UNIVERZA V LJUBLJANI
FAKULTETA ZA MATEMATIKO IN FIZIKO

Finančna matematika – 2. stopnja

Ana Hvalica

Hierarhični linearni strukturni model ter njegova uporaba pri modeliranju verjetnosti propada podjetja

Magistrsko delo

Mentor: prof. dr. Matjaž Omladič

Ljubljana, 2015

Podpisana Ana Hvalica izjavljam:

– da sem magistrsko delo z naslovom *Hierarhični linearni strukturni model ter njegova uporaba pri modeliranju verjetnosti propada podjetja* izdelala samostojno pod mentorstvom prof. dr. Matjaža Omladiča

– da Fakulteti za matematiko in fiziko Univerze v Ljubljani dovoljujem objavo elektronske oblike svojega dela na spletnih straneh.

Ljubljana, 2015 Podpis: ....

Zahvala

Za pomoč, nasvete in strokovne usmeritve pri izdelavi te magistrske naloge se zahvaljujem mentorju, prof. dr. Matjažu Omladiču.

Zahvalila bi se rada tudi svojim staršem, mami Nives ter očetu Jožetu, ker so me navdušili nad matametiko ter vzpodbujali tekom celotnega študija. Hvala tudi sestri Nuši, za vsa bodrenja ter skupna učenja, čeprav vsaka svojega področja.

Vsekakor pa gre zahvala tudi Mateju, ki mi je znal prisluhniti in pomagati uresničetvati zastavljene življenske cilje.

Na koncu pa se zahvaljujem še vsem, ki so kakorkoli pomagali pri pisanju te magistrske naloge.

Kazalo

Program dela

V magistrskem delu predstavite teoretične osnove hierarhičnega linearnega strukturnega modela. Prikažite primer, kjer bi algoritem lahko uporabili.

Osnovna literatura:

- S. Rabe-Hesketh, A. Skrondal, A. Pickles , *Generalized Multilevel structural equation modeling*, Psyhometrika **2** (2004) 167-190.
- S. Rabe-Hesketh, A. Skrondal, *Generalized Linear Latent and Mixed Models with Composite links and Exploded Likelihoods*, Statistical Modeling (2004) 27-39.

Ljubljana, 2015

prof. dr. Matjaž Omladič

Uporaba hierarhičnega linearnega strukturnega modela pri modeliranju verjetnosti propada podjetja

Povzetek

Delo magistrske naloge temelji na obravnavi hierarhičnega linearnega strukturnega modela ter njegovi uporabi pri modeliranju verjetnosti propada podjetja. Glavne težave pri vsakem modeliranju opazovanih podatkov namreč izhajajo iz neupoštevanja osnovnih predpostavk. Pri linearni regresiji ter modelih, ki na njej temelijo, sta kršeni predpostavki homoskedastičnosti ter neodvisnosti med podatki. Njuno neizpolnjevanje povzroča napačne vrednosti koeficientov in pripadajoče vrednosti standardnih napak. V tem delu predstavimo primer 3000 podjetijh iz ameriškega indeksa Russell 3000, kjer jasno pokažemo, da se težavi mulitkolinearnosti in heteroskedastičnosti pojavljajo tudi na teh podatkih. Model, ki težave multikolinearnosti odpravlja, je linearni strukturni model. Z njegovo pomočjo pridobimo pravilne vrednosti koeficientov z nižjimi vrednostmi standardnih odklonov (le redko se namreč zgodi, da kateri izmed koeficientov ne bo statistično značilen). Vse to linearni strukturni model dosega tako, da ne minimizira razdalje med opazovanji ter teoretično vrednostjo, temveč minimizira razliko med teoretično ter empirično variančno-kovariančno matriko. Žal pa linearni strukturni model ne upošteva hierarhične strukture podatkov, kar je pri modeliranju verjetnosti propada podjetja ključnega pomena. Hierarhična struktura izhaja iz samega delovanja podjetij, saj se ta razlikujejo glede na industrijsko panogo, v kateri delujejo. To težavo rešuje nekoliko bolj splošen hierarhični linearni strukturni model. Gre namreč za kombinacijo posplošenih linearnih mešanih modelov ter linearnega strukturnega modela. Zaradi svoje strukture se na danih podatkih prilega boljše kot ostali primerljivi modeli.

