=	742,051	$p$ -hodnota = 0.
$\chi^2(56)$	=	764,21	$p$ -hodnota = 0.
$\chi^2(56)$	=	667,738	$p$ -hodnota = 0.
$\chi^2(56)$	=	727,935	$p$ -hodnota = 0.
$\chi^2(56)$	=	765,039	$p$ -hodnota = 0.
$\chi^2(56)$	=	879,951	$p$ -hodnota = 0.
$\chi^2(56)$	=	1112,13	$p$ -hodnota = 0.
$\chi^2(56)$	=	1527,24	$p$ -hodnota = 0.
$\chi^2(56)$	=	1090,1	$p$ -hodnota = 0.
$\chi^2(56)$	=	1148,69	$p$ -hodnota = 0.

<b>n = 2</b>			
$\chi^2(112)$	=	2483,24	$p$ -hodnota = 0.
$\chi^2(112)$	=	2557,92	$p$ -hodnota = 0.
$\chi^2(112)$	=	2719,15	$p$ -hodnota = 0.
$\chi^2(112)$	=	2848,24	$p$ -hodnota = 0.
$\chi^2(112)$	=	3039,11	$p$ -hodnota = 0.
$\chi^2(112)$	=	3193	$p$ -hodnota = 0.
$\chi^2(112)$	=	3284,09	$p$ -hodnota = 0.
$\chi^2(112)$	=	3316,21	$p$ -hodnota = 0.
$\chi^2(112)$	=	3175,12	$p$ -hodnota = 0.
$\chi^2(112)$	=	2983,36	$p$ -hodnota = 0.
$\chi^2(112)$	=	3028,8	$p$ -hodnota = 0.
$\chi^2(112)$	=	3222,05	$p$ -hodnota = 0.
$\chi^2(112)$	=	3411,54	$p$ -hodnota = 0.
$\chi^2(112)$	=	3717,32	$p$ -hodnota = 0.
$\chi^2(112)$	=	4227,01	$p$ -hodnota = 0.
$\chi^2(112)$	=	4642,9	$p$ -hodnota = 0.
$\chi^2(112)$	=	4760,33	$p$ -hodnota = 0.

<b>n = 3</b>			
$\chi^2(168)$	=	6091,39	$p$ -hodnota = 0.
$\chi^2(168)$	=	6323,92	$p$ -hodnota = 0.
$\chi^2(168)$	=	6604,4	$p$ -hodnota = 0.
$\chi^2(168)$	=	6965,56	$p$ -hodnota = 0.
$\chi^2(168)$	=	7292	$p$ -hodnota = 0.
$\chi^2(168)$	=	7695,97	$p$ -hodnota = 0.
$\chi^2(168)$	=	7827,33	$p$ -hodnota = 0.
$\chi^2(168)$	=	7721,8	$p$ -hodnota = 0.
$\chi^2(168)$	=	7372,23	$p$ -hodnota = 0.
$\chi^2(168)$	=	7297,33	$p$ -hodnota = 0.
$\chi^2(168)$	=	7450,55	$p$ -hodnota = 0.
$\chi^2(168)$	=	7932,63	$p$ -hodnota = 0.
$\chi^2(168)$	=	8389,42	$p$ -hodnota = 0.
$\chi^2(168)$	=	9181,14	$p$ -hodnota = 0.
$\chi^2(168)$	=	10152,7	$p$ -hodnota = 0.
$\chi^2(168)$	=	11017,6	$p$ -hodnota = 0.

<b>n = 4</b>			
$\chi^2(224)$	=	11610,5	$p$ -hodnota = 0.
$\chi^2(224)$	=	11971,4	$p$ -hodnota = 0.
$\chi^2(224)$	=	12583,7	$p$ -hodnota = 0.
$\chi^2(224)$	=	13167,4	$p$ -hodnota = 0.
$\chi^2(224)$	=	13770,3	$p$ -hodnota = 0.
$\chi^2(224)$	=	14248,7	$p$ -hodnota = 0.
$\chi^2(224)$	=	14244,5	$p$ -hodnota = 0.
$\chi^2(224)$	=	13972,6	$p$ -hodnota = 0.
$\chi^2(224)$	=	13793,2	$p$ -hodnota = 0.
$\chi^2(224)$	=	13783,3	$p$ -hodnota = 0.
$\chi^2(224)$	=	14299,3	$p$ -hodnota = 0.
$\chi^2(224)$	=	15198,5	$p$ -hodnota = 0.
$\chi^2(224)$	=	16349,6	$p$ -hodnota = 0.
$\chi^2(224)$	=	18021,3	$p$ -hodnota = 0.
$\chi^2(224)$	=	19505	$p$ -hodnota = 0.

<b>n = 5</b>			
$\chi^2(280)$	=	18903,9	$p$ -hodnota = 0.
$\chi^2(280)$	=	19655,6	$p$ -hodnota = 0.
$\chi^2(280)$	=	20595,6	$p$ -hodnota = 0.
$\chi^2(280)$	=	21542	$p$ -hodnota = 0.
$\chi^2(280)$	=	22205,2	$p$ -hodnota = 0.
$\chi^2(280)$	=	22538	$p$ -hodnota = 0.
$\chi^2(280)$	=	22424,7	$p$ -hodnota = 0.
$\chi^2(280)$	=	22411,6	$p$ -hodnota = 0.
$\chi^2(280)$	=	22309,4	$p$ -hodnota = 0.
$\chi^2(280)$	=	22702,9	$p$ -hodnota = 0.
$\chi^2(280)$	=	23717,4	$p$ -hodnota = 0.
$\chi^2(280)$	=	25522,2	$p$ -hodnota = 0.
$\chi^2(280)$	=	27804,1	$p$ -hodnota = 0.
$\chi^2(280)$	=	30130,7	$p$ -hodnota = 0.

<b>n = 6</b>			
$\chi^2(336)$	=	28448,8	$p$ -hodnota = 0.
$\chi^2(336)$	=	29608,1	$p$ -hodnota = 0.
$\chi^2(336)$	=	30968,4	$p$ -hodnota = 0.
$\chi^2(336)$	=	31998,2	$p$ -hodnota = 0.
$\chi^2(336)$	=	32481,5	$p$ -hodnota = 0.
$\chi^2(336)$	=	32730,4	$p$ -hodnota = 0.
$\chi^2(336)$	=	32981,4	$p$ -hodnota = 0.
$\chi^2(336)$	=	33103,1	$p$ -hodnota = 0.
$\chi^2(336)$	=	33499,1	$p$ -hodnota = 0.
$\chi^2(336)$	=	34431,7	$p$ -hodnota = 0.
$\chi^2(336)$	=	36486,1	$p$ -hodnota = 0.
$\chi^2(336)$	=	39656,2	$p$ -hodnota = 0.
$\chi^2(336)$	=	42687,4	$p$ -hodnota = 0.

<b>n = 7</b>			
$\chi^2(392)$	=	40245,7	$p$ -hodnota = 0.
$\chi^2(392)$	=	41874,6	$p$ -hodnota = 0.
$\chi^2(392)$	=	43333,6	$p$ -hodnota = 0.
$\chi^2(392)$	=	44171,4	$p$ -hodnota = 0.
$\chi^2(392)$	=	44578,2	$p$ -hodnota = 0.
$\chi^2(392)$	=	45267,9	$p$ -hodnota = 0.
$\chi^2(392)$	=	45737,2	$p$ -hodnota = 0.
$\chi^2(392)$	=	46503,3	$p$ -hodnota = 0.
$\chi^2(392)$	=	47519	$p$ -hodnota = 0.
$\chi^2(392)$	=	49598,7	$p$ -hodnota = 0.
$\chi^2(392)$	=	53166,1	$p$ -hodnota = 0.
$\chi^2(392)$	=	57180,9	$p$ -hodnota = 0.

<b>n = 8</b>			
$\chi^2(448)$	=	54586	$p$ -hodnota = 0.
$\chi^2(448)$	=	56322,3	$p$ -hodnota = 0.
$\chi^2(448)$	=	57589	$p$ -hodnota = 0.
$\chi^2(448)$	=	58364,2	$p$ -hodnota = 0.
$\chi^2(448)$	=	59288,9	$p$ -hodnota = 0.
$\chi^2(448)$	=	60280,6	$p$ -hodnota = 0.
$\chi^2(448)$	=	61538	$p$ -hodnota = 0.
$\chi^2(448)$	=	63058,1	$p$ -hodnota = 0.
$\chi^2(448)$	=	65371,1	$p$ -hodnota = 0.
$\chi^2(448)$	=	69104,2	$p$ -hodnota = 0.
$\chi^2(448)$	=	73572,6	$p$ -hodnota = 0.

<b>n = 9</b>			
$\chi^2(504)$	=	70733,4	$p$ -hodnota = 0.
$\chi^2(504)$	=	72237,7	$p$ -hodnota = 0.
$\chi^2(504)$	=	73429,3	$p$ -hodnota = 0.
$\chi^2(504)$	=	74807,2	$p$ -hodnota = 0.
$\chi^2(504)$	=	76056,5	$p$ -hodnota = 0.
$\chi^2(504)$	=	77995,3	$p$ -hodnota = 0.
$\chi^2(504)$	=	80208,1	$p$ -hodnota = 0.
$\chi^2(504)$	=	83226,9	$p$ -hodnota = 0.
$\chi^2(504)$	=	87394,5	$p$ -hodnota = 0.
$\chi^2(504)$	=	92044,9	$p$ -hodnota = 0.

<b>n = 10</b>			
$\chi^2(560)$	=	87442	$p$ -hodnota = 0.
$\chi^2(560)$	=	88865,8	$p$ -hodnota = 0.
$\chi^2(560)$	=	90695,4	$p$ -hodnota = 0.
$\chi^2(560)$	=	92449,3	$p$ -hodnota = 0.
$\chi^2(560)$	=	94783,5	$p$ -hodnota = 0.
$\chi^2(560)$	=	97793,8	$p$ -hodnota = 0.
$\chi^2(560)$	=	101701	$p$ -hodnota = 0.
$\chi^2(560)$	=	106624	$p$ -hodnota = 0.
$\chi^2(560)$	=	111697	$p$ -hodnota = 0.

<b>n = 11</b>			
$\chi^2(616)$	=	104336	$p$ -hodnota = 0.
$\chi^2(616)$	=	106428	$p$ -hodnota = 0.
$\chi^2(616)$	=	108650	$p$ -hodnota = 0.
$\chi^2(616)$	=	111572	$p$ -hodnota = 0.
$\chi^2(616)$	=	115054	$p$ -hodnota = 0.
$\chi^2(616)$	=	119864	$p$ -hodnota = 0.
$\chi^2(616)$	=	125776	$p$ -hodnota = 0.
$\chi^2(616)$	=	131593	$p$ -hodnota = 0.

<b>n = 12</b>			
$\chi^2(672)$	=	121521	$p$ -hodnota = 0.
$\chi^2(672)$	=	124003	$p$ -hodnota = 0.
$\chi^2(672)$	=	127423	$p$ -hodnota = 0.
$\chi^2(672)$	=	131506	$p$ -hodnota = 0.
$\chi^2(672)$	=	136846	$p$ -hodnota = 0.
$\chi^2(672)$	=	143704	$p$ -hodnota = 0.
$\chi^2(672)$	=	150511	$p$ -hodnota = 0.

<b>n = 13</b>			
$\chi^2(728)$	=	140014	$p$ -hodnota = 0.
$\chi^2(728)$	=	143766	$p$ -hodnota = 0.
$\chi^2(728)$	=	148407	$p$ -hodnota = 0.
$\chi^2(728)$	=	154451	$p$ -hodnota = 0.
$\chi^2(728)$	=	161947	$p$ -hodnota = 0.
$\chi^2(728)$	=	169737	$p$ -hodnota = 0.

<b>n = 14</b>			
$\chi^2(784)$	=	160521	$p$ -hodnota = 0.
$\chi^2(784)$	=	165523	$p$ -hodnota = 0.
$\chi^2(784)$	=	172211	$p$ -hodnota = 0.
$\chi^2(784)$	=	180474	$p$ -hodnota = 0.
$\chi^2(784)$	=	188915	$p$ -hodnota = 0.

<b>n = 15</b>			
$\chi^2(840)$	=	180663	$p$ -hodnota = 0.
$\chi^2(840)$	=	187804	$p$ -hodnota = 0.
$\chi^2(840)$	=	196771	$p$ -hodnota = 0.
$\chi^2(840)$	=	205969	$p$ -hodnota = 0.

<b>n = 16</b>			
$\chi^2(896)$	=	210147	$p$ -hodnota = 0.
$\chi^2(896)$	=	219868	$p$ -hodnota = 0.
$\chi^2(896)$	=	230044	$p$ -hodnota = 0.

<b>n = 17</b>			
$\chi^2(952)$	=	245380	$p$ -hodnota = 0.
$\chi^2(952)$	=	256419	$p$ -hodnota = 0.

<b>n = 18</b>			
$\chi^2(1008)$	=	285396	$p$ -hodnota = 0.

Vzhledem k tomu, že jsme pro testování použili data za osmnáctileté období, můžeme usuzovat, že předpoklad časové homogenity nebude v praxi obvykle splněn a to ani pro data jiných agentur nebo z interního hodnocení. Je tedy nutné k tomuto faktu přihlédnout v momentě, kdy se společnost rozhodne na základě úvěrové migrace modelovat vývoj pravděpodobností selhání do budoucnosti. Nesporná výhoda tohoto přístupu spočívá v jeho jednoduchosti. Jak ale dokládá námi provedený test na reálných datech, tato jednoduchost je vykoupena značně nereálným předpokladem, který může vést k nespolehlivým výsledkům v modelování budoucího vývoje.

## 5.2 Modelování úvěrového rizika pomocí analýzy přežití

Když se banky rozhodují, zda klientovi poskytnou úvěr či nikoliv, jsou postaveny před úkol, jak určit riziko selhání spojené s daným klientem. Jednu z možností jak k tomuto problému přistupovat nabízí model popsáný v kapitole 4.2 resp. 4.3. Způsob, jakým je možné model použít na reálných datech, ukážeme v následující části.

Máme k dispozici data z archivu Univerzity Mnichov (viz. [6]). Jedná se o soubor tisíce různých dlužníků, kde ke každému z nich máme evidovány různé charakteristiky. Pro účely našeho modelu jsme vybrali veličiny, u kterých je možné předpokládat, že mají významný vliv na selhání klienta. Využijeme veličiny popisující, zda klient úvěr splatil nebo ne, jak dlouho úvěr trval, hodnotu úvěru, předchozí platební morálku klienta, procento hodnoty splátky z disponibilního příjmu klienta a věk klienta. Pro potřeby modelu musíme data převést do matematické podoby.

Začneme s veličinou určující, zda klient úvěr splatil nebo ne. Tento jev můžeme vyjádřit pro dlužníky označené indexem  $i = 1, \dots, n$  pomocí indikátoru  $\chi_i(t)$ . V tomto konkrétním případě  $n = 1000$ . Indikátor  $\chi_i(t)$  bude nabývat pouze dvou hodnot v závislosti na tom, zda klient úvěr splatil nebo ne, tedy

$$\chi_i(t) = \begin{cases} 0 & \text{nedošlo k selhání dlužníka } i \text{ v čase } t \\ 1 & \text{klient v } t \text{ selhal.} \end{cases}$$