Math. Subj. Class. (2015): 62H10, 62H86, 62H25, 62J12 , 62N02

Ključne besede: kreditna analiza, linearni strukturni model, hierahični linearni strukturni model, težave multikolinearnosti, modeliranje verjetnosti propada podjetja, indeks Russell 3000, napovedni model

Hierarchical structural equation model for the estimation of default probabilities

Abstract

In my thesis I have focused on the use of hierarchical structural equation model in the estimation of default probabilities. The main issues arise from the violations of main assumptions. When using linear regression and models based on it we are most often faced with violations of homoscedasticity and no multicollinearity. Breaching those assumptions lead to biased value of coefficients and inflated standard errors. We have used a database with 3000 constituent securities from Russell 3000 index (US stock market). This example will clearly demonstrate the problems of heteroscedasticity and multicollinearity in practice. The model that overcomes these problems is structural equation model. When using structural equation model, we get the correct values of coefficients with significantly lower standard errors. It does so with the minimization of theoretical and empirical variance-covariance matrix instead of minimization of the difference between theoretical and empirical values. Unfortunately, the model does not incorporate hierarchical structure of the data. This has significant impact on the validity of the model, given the importance of industry effect in default probability modelling. Therefore, we extend the structural equation model with a general linear mixed effect model to obtain hierarchical structural equation model. Given its structure, the model offers a better fit with the same input data, as compared to other models.

Math. Subj. Class. (2015): 62H10, 62H86, 62H25, 62J12 , 62N02

Keywords: Credit analysis, Structural equation model, Hierarchical structural equation model, Multicollinearity problems, Estimation of default probabilities, Rusell 3000 Index, Predictive modelling

1 Uvod

Vsak pojav je odvisen od vrste dejavnikov, ki nanj vplivajo. Pri preučevanju pojavov si zato pomagamo z najrazličnejšimi regresijami, kjer preučujemo odnose med spremenljivkami. Še bolj pa so zanimivi modeli, s pomočjo katerih želimo nek rezultat napovedati. Osrednja dogma vsakega takega raziskovanja pa je, da poskušamo pri modelu optimizirati napovedovalno moč in kompleksnost modela. Želimo torej izbrati enostavnejši statistični model, ki je kljub vsemu še vedno pravilen (torej apliciran pravilno glede na dane podatke). Prav pri pravilni izpeljavi modela lahko nastopijo težave, saj nas najenostavnejši modeli, kot je na primer linearna regresija, lahko hitro zavedejo in pripeljejo do napačnih sklepanj. Največje kršitve modelov povzročajo neizpolnjene predpostavke homoskedastičnosti ter neodvisnost podatkov. Medtem ko heteroskedastičnost povroča napačne ocene standardnih odklonov, ima multikolinearnost še hujše posledice. Prav zaradi nje so koeficienti pri opazovanih relacijah slučajnih spremenljivk lahko povsem napačni. Hkrati pa se žal velikokrat izkaže, da konceptualno pomembnih spremenljivk ne moremo direktno opazovati. Takim spremenljivkam pravimo latentne. Njihove lastnosti moramo izpeljati s pomočjo statističnih modelov in znanih spremenljivk. Dva izmed modelov, ki sta sposobna delovati pod prisotnostjo latentnih spremenljivk ter multikolinearnosti, sta linearni strukturni model (SEM) in njegova razširitev, hierarhični linearni strukturni model (HSEM). Modela to dosegata tako, da ne minimizirata razdalje med opazovanji ter teoretično vrednostjo, ampak razliko med teoretično ter empirično variančno-kovariančno matriko. Bistvena razlika med njima pa je, da HSEM model omogoča hierarhično strukturo podatkov. Dovoljuje torej, da imamo podatke v prvi ravni kot osnovne enote, v drugi pa že razvrščene v različne skupine, ki se nato združujejo v večje skupine. Oba modela, s svojimi prednostimi in slabostmi, bomo podrobneje spoznali v magistrski nalogi.

Največji zagon je statistična obdelava podatkov doživela s pojavom zmogljivih računalnikov. To je prineslo velik potencial pri analizi večjih količin podatkov. To so izkoristile tudi banke, ki lahko ob enaki količini prejetih depozitov izdajo več kreditov. Na ta način so veliko bolj profitabilne, hkrati pa tudi ranljive na izjemne dogodke, kot je bila zadnja finančna kriza leta 2009. Prav zato je modeliranje verjetnosti propada kreditojemalcev postalo ključnega pomena. Zaradi vpliva bančništva na celotno gospodarstvo je pravilna ocena tveganosti kreditojemalcev postala izjemnega pomena tudi za centralne banke. To je glavni razlog, da mora biti vsak novo razvit model pred uporabo predhodno odobren s strani centralne banke. Banke pa imajo proste roke pri izbiri specifičnega modela ter spremenljivk v njem. V nadaljevanju si bomo ogledali kreditno analizo podjetij ter uporabnost teorije za konstrukcijo modelov predvidevanja propada podjetij. Ogedali si bomo tudi ocene kreditnega tveganja, ki jih izdajajo bonitetne agencije, ter pokazali, zakaj je za banko kljub temu smotrno oblikovati lasten model.

V želji po enostavnosti modela verjetnosti propada podjetja bomo naprej začeli s probit in logit modelom. Oba sta namreč najpopularnejši metodi pri modeliranju diskretnih slučajnih spremenjivk z zveznimi. Torej imata vnaprej predvidljivo obliko, ki jo glede na vhodne podatke pričakujemo. Oba pa, tako kot linearna regresija, predpostavljata neodvisnost podatkov. Na primeru podatkov 3000 podjetij z ameriškega trga, ki so del indeksa Russell 3000, si bomo ogledali, kako dobro lahko napovemo propad podjetja z različnimi modeli. Kot neodvisne slučajne spremenljivke bodo v modelu nastopali kazalniki podjetij, ki jih lahko izračunamo s pomočjo podatkov iz računovodskih izkazov podjetja. V nadaljevanju bo prikazan tudi primer modeliranja s pomočjo SEM modela. Čeprav SEM model izhaja iz ekonometričnih krogov, je večjo pozornost pritegnil v psihologiji ter sociologiji, kjer imajo raziskovalci več problemov z latentnimi spremenljivkami. V zadnjem času pa se njegova popularnost vrača tudi v ekonometrične modele. Zanimalo nas bo, ali se osnovne težave multikolinearnosti v podatkih sploh pojavljajo, ter kako SEM model uporabiti. Splošna lastnost finančnih kazalnikov, s pomočjo katerih bodo modeli konstruirani je, da so njihove vrednosti zelo odvisne od industrijske panoge, v kateri se podjetje nahaja. Zato si bomo ogledali, kakšna je razlika, v kolikor v model vključimo učinek vrste industrije. Najenostavnejši izmed pristopov bi bila ločena statistična obdelava podatkov za vsako industrijo posebej, vendar na ta način izgubimo pomembne povezanosti med podjetji različnih industrij. Zato bomo SEM model razširili s HSEM modelom ter preverili dobljene rezultate. Poudarek pa bo tudi na samem prileganju podatkom. Primerjali bomo torej, koliko odstopajo naše napovedi (izpeljane s pomočjo napovednega modela) od realizacije v naslednjem letu.

V drugem poglavju magistrske naloge bomo spoznali teoretično ozadje linearnega strukturnega modela. V tretjem poglavju bo na kratko predstavljen posplošeni linearni mešani model. Njuna združitev, hierarhični linearni strukturni model, bo izpeljan v četrtem poglavju. Nato bomo našo pozornost v petem poglavju osredotočili na teorijo kreditne analize verjetnosti propada podjetij. V šestem poglavju bomo podjetja iz indeksa Russell 3000 uporabili za praktično izpelavo prej omenjenih modelov. V zaključku pa bomo interpretirali dobljene rezultate.