Je-li  $\chi_i = 0$ , jedná se o pozorování, které bylo z výběru vyjmuto dříve a je tedy cenzorované. Další parametr, který určuje délku trvání úvěru, tak má vlastně dvojitý význam. Pokud  $\chi_i = 0$ , vyjadřuje hodnotu časového cenzoru  $T_i$ . V druhém případě pak vyjadřuje skutečnou délku trvání dluhu  $X_i$ . Rozdělení doby do selhání ale neznáme, nemůžeme tak určit funkci spolehlivosti  $S(t)$ , která je doplnkem distribuční funkce do jedné,  $S(t) = P(T > t)$ . Využijeme k tomu *Kaplan-Meierův odhad funkce spolehlivosti*. Nejdříve musíme však data uspořádat. Máme k dispozici dvojici pozorování

$$(W_1, I_1), \dots, (W_n, I_n),$$

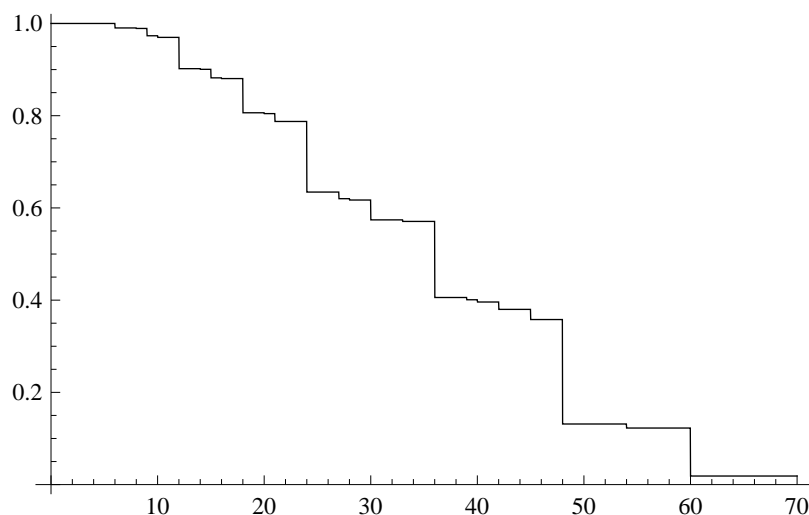
kde  $W_i = \min(X_i, T_i)$  a  $I_i = \chi\{X_i \leq T_i\}$ . Platí

$$\begin{aligned} \chi_i(t) &= 0 & W_i = T_i, I_i = 0 \\ \chi_i(t) &= 1 & W_i = X_i, I_i = 1. \end{aligned}$$

Dvojice  $(W_1, I_1), \dots, (W_n, I_n)$  uspořádáme lexikograficky podle hodnoty  $W_i$ . Výsledkem je posloupnost  $(W_{(1)}, I_{(1)}), \dots, (W_{(n)}, I_{(n)})$ , ve které  $I_{(i)}$  označuje veličinu příslušnou k  $W_{(i)}$ . V případě, že nedošlo ke shodám mezi  $W_i$ ,  $\forall i$ , platí  $W_{(1)} < W_{(2)} < \dots < W_{(n)}$ . Pokud se v posloupnosti vyskytly shody, můžeme pořadí určit pomocí druhé veličiny. Jsou-li i ty shodné, na pořadí takových veličin nezáleží. Kaplan-Meierův odhad funkce spolehlivosti (viz. [7]) pak můžeme uvést ve tvaru

$$\widehat{S}(t) = \prod_{i: W_{(i)} \leq t} \left( \frac{n-i}{n-i+1} \right)^{I_{(i)}} \quad t \leq W_{(n)}.$$

Obrázek 5.2 znázorňuje Kaplan-Meierův odhad funkce spolehlivosti v závislosti na době přežití  $t$  pro naše konkrétní data.



Obr. 5.2: Kaplan-Meierův odhad funkce spolehlivosti

Další veličiny, které charakterizují jednotlivé dlužníky, matematicky interpretujeme následujícím způsobem:

- (i)  $v_1^{(i)} = 0$  pokud hodnota úvěru nepřesáhla 2500 DM a  $v_1^{(i)} = 1$  pokud si dlužník půjčuje více než 2500 DM;
- (ii)  $v_2^{(i)} = 0$  pokud klient neměl dříve potíže se splácením a  $v_2^{(i)} = 1$  pokud klient měl dříve potíže se splácením nebo jeho dřívější platební morálku neznáme;
- (iii)  $v_3^{(i)} = 0$  pokud splátka půjčky je méně než 25% měsíčního příjmu a  $v_3^{(i)} = 1$  pokud splátka úvěru je větší nebo rovna čtvrtině měsíčního příjmu;
- (iv)  $v_4^{(i)} = 0$  pro dlužníka ve věku menším než 40 let a  $v_4^{(i)} = 1$  pro dlužníka staršího než 40 let.

Vektor všech vysvětlujících veličin budeme uvažovat ve tvaru

$$\mathbf{v}^{(i)} = (1, v_1^{(i)}, v_2^{(i)}, v_3^{(i)}, v_4^{(i)})^\top.$$

Pro jednoduchost budeme předpokládat, že parametry  $p^{(i)}$  nezávisí na žádných dalších vysvětlujících proměnných. Můžeme tedy pro  $i = 1, \dots, n$  psát  $p^{(i)} = p$ . Parametry  $\boldsymbol{\lambda}^{(i)}$  závisí na vysvětlujících proměnných  $\mathbf{v}^{(i)}$  pomocí funkce  $\psi_v$ . Budeme předpokládat, že funkce  $\psi_v$  je tvaru

$$\psi_v(x) = \exp(-x).$$

Za  $x$  pak dosadíme lineární kombinaci vektoru vysvětlujících proměnných  $\mathbf{v}^{(i)}$  a vektoru odhadovaných parametrů  $\boldsymbol{\delta}$ . Složky vektoru odhadovaných parametrů



$\boldsymbol{\delta} = (\delta_0, \delta_1, \delta_2, \delta_3, \delta_4)$  vyjadřují vliv jednotlivých vysvětlujících proměnných. Máme pak

$$\boldsymbol{\lambda}^{(i)} = \exp(-\boldsymbol{\delta}^\top \cdot \mathbf{v}^{(i)}).$$

Rozdělení náhodné veličiny  $Z$  budeme předpokládat exponenciální, takže její distribuční funkce má tvar  $F_Z(t) = 1 - e^{-\boldsymbol{\lambda}^{(i)} t}$ .

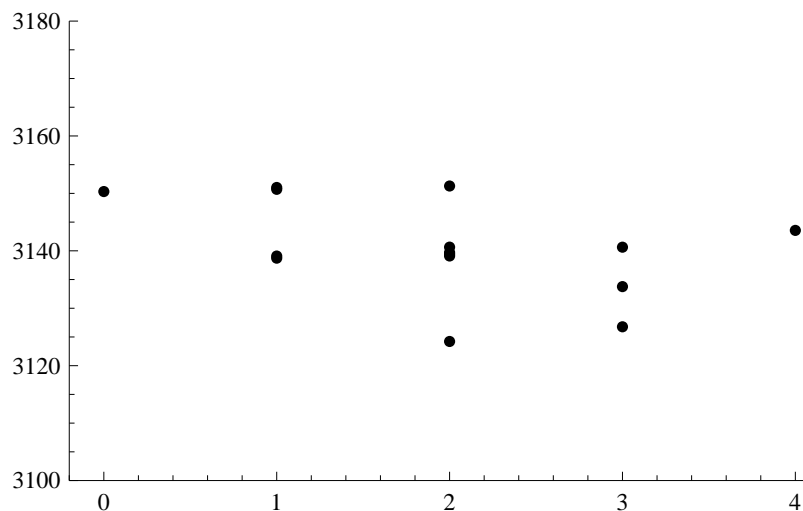
Některé z námi zvolených vysvětlujících proměnných mohou mít v modelu větší vliv než jiné, které naopak mohou být zbytečné. K tomu, abychom rozhodli, kolik vysvětlujících proměnných a které do modelu zahrneme, využijeme *Akaikeovo informační kritérium*. Mějme  $m$  modelů  $M_1, \dots, M_m$  takových, že model  $M_j$  má  $k_j$  parametrů,  $\boldsymbol{\Theta}_j = (\Theta_{j1}, \dots, \Theta_{jk_j})$ , a věrohodnostní funkci  $L_j(\boldsymbol{\Theta})$ . Na základě Akaikeova kritéria  $AIC(M_j)$  vybereme model, který minimalizuje

$$AIC(M_j) = -2 \ln L(\widehat{\boldsymbol{\Theta}}_j) + 2k_j,$$

kde  $\widehat{\boldsymbol{\Theta}}_j$  označuje maximálně věrohodný odhad parametru  $\boldsymbol{\Theta}_j$ , (viz. [10]). Hodnota  $AIC$  má v podstatě význam jakési penalizace maximální hodnoty logaritmicke věrohodnostní funkce rovné počtu parametrů  $k_j$ . Preferovaný pak bude model, pro který je penalizovaná logaritmicke věrohodnostní funkce

$$\ln L(\widehat{\boldsymbol{\Theta}}) - k_j$$

největší. V našem případě jsme vyčíslili  $AIC$  postupně pro modely s jednou, dvěma, třemi a čtyřmi vysvětlujícími proměnnými ve všech možných kombinacích. V závislosti na počtu použitých proměnných jsme odhadovali i odpovídající počet parametrů. Pro jednu proměnnou jsme uvažovali vztah mezi ní a vektorem parametrů  $\boldsymbol{\delta}$  ve tvaru  $\delta_0 + \delta_1 \cdot v_l^{(i)}$  pro  $l = 1, 2, 3, 4$ . Pro dvě proměnné odhadujeme  $\delta_0, \delta_1$  a  $\delta_2$  atd. Výsledky jsme zaznamenali v následujícím grafu.



Obr. 5.3:  $AIC$  postupně pro modely s parametry  $\delta_0, \delta_1$  a  $\delta_2$  až  $\delta_0, \dots, \delta_4$

Z grafu je patrné, že nejmenší hodnotu  $AIC$  nabývá v případě jednoho modelu se dvěma vektory proměnných, přičemž se jedná o parametry vyjadřující závislost na hodnotě úvěru a předchozí platební morálce klienta. Tento model se jeví jako nejlepší ze všech uvažovaných.

Nyní budeme postupně testovat, zda některý z parametrů  $\delta_0$ ,  $\delta_1$  a  $\delta_2$  nemůže být roven nule. Budeme testovat model  $M$  s parametry  $\Theta_M = (\delta_0, \delta_1, \delta_2)$  proti užšímu modelu  $\widetilde{M}$ . Model  $\widetilde{M}$  s vektorem parametrů  $\Theta_{\widetilde{M}}$  vznikne z  $M$  položením  $\delta_j = 0$  postupně pro všechna  $j = 0, 1, 2$ . Testová statistika

$$-2 \ln \frac{L(\Theta_{\widetilde{M}})}{L(\Theta_M)}$$

má asymptoticky  $\chi^2$ -kvadrát rozdělení s  $q$  stupni volnosti (viz. [2]), kde  $q$  vyjadřuje rozdíl v počtu parametrů mezi modely,  $q = \dim(\Theta_M) - \dim(\Theta_{\widetilde{M}})$ . V tomto konkrétním případě je  $q = 1$ . Pro všechna  $j$  jsou  $p$ -hodnoty testů menší než hladina 5%, takže hypotézu o nulovosti některého z parametrů zamítáme. Výsledky jsou uvedeny v tabulce 5.1.

	statistika	$p$ -hodnota
$\delta_0$	30,10	$4.09675 \times 10^{-8}$
$\delta_1$	16,51	0.0000399477
$\delta_2$	16,87	0.0000485141

Tabulka 5.1: Výsledky testu o nulovosti některého z parametrů

Pro výsledný model, který využívá vysvětlující proměnné popisující hodnotu úvěru a předchozí platební morálku klienta, mají odhady metodou maximální věrohodnosti hodnoty uvedené v tabulce 5.2.

$p$	0.
$\delta_0$	4.06592
$\delta_1$	0.484884
$\delta_2$	-0.664428

Tabulka 5.2: Hodnoty parametrů pro vybraný model

Abychom získali intervaly spolehlivosti uvedené v kapitole 4.2 a 4.3, musíme znát rozdělení časového cenzoru  $T$ . V [2] se uvádí jako nejlépe vystihující rozdělení náhodného cenzoru  $T$  rozdělení exponenciální.

## Závěr

V práci jsme se zabývali možnostmi, jak modelovat úvěrové riziko. Věnovali jsme se hlavně dvěma veličinám, pomocí nichž lze úvěrové riziko vyjádřit. Nejdříve to byla pravděpodobnost selhání a poté doba do selhání.

Ukázali jsme tři metody, jak stanovit časovou strukturu pravděpodobnosti selhání z hodnot dostupných na trhu. První metoda byla matematicky poměrně snadná, ale v praxi se příliš nepoužívá. Výsledky, které se pomocí ní získají, příliš neodpovídají reálnému očekávání. A tak jsme se podrobněji věnovali dalšímu postupu, který vychází z úvěrové migrace a využívá ratingová ohodnocení protistran. Tento přístup je poměrně rozšířený, protože umožňuje snadno získat pravděpodobnosti selhání pro různé časové intervaly. Nevýhoda tohoto postupu ale spočívá v předpokladu, že Markovův řetězec, kterým se úvěrová migrace mezi kategoriemi ohodnocení popisuje, je homogenní. K ověření platnosti tohoto předpokladu jsme uvedli dvě statistické metody, pomocí nichž lze zjistit, zda Markovův řetězec je homogenní. Ukázali jsme pak na datech jedné z ratingových agentury jejich použití a zároveň došli k nepříznivému závěru, že předpoklad homogenního markovského řetězce není většinou v praxi splněn.

Třetí metoda vychází z podobnosti odvození úvěrové křivky, která zobrazuje časovou strukturu pravděpodobnosti selhání, s odvozením křivky výnosové. Využívá postupy používané k vyhlazování výnosových křivek ke stanovení časové struktury pravděpodobnosti selhání. Tento přístup se nejčastěji uplatňuje u úvěrových derivátů.

Poté jsme se věnovali době do selhání. Popsali jsme model, který využívá náhodné cenzorování k odvození rozdělení doby do selhání. Takto navržený model umožňuje provádět analýzu portfolia půjček ještě před okamžikem, než všechny vyprší nebo nebudou případně splaceny. Navíc je použitelný i v případě, že v portfoliu dochází k selhání jen zřídka. Tato situace je typická pro úvěry a s nimi neodmyslitelně spojené úvěrové riziko. Použití navrženého modelu jsme následně ukázali na reálných datech.

# Příloha

## Data pro $\chi^2$ -test dobré shody a test poměrem věrohodností

### Matice četností

$$N^{(1)} = \begin{pmatrix} 42 & 5 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 816 & 15 & 1 & 0 & 0 & 0 & 0 \\ 0 & 34 & 2740 & 33 & 2 & 0 & 1 & 0 \\ 0 & 2 & 13 & 973 & 7 & 2 & 3 & 0 \\ 0 & 0 & 0 & 5 & 44 & 3 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \end{pmatrix} \quad N^{(2)} = \begin{pmatrix} 35 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 880 & 12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 34 & 2756 & 41 & 2 & 0 & 0 & 0 \\ 0 & 1 & 26 & 1006 & 11 & 3 & 1 & 1 \\ 0 & 0 & 0 & 5 & 42 & 0 & 1 & 0 \\ 0 & 0 & 0 & 5 & 0 & 13 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & 1 \end{pmatrix}$$

$$N^{(3)} = \begin{pmatrix} 37 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 960 & 11 & 0 & 0 & 0 & 0 & 0 \\ 0 & 27 & 2786 & 34 & 1 & 0 & 0 & 0 \\ 0 & 4 & 17 & 1043 & 3 & 2 & 1 & 0 \\ 0 & 0 & 0 & 5 & 45 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 13 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 & 0 \end{pmatrix} \quad N^{(4)} = \begin{pmatrix} 45 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1022 & 4 & 7 & 0 & 0 & 0 & 0 \\ 0 & 42 & 2828 & 18 & 0 & 0 & 2 & 0 \\ 0 & 0 & 19 & 1129 & 7 & 0 & 0 & 0 \\ 0 & 0 & 0 & 3 & 44 & 1 & 0 & 2 \\ 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 7 & 1 \end{pmatrix}$$

$$N^{(5)} = \begin{pmatrix} 46 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1114 & 17 & 0 & 0 & 0 & 0 & 0 \\ 0 & 20 & 2889 & 76 & 3 & 1 & 0 & 0 \\ 0 & 2 & 14 & 1166 & 12 & 2 & 6 & 0 \\ 0 & 0 & 0 & 0 & 44 & 4 & 0 & 0 \\ 0 & 0 & 0 & 2 & 1 & 8 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 4 & 3 \end{pmatrix} \quad N^{(6)} = \begin{pmatrix} 46 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1171 & 69 & 0 & 0 & 0 & 0 & 0 \\ 0 & 22 & 2964 & 45 & 3 & 0 & 0 & 0 \\ 0 & 0 & 14 & 1269 & 3 & 1 & 3 & 0 \\ 0 & 0 & 0 & 2 & 57 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 12 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 10 & 1 \end{pmatrix}$$

$$N^{(7)} = \begin{pmatrix} 45 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 2 & 1255 & 12 & 0 & 0 & 0 & 0 & 0 \\ 0 & 26 & 3104 & 50 & 1 & 0 & 0 & 0 \\ 0 & 1 & 16 & 1393 & 9 & 2 & 0 & 0 \\ 0 & 0 & 0 & 8 & 44 & 4 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 9 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 6 & 2 & 5 \end{pmatrix} \quad N^{(8)} = \begin{pmatrix} 46 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1311 & 7 & 0 & 0 & 0 & 0 & 0 \\ 0 & 15 & 3139 & 21 & 0 & 0 & 0 & 0 \\ 0 & 0 & 34 & 1427 & 10 & 1 & 1 & 0 \\ 0 & 0 & 0 & 3 & 43 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 6 & 13 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 0 \end{pmatrix}$$

$$N^{(9)} = \begin{pmatrix} 48 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1356 & 16 & 0 & 0 & 0 & 0 & 0 \\ 1 & 29 & 3051 & 38 & 0 & 0 & 0 & 0 \\ 0 & 0 & 12 & 1449 & 11 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 48 & 2 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 12 & 1 & 2 \\ 0 & 0 & 0 & 0 & 0 & 1 & 4 & 1 \end{pmatrix} \quad N^{(10)} = \begin{pmatrix} 50 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 1389 & 20 & 0 & 0 & 0 & 0 & 0 \\ 0 & 25 & 2909 & 65 & 0 & 1 & 0 & 0 \\ 0 & 0 & 26 & 1395 & 14 & 4 & 0 & 0 \\ 0 & 1 & 0 & 15 & 39 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 2 & 10 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 3 & 0 \end{pmatrix}$$

$$N^{(11)} = \begin{pmatrix} 51 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 8 & 1408 & 15 & 0 & 0 & 0 & 0 & 0 \\ 0 & 54 & 2812 & 25 & 2 & 1 & 0 & 0 \\ 0 & 1 & 35 & 1386 & 8 & 0 & 1 & 0 \\ 0 & 0 & 0 & 7 & 38 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 & 10 & 2 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 2 & 0 \end{pmatrix} \quad N^{(12)} = \begin{pmatrix} 59 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 1512 & 11 & 0 & 0 & 0 & 0 & 0 \\ 0 & 48 & 2888 & 8 & 0 & 0 & 0 & 0 \\ 0 & 0 & 34 & 1437 & 3 & 2 & 0 & 0 \\ 0 & 0 & 0 & 5 & 38 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 13 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 \end{pmatrix}$$

$$N^{(13)} = \begin{pmatrix} 67 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 7 & 1610 & 5 & 0 & 0 & 0 & 0 & 0 \\ 1 & 67 & 3018 & 7 & 0 & 3 & 0 & 0 \\ 0 & 1 & 24 & 1429 & 9 & 0 & 1 & 0 \\ 0 & 0 & 0 & 2 & 29 & 1 & 1 & 0 \\ 0 & 0 & 0 & 1 & 3 & 8 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 5 & 0 \end{pmatrix} \quad N^{(14)} = \begin{pmatrix} 79 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 3 & 1785 & 10 & 0 & 0 & 0 & 0 & 0 \\ 0 & 54 & 3435 & 19 & 0 & 1 & 0 & 0 \\ 0 & 1 & 96 & 1513 & 13 & 2 & 0 & 0 \\ 0 & 0 & 0 & 3 & 38 & 3 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 3 & 4 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & 1 \end{pmatrix}$$

$$N^{(15)} = \begin{pmatrix} 73 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 19 & 1972 & 13 & 0 & 0 & 0 & 0 & 0 \\ 0 & 218 & 3768 & 16 & 0 & 0 & 0 & 0 \\ 0 & 1 & 53 & 1675 & 10 & 4 & 1 & 0 \\ 0 & 0 & 1 & 1 & 48 & 2 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 6 & 1 & 2 \\ 0 & 0 & 0 & 0 & 1 & 2 & 4 & 1 \end{pmatrix} \quad N^{(16)} = \begin{pmatrix} 107 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 1 & 2342 & 61 & 0 & 0 & 0 & 0 & 0 \\ 0 & 317 & 4251 & 16 & 1 & 0 & 0 & 0 \\ 0 & 8 & 127 & 1859 & 7 & 0 & 0 & 0 \\ 0 & 1 & 0 & 4 & 57 & 1 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 13 & 2 & 1 \\ 0 & 0 & 0 & 0 & 0 & 0 & 4 & 2 \end{pmatrix}$$

$$N^{(17)} = \begin{pmatrix} 117 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 13 & 2774 & 23 & 0 & 0 & 0 & 0 & 0 \\ 0 & 100 & 4657 & 37 & 0 & 0 & 0 & 0 \\ 0 & 0 & 44 & 1897 & 15 & 3 & 0 & 0 \\ 0 & 0 & 0 & 3 & 59 & 1 & 0 & 1 \\ 0 & 0 & 0 & 0 & 0 & 12 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 \end{pmatrix}$$

$$N^{(18)} = \begin{pmatrix} 137 & 3 & 0 & 0 & 0 & 0 & 0 & 0 \\ 8 & 2942 & 37 & 0 & 0 & 0 & 0 & 0 \\ 0 & 83 & 4827 & 92 & 3 & 0 & 0 & 0 \\ 0 & 0 & 51 & 1932 & 18 & 2 & 1 & 0 \\ 0 & 0 & 0 & 4 & 64 & 7 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 15 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 6 & 0 \end{pmatrix}$$

## Matice relativních četností

$$\hat{P}^{(1)} = \begin{pmatrix} 0.8936 & 0.1064 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0012 & 0.9796 & 0.018 & 0.0012 & 0 & 0 & 0 & 0 \\ 0 & 0.0121 & 0.9751 & 0.0117 & 0.0007 & 0 & 0.0004 & 0 \\ 0 & 0.002 & 0.013 & 0.973 & 0.007 & 0.002 & 0.003 & 0 \\ 0 & 0 & 0 & 0.0962 & 0.8462 & 0.0577 & 0 & 0 \\ 0 & 0 & 0 & 0.0625 & 0 & 0.9375 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(2)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0022 & 0.9843 & 0.0134 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.012 & 0.9728 & 0.0145 & 0.0007 & 0 & 0 & 0 \\ 0 & 0.001 & 0.0248 & 0.959 & 0.0105 & 0.0029 & 0.001 & 0.001 \\ 0 & 0 & 0 & 0.1042 & 0.875 & 0 & 0.0208 & 0 \\ 0 & 0 & 0 & 0.2632 & 0 & 0.6842 & 0.0526 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.8571 & 0.1429 \end{pmatrix}$$

$$\hat{P}^{(3)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.001 & 0.9877 & 0.0113 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0095 & 0.9782 & 0.0119 & 0.0004 & 0 & 0 & 0 \\ 0 & 0.0037 & 0.0159 & 0.9748 & 0.0028 & 0.0019 & 0.0009 & 0 \\ 0 & 0 & 0 & 0.0962 & 0.8654 & 0.0192 & 0.0192 & 0 \\ 0 & 0 & 0 & 0 & 0.0667 & 0.8667 & 0.0667 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(4)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.001 & 0.9884 & 0.0039 & 0.0068 & 0 & 0 & 0 & 0 \\ 0 & 0.0145 & 0.9785 & 0.0062 & 0 & 0 & 0.0007 & 0 \\ 0 & 0 & 0.0165 & 0.9775 & 0.0061 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.06 & 0.88 & 0.02 & 0 & 0.04 \\ 0 & 0 & 0 & 0 & 0 & 1. & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.875 & 0.125 \end{pmatrix}$$

$$\hat{P}^{(5)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0018 & 0.9832 & 0.015 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0067 & 0.9665 & 0.0254 & 0.001 & 0.0003 & 0 & 0 \\ 0 & 0.0017 & 0.0116 & 0.97 & 0.01 & 0.0017 & 0.005 & 0 \\ 0 & 0 & 0 & 0 & 0.9167 & 0.0833 & 0 & 0 \\ 0 & 0 & 0 & 0.1667 & 0.0833 & 0.6667 & 0.0833 & 0 \\ 0 & 0 & 0 & 0 & 0.125 & 0 & 0.5 & 0.375 \end{pmatrix}$$

$$\hat{P}^{(6)} = \begin{pmatrix} 0.9787 & 0.0213 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.9444 & 0.0556 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0073 & 0.9769 & 0.0148 & 0.001 & 0 & 0 & 0 \\ 0 & 0 & 0.0109 & 0.9837 & 0.0023 & 0.0008 & 0.0023 & 0 \\ 0 & 0 & 0 & 0.0333 & 0.95 & 0 & 0.0167 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9231 & 0.0769 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.9091 & 0.0909 \end{pmatrix}$$

$$\hat{P}^{(7)} = \begin{pmatrix} 0.9783 & 0.0217 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0016 & 0.989 & 0.0095 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0082 & 0.9758 & 0.0157 & 0.0003 & 0 & 0 & 0 \\ 0 & 0.0007 & 0.0113 & 0.9803 & 0.0063 & 0.0014 & 0 & 0 \\ 0 & 0 & 0 & 0.1404 & 0.7719 & 0.0702 & 0.0175 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.9 & 0.1 & 0 \\ 0 & 0 & 0 & 0 & 0.0714 & 0.4286 & 0.1429 & 0.3571 \end{pmatrix}$$

$$\hat{P}^{(8)} = \begin{pmatrix} 0.9787 & 0.0213 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0008 & 0.9939 & 0.0053 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0047 & 0.9887 & 0.0066 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0231 & 0.9688 & 0.0068 & 0.0007 & 0.0007 & 0 \\ 0 & 0 & 0 & 0.0612 & 0.8776 & 0.0612 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.3 & 0.65 & 0.05 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(9)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0007 & 0.9876 & 0.0117 & 0 & 0 & 0 & 0 & 0 \\ 0.0003 & 0.0093 & 0.9782 & 0.0122 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0081 & 0.983 & 0.0075 & 0.0007 & 0.0007 & 0 \\ 0 & 0 & 0 & 0.0196 & 0.9412 & 0.0392 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0625 & 0.75 & 0.0625 & 0.125 \\ 0 & 0 & 0 & 0 & 0 & 0.1667 & 0.6667 & 0.1667 \end{pmatrix}$$

$$\hat{P}^{(10)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0007 & 0.9851 & 0.0142 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0083 & 0.9697 & 0.0217 & 0 & 0.0003 & 0 & 0 \\ 0 & 0 & 0.0181 & 0.9694 & 0.0097 & 0.0028 & 0 & 0 \\ 0 & 0.0182 & 0 & 0.2727 & 0.7091 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0.0714 & 0.1429 & 0.7143 & 0 & 0.0714 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(11)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0056 & 0.9839 & 0.0105 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0187 & 0.9717 & 0.0086 & 0.0007 & 0.0003 & 0 & 0 \\ 0 & 0.0007 & 0.0245 & 0.9686 & 0.0056 & 0 & 0.0007 & 0 \\ 0 & 0 & 0 & 0.1489 & 0.8085 & 0.0213 & 0.0213 & 0 \\ 0 & 0 & 0 & 0 & 0.0769 & 0.7692 & 0.1538 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(12)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0046 & 0.9882 & 0.0072 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0163 & 0.981 & 0.0027 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.023 & 0.9736 & 0.002 & 0.0014 & 0 & 0 \\ 0 & 0 & 0 & 0.1163 & 0.8837 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0714 & 0.9286 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(13)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0043 & 0.9926 & 0.0031 & 0 & 0 & 0 & 0 & 0 \\ 0.0003 & 0.0216 & 0.9748 & 0.0023 & 0 & 0.001 & 0 & 0 \\ 0 & 0.0007 & 0.0164 & 0.9761 & 0.0061 & 0 & 0.0007 & 0 \\ 0 & 0 & 0 & 0.0606 & 0.8788 & 0.0303 & 0.0303 & 0 \\ 0 & 0 & 0 & 0.0769 & 0.2308 & 0.6154 & 0.0769 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(14)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0017 & 0.9928 & 0.0056 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0154 & 0.9789 & 0.0054 & 0 & 0.0003 & 0 & 0 \\ 0 & 0.0006 & 0.0591 & 0.9311 & 0.008 & 0.0012 & 0 & 0 \\ 0 & 0 & 0 & 0.0682 & 0.8636 & 0.0682 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.125 & 0.375 & 0.5 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0.8571 & 0.1429 \end{pmatrix}$$

$$\hat{P}^{(15)} = \begin{pmatrix} 0.9733 & 0.0133 & 0.0133 & 0 & 0 & 0 & 0 & 0 \\ 0.0095 & 0.984 & 0.0065 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0545 & 0.9415 & 0.004 & 0 & 0 & 0 & 0 \\ 0 & 0.0006 & 0.0304 & 0.9604 & 0.0057 & 0.0023 & 0.0006 & 0 \\ 0 & 0 & 0.0192 & 0.0192 & 0.9231 & 0.0385 & 0 & 0 \\ 0 & 0 & 0.1 & 0 & 0 & 0.6 & 0.1 & 0.2 \\ 0 & 0 & 0 & 0 & 0.125 & 0.25 & 0.5 & 0.125 \end{pmatrix}$$

$$\hat{P}^{(16)} = \begin{pmatrix} 0.9907 & 0.0093 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0004 & 0.9742 & 0.0254 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0691 & 0.9272 & 0.0035 & 0.0002 & 0 & 0 & 0 \\ 0 & 0.004 & 0.0635 & 0.929 & 0.0035 & 0 & 0 & 0 \\ 0 & 0.0156 & 0 & 0.0625 & 0.8906 & 0.0156 & 0.0156 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0.8125 & 0.125 & 0.0625 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0.6667 & 0.3333 \end{pmatrix}$$



$$\hat{P}^{(17)} = \begin{pmatrix} 1. & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0046 & 0.9872 & 0.0082 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0209 & 0.9714 & 0.0077 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0.0225 & 0.9684 & 0.0077 & 0.0015 & 0 & 0 \\ 0 & 0 & 0 & 0.0469 & 0.9219 & 0.0156 & 0 & 0.0156 \\ 0 & 0 & 0 & 0 & 0 & 1. & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

$$\hat{P}^{(18)} = \begin{pmatrix} 0.9786 & 0.0214 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0.0027 & 0.9849 & 0.0124 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0.0166 & 0.9644 & 0.0184 & 0.0006 & 0 & 0 & 0 \\ 0 & 0 & 0.0254 & 0.9641 & 0.009 & 0.001 & 0.0005 & 0 \\ 0 & 0 & 0 & 0.0533 & 0.8533 & 0.0933 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0.0625 & 0.9375 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1. & 0 \end{pmatrix}$$

## Matice $n$ -tého řádu

$$P(1) = \begin{pmatrix} 0.9617 & 0.0111 & 0.0009 & 0. & 0. & 0. & 0. & 0. \\ 0.0027 & 0.9464 & 0.0123 & 0.0003 & 0. & 0. & 0. & 0. \\ 0. & 0.0189 & 0.9161 & 0.01 & 0.0003 & 0.0001 & 0. & 0. \\ 0. & 0.0008 & 0.0233 & 0.9024 & 0.006 & 0.0011 & 0.0007 & 0. \\ 0. & 0.0019 & 0.001 & 0.0729 & 0.7879 & 0.0326 & 0.0067 & 0.0029 \\ 0. & 0. & 0.0036 & 0.0356 & 0.0641 & 0.7011 & 0.0605 & 0.0214 \\ 0. & 0. & 0. & 0. & 0.0224 & 0.0672 & 0.6642 & 0.1194 \end{pmatrix}$$

$$P(2) = \begin{pmatrix} 0.9389 & 0.0165 & 0.0019 & 0. & 0. & 0. & 0. & 0. \\ 0.0056 & 0.8993 & 0.0205 & 0.0014 & 0. & 0. & 0. & 0. \\ 0.0001 & 0.0353 & 0.8393 & 0.0178 & 0.0006 & 0.0002 & 0.0002 & 0. \\ 0. & 0.0038 & 0.0415 & 0.8103 & 0.0105 & 0.0022 & 0.0013 & 0.0004 \\ 0. & 0.0021 & 0.0041 & 0.1325 & 0.618 & 0.0404 & 0.0104 & 0.0072 \\ 0. & 0. & 0.0038 & 0.083 & 0.083 & 0.5094 & 0.0453 & 0.0453 \\ 0. & 0. & 0. & 0. & 0.0781 & 0.0781 & 0.4688 & 0.1641 \end{pmatrix}$$

$$P(3) = \begin{pmatrix} 0.9254 & 0.0219 & 0.0033 & 0. & 0. & 0. & 0. & 0. \\ 0.0086 & 0.857 & 0.0257 & 0.0035 & 0. & 0. & 0. & 0. \\ 0.0001 & 0.0493 & 0.7677 & 0.0236 & 0.0009 & 0.0002 & 0.0002 & 0.0001 \\ 0. & 0.0063 & 0.0568 & 0.7238 & 0.0134 & 0.0033 & 0.0017 & 0.0008 \\ 0. & 0.0022 & 0.0089 & 0.1654 & 0.4828 & 0.0455 & 0.0122 & 0.0111 \\ 0. & 0. & 0.004 & 0.1355 & 0.0637 & 0.3625 & 0.0438 & 0.0558 \\ 0. & 0. & 0. & 0.0083 & 0.1322 & 0.0413 & 0.3554 & 0.2066 \end{pmatrix}$$

$$P(4) = \begin{pmatrix} 0.9134 & 0.0263 & 0.0038 & 0. & 0. & 0. & 0. & 0. \\ 0.0123 & 0.8174 & 0.0318 & 0.0028 & 0. & 0. & 0. & 0. \\ 0.0002 & 0.0576 & 0.7019 & 0.029 & 0.0013 & 0.0002 & 0.0002 & 0.0002 \\ 0. & 0.0085 & 0.0664 & 0.6436 & 0.0153 & 0.0042 & 0.0019 & 0.0015 \\ 0. & 0.0012 & 0.0096 & 0.1964 & 0.3713 & 0.0407 & 0.012 & 0.012 \\ 0. & 0. & 0.0086 & 0.1588 & 0.0515 & 0.2532 & 0.0343 & 0.0601 \\ 0. & 0. & 0. & 0.0702 & 0.0877 & 0.0526 & 0.2632 & 0.2368 \end{pmatrix}$$

$$P(5) = \begin{pmatrix} 0.906 & 0.0309 & 0.0042 & 0. & 0. & 0. & 0. & 0. \\ 0.0157 & 0.7778 & 0.0376 & 0.0028 & 0. & 0.0001 & 0.0001 & 0. \\ 0.0002 & 0.0624 & 0.6437 & 0.0324 & 0.0017 & 0.0003 & 0.0002 & 0.0002 \\ 0. & 0.0108 & 0.0725 & 0.5728 & 0.0161 & 0.0045 & 0.0021 & 0.002 \\ 0. & 0.0013 & 0.0102 & 0.2074 & 0.2971 & 0.0359 & 0.009 & 0.0115 \\ 0. & 0. & 0.009 & 0.1614 & 0.0314 & 0.2063 & 0.0269 & 0.0583 \\ 0. & 0. & 0. & 0.1442 & 0.0192 & 0.0481 & 0.2212 & 0.2596 \end{pmatrix}$$

$$P(6) = \begin{pmatrix} 0.8975 & 0.0363 & 0.0047 & 0. & 0. & 0. & 0. & 0. \\ 0.0192 & 0.7409 & 0.0419 & 0.0026 & 0. & 0.0001 & 0.0001 & 0. \\ 0.0002 & 0.0673 & 0.5885 & 0.0348 & 0.0019 & 0.0004 & 0.0003 & 0.0003 \\ 0.0001 & 0.0107 & 0.077 & 0.5101 & 0.0159 & 0.0047 & 0.002 & 0.0024 \\ 0. & 0.0027 & 0.0123 & 0.218 & 0.2275 & 0.0327 & 0.0095 & 0.0109 \\ 0. & 0. & 0.0142 & 0.1517 & 0.0237 & 0.1659 & 0.0142 & 0.0711 \\ 0. & 0. & 0. & 0.1458 & 0. & 0.0521 & 0.2083 & 0.2604 \end{pmatrix}$$

$$P(7) = \begin{pmatrix} 0.8889 & 0.0388 & 0.0053 & 0. & 0. & 0. & 0. & 0. \\ 0.0225 & 0.7056 & 0.0451 & 0.0024 & 0. & 0.0002 & 0.0001 & 0. \\ 0.0002 & 0.0723 & 0.5385 & 0.0361 & 0.002 & 0.0004 & 0.0003 & 0.0003 \\ 0.0001 & 0.0111 & 0.0805 & 0.4539 & 0.0149 & 0.0048 & 0.0016 & 0.003 \\ 0. & 0.0043 & 0.0145 & 0.224 & 0.1806 & 0.0289 & 0.0087 & 0.0101 \\ 0. & 0. & 0.0102 & 0.1327 & 0.0153 & 0.1378 & 0.0051 & 0.0816 \\ 0. & 0. & 0. & 0.1319 & 0. & 0.0549 & 0.1868 & 0.2747 \end{pmatrix}$$

$$P(8) = \begin{pmatrix} 0.8858 & 0.0394 & 0.0039 & 0. & 0. & 0. & 0. & 0. \\ 0.0253 & 0.6713 & 0.0475 & 0.0025 & 0. & 0.0003 & 0.0001 & 0. \\ 0.0002 & 0.076 & 0.4925 & 0.0371 & 0.0021 & 0.0004 & 0.0003 & 0.0004 \\ 0.0001 & 0.0116 & 0.083 & 0.4 & 0.0139 & 0.0043 & 0.0016 & 0.0033 \\ 0. & 0.0062 & 0.0203 & 0.2122 & 0.1435 & 0.0265 & 0.0078 & 0.0109 \\ 0. & 0. & 0.0055 & 0.1105 & 0.0055 & 0.1105 & 0. & 0.0939 \\ 0. & 0. & 0. & 0.1412 & 0. & 0.0471 & 0.1765 & 0.2941 \end{pmatrix}$$

$$P(9) = \begin{pmatrix} 0.884 & 0.035 & 0.0022 & 0. & 0. & 0. & 0. & 0. \\ 0.0281 & 0.6393 & 0.0497 & 0.0025 & 0. & 0.0004 & 0.0001 & 0. \\ 0.0003 & 0.0784 & 0.451 & 0.0383 & 0.0023 & 0.0004 & 0.0003 & 0.0005 \\ 0.0001 & 0.012 & 0.0832 & 0.3534 & 0.012 & 0.004 & 0.0018 & 0.0037 \\ 0. & 0.0068 & 0.029 & 0.186 & 0.1212 & 0.0239 & 0.0068 & 0.0119 \\ 0. & 0. & 0. & 0.0964 & 0. & 0.0904 & 0. & 0.1084 \\ 0. & 0. & 0. & 0.1585 & 0. & 0.0366 & 0.1585 & 0.3049 \end{pmatrix}$$

$$P(10) = \begin{pmatrix} 0.8796 & 0.0369 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0314 & 0.611 & 0.0516 & 0.0031 & 0. & 0.0005 & 0.0001 & 0. \\ 0.0004 & 0.0792 & 0.4177 & 0.038 & 0.0024 & 0.0003 & 0.0004 & 0.0006 \\ 0. & 0.0123 & 0.0831 & 0.3165 & 0.01 & 0.0037 & 0.0018 & 0.004 \\ 0. & 0.0076 & 0.0363 & 0.1527 & 0.1107 & 0.0172 & 0.0057 & 0.0153 \\ 0. & 0. & 0. & 0.06 & 0. & 0.0867 & 0. & 0.1133 \\ 0. & 0. & 0. & 0.1711 & 0. & 0.0263 & 0.1447 & 0.3289 \end{pmatrix}$$

$$P(11) = \begin{pmatrix} 0.8747 & 0.0362 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0353 & 0.586 & 0.0522 & 0.0039 & 0. & 0.0006 & 0.0001 & 0. \\ 0.0005 & 0.0817 & 0.393 & 0.0371 & 0.0025 & 0.0003 & 0.0006 & 0.0008 \\ 0. & 0.011 & 0.0851 & 0.2861 & 0.0083 & 0.0031 & 0.0016 & 0.0045 \\ 0. & 0.0087 & 0.039 & 0.1255 & 0.1017 & 0.0195 & 0.0043 & 0.0173 \\ 0. & 0. & 0. & 0.0458 & 0. & 0.0763 & 0. & 0.1069 \\ 0. & 0. & 0. & 0.1714 & 0. & 0.0143 & 0.1286 & 0.3429 \end{pmatrix}$$

$$P(12) = \begin{pmatrix} 0.8622 & 0.0385 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0397 & 0.5656 & 0.0542 & 0.0042 & 0. & 0.0005 & 0.0001 & 0. \\ 0.0005 & 0.0849 & 0.3738 & 0.0366 & 0.0029 & 0.0004 & 0.0007 & 0.0009 \\ 0. & 0.0095 & 0.0875 & 0.2656 & 0.0072 & 0.0028 & 0.0012 & 0.0052 \\ 0. & 0.0099 & 0.047 & 0.1114 & 0.0916 & 0.0248 & 0.005 & 0.0173 \\ 0. & 0. & 0. & 0.0091 & 0. & 0.0636 & 0. & 0.1091 \\ 0. & 0. & 0. & 0.1667 & 0. & 0. & 0.1061 & 0.3636 \end{pmatrix}$$

$$P(13) = \begin{pmatrix} 0.8496 & 0.0451 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0441 & 0.5486 & 0.0555 & 0.0052 & 0. & 0.0006 & 0. & 0. \\ 0.0006 & 0.0895 & 0.3572 & 0.0356 & 0.0031 & 0.0005 & 0.001 & 0.001 \\ 0. & 0.0079 & 0.0902 & 0.2449 & 0.0066 & 0.0025 & 0.0007 & 0.0059 \\ 0. & 0.0118 & 0.0473 & 0.1006 & 0.0799 & 0.0325 & 0. & 0.0207 \\ 0. & 0. & 0. & 0.0103 & 0. & 0.0412 & 0. & 0.1134 \\ 0. & 0. & 0. & 0.1176 & 0. & 0. & 0.1176 & 0.3725 \end{pmatrix}$$

$$P(14) = \begin{pmatrix} 0.8349 & 0.0505 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.049 & 0.537 & 0.056 & 0.0054 & 0. & 0.0006 & 0. & 0. \\ 0.0007 & 0.0931 & 0.3431 & 0.0367 & 0.0035 & 0.0005 & 0.0011 & 0.0013 \\ 0. & 0.0079 & 0.0925 & 0.2252 & 0.0062 & 0.0028 & 0.0002 & 0.0065 \\ 0. & 0.011 & 0.0476 & 0.0842 & 0.0696 & 0.0403 & 0. & 0.022 \\ 0. & 0. & 0.0122 & 0.0244 & 0. & 0.0122 & 0. & 0.1098 \\ 0. & 0. & 0. & 0.05 & 0. & 0. & 0.125 & 0.4 \end{pmatrix}$$

$$P(15) = \begin{pmatrix} 0.8081 & 0.0581 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0543 & 0.5255 & 0.0569 & 0.0052 & 0. & 0.0005 & 0. & 0. \\ 0.001 & 0.0964 & 0.3318 & 0.0368 & 0.0036 & 0.0008 & 0.0009 & 0.0015 \\ 0. & 0.0103 & 0.0966 & 0.2098 & 0.0059 & 0.0035 & 0. & 0.0073 \\ 0. & 0.009 & 0.0498 & 0.0769 & 0.0543 & 0.0362 & 0. & 0.0226 \\ 0. & 0. & 0.029 & 0.0435 & 0. & 0. & 0. & 0.1014 \\ 0. & 0. & 0. & 0.0645 & 0. & 0. & 0.129 & 0.3548 \end{pmatrix}$$

$$P(16) = \begin{pmatrix} 0.7953 & 0.063 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0561 & 0.5084 & 0.0578 & 0.0054 & 0. & 0.0004 & 0. & 0. \\ 0.0011 & 0.0972 & 0.3112 & 0.0362 & 0.0036 & 0.0009 & 0.0009 & 0.0015 \\ 0. & 0.0123 & 0.096 & 0.1897 & 0.006 & 0.0036 & 0. & 0.0087 \\ 0. & 0.006 & 0.0539 & 0.0778 & 0.0419 & 0.024 & 0. & 0.012 \\ 0. & 0. & 0.0377 & 0.0566 & 0. & 0. & 0. & 0.0943 \\ 0. & 0. & 0. & 0.0952 & 0. & 0. & 0.1429 & 0.3333 \end{pmatrix}$$

$$P(17) = \begin{pmatrix} 0.7667 & 0.0778 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0588 & 0.4939 & 0.0577 & 0.0044 & 0. & 0. & 0. & 0. \\ 0.0013 & 0.0947 & 0.2878 & 0.0357 & 0.004 & 0.0012 & 0.0007 & 0.0017 \\ 0. & 0.0141 & 0.0899 & 0.1716 & 0.0054 & 0.0041 & 0. & 0.0095 \\ 0. & 0. & 0.0636 & 0.0818 & 0.0364 & 0.0091 & 0. & 0.0091 \\ 0. & 0. & 0.0541 & 0.0541 & 0. & 0. & 0. & 0.0811 \\ 0. & 0. & 0. & 0.0833 & 0. & 0. & 0.1667 & 0.4167 \end{pmatrix}$$

$$P(18) = \begin{pmatrix} 0.7021 & 0.1277 & 0. & 0. & 0. & 0. & 0. & 0. \\ 0.0608 & 0.4822 & 0.054 & 0.0057 & 0. & 0. & 0. & 0. \\ 0.0013 & 0.0927 & 0.2642 & 0.0357 & 0.0037 & 0.002 & 0.0007 & 0.0017 \\ 0. & 0.0149 & 0.0849 & 0.1549 & 0.0056 & 0.0047 & 0. & 0.0103 \\ 0. & 0. & 0.0727 & 0.1091 & 0.0182 & 0. & 0. & 0.0182 \\ 0. & 0. & 0.0625 & 0.0625 & 0. & 0. & 0. & 0.0625 \\ 0. & 0. & 0. & 0. & 0. & 0. & 0.25 & 0.5 \end{pmatrix}$$

# Literatura

- [1] Anděl J.: *Matematická statistika*, Státní nakladatelství technické literatury, Praha, 1985.
- [2] Beran J., Djaïdja A.-Y. K.: *Credit risk modeling based on survival analysis with immunes*, Statistical Methodology 4, 2007.
- [3] Bluhm Ch., Overbeck L., Wagner Ch.: *An introduction to credit risk modeling*, Chapman & Hall/CRC, Boca Raton, 2003.
- [4] Cipra T.: *Finanční ekonometrie*, Ekopress, Praha, 2008.
- [5] Česká národní banka: *Historie diskontní sazby*,  
[http://www.cnb.cz/m2export/sites/www.cnb.cz/cs/faq/faq\\_diskontni\\_historie.txt](http://www.cnb.cz/m2export/sites/www.cnb.cz/cs/faq/faq_diskontni_historie.txt), citováno 2009.
- [6] Datensatz-Archiv des Instituts für Statistik der Ludwig-Maximilians-Universität München und des Sonderforschungsbereichs 386: *Kreditscoring zur Klassifikation von Kreditnehmern*,  
<http://www.stat.uni-muenchen.de/service/datenarchiv/kredit/kredit.asc>, citováno 2009.
- [7] Hurt J.: *Teorie spolehlivosti*, Státní pedagogické nakladatelství, Praha, 1984.
- [8] Hurt J.: *Yield curves with Mathematica 6.0*,  
<http://library.wolfram.com/infocenter/Conferences/6956/>,  
Wolfram Technology Conference, 2007.
- [9] Mandl P.: *Pravděpodobnostní dynamické modely*, Academia, Praha, 1985.
- [10] McNeil A., Rüdiger F., Embrechts P.: *Quantitative risk management: concepts, techniques and tools*, Princeton University Press, Princeton, N. J., 2005.
- [11] Nelson Ch. R., Siegel A. F.: *Parsimonious modeling of yield curves*, The Journal of Business **60** No. 4, 1987.
- [12] Prášková Z., Lachout P.: *Základy náhodných procesů*, Nakladatelství Karolinum, Praha, 2005.

- [13] Standard and Poor's: *U.S. Municipal rating transitions and defaults 1986 - 2003*,  
<http://www2.standardandpoors.com/portal/site/sp/en/us/page.article/3,1,1,0,1139588928545.html>, 2004.
- [14] Tan B., Yilmaz K.: *Markov chain test for time dependence and homogeneity: an analytical and empirical evaluation*, European Journal of Operational Research **137**, Issue 3, 2002.