2 Linearni strukturni model

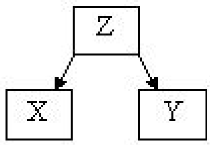
Zgodovina linearnega strukturnega modela se začne leta 1918 s prvo definicijo njegovega predhodnika: analize po poti (ang. path analysis). Definiral jo je genetik Sewall Wright, ko je modeliral velikost kosti zajcev. V psihologiji pa se je iz teorij obnašanja in sposobnosti razvila faktorska analiza (ang. factor analysis). Kot oče faktorske analize se največkrat navaja Spearman (1904), čeprav je nekaj podobnega objavil že Pearson (1901). Šele sedemdseta leta pa so prinesla združitev ter pravi razvoj tega, kar danes poznamo kot metodo linearnega strukturnega modela (SEM-ang. structural equation model). Vse se je začelo, ko je Jöreskog leta 1970 objavil posplošeno metodo za analizo kovariančnih struktur. Ključna članka, objavljena tem času, sta Hauserjev ter Goldbergerjev iz leta 1971 ter Jöreskogov iz leta 1973. Prav Jöreskog pa je leta 1976 prvi razvil računalniški program (LISREL- linear structural relations) za empirično uporabo modela. Slednje je metodi SEM prinašalo vedno večjo razpoznavnost in priljubljenost. Čeprav so prvi članki o tej metodi izhajali iz ekonometričnih krogov, je metoda veliko večjo pozornost pritegnila v psihologiji in sociologiji, šele v zadnjem času pa se njen pomen veča tudi v ekonometriji.

Kaj torej sploh je metoda SEM? Največkrat se jo omenja kot alternativno, v literaturi že dobro ustaljeno multiplo regresijo (GLM - ang. general linear model). Velike težave pri slednji pa nastanejo zaradi multikolinearnosti ter neupoštevanja vseh odnosov med spremenljivkami.

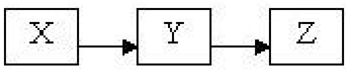
2.1 Odnosi med spremenljivkami

Prvi korak pri metodi SEM je natančna specifikacija odnosov med spremenjivkami in njihova predstavitev za vizualizacijo modela. Na diagramu moramo namreč prikazati vse odnose povezanosti (kovariranosti) ter odvisnosti (vzročnosti). Ločimo naslednje tipe povezanosti:

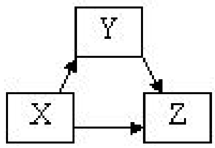
- *navidezna povezanost* med spemenljivkama X in Y nastane zaradi skupne vzročne spremenljivke Z, ki vpliva na vsako posebej;



- *direktna ter indirektna vzročna povezanost* slednja nastane, ko spremenljivka X vpliva na spremeljivko Z preko spremenljivke Y, ki jo imenujemo tudi intervenirajoča spremenljivka; torej ali

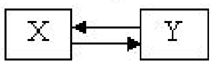


ali pa

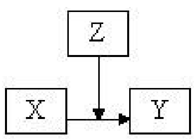


.

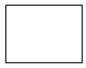
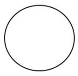
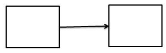
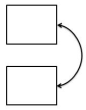
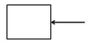
- *recipročna vzročna povezanost* pomeni, da spremenljivki vzročno vplivata druga na drugo, navadno vsaka z različnim učinkom;



- *pogojna povezanost* je takrat, ko na povezavo dveh spremenljivk vpliva tretja; torej sta spremenljivki povezani le v primeru, ko se tretja zgodi.



Do natančne opredelitve odnosov raziskovalec pride s preučevanjem teorije. V naslednjem koraku pa s pomočjo modela SEM preveri, ali so domneve pravilne. Za predstavitev modelov bomo uporabljali standardno notacijo (uporabljajo jo tudi programski jeziki, kot so LISREL ter MPlus) ter označili z:

- kvadratom vse direktno opazovane spremenljivke, 
- krogom vse latentne spremenljivke, 
- enojno puščico direktno povezanost , 
- dvojno puščico kovariiančno odvisnost, 
- s puščico, ki kaže proti spremenljivki, pa predstavimo napako. 

2.2 Specifikacija modela

Ideja metode SEM je, da iz poznanih koreliranih podatkov generiramo latentne spremenljivke. V naslednjem koraku pa opazujemo odnose med latentnimi spremenljivkami ter odvisnimi spremenljivkami. Na ta način lahko z metodo SEM izvedemo regresijo tudi na zelo koreliranih podatkih in hkrati opazujemo tako direktne kot tudi indirektne povezave med neodvisnimi ter odvisnimi spremenljivkami. Ideja SEM modela je torej podobna ideji faktorske analize z razliko, da SEM lahko vključuje več odvisnosti spremenljivk na različnejših mestih (npr. med opazovanimi ter neopazovanimi spremenljivkami). Hkrati pa lahko z vključitvijo metode preučevanja po poti jasno nakažemo odnose med spremenjivkami in s tem intiutivno predstavimo dogajanje. Velika razlika je tudi v uporabnosti same metode, saj metoda SEM ni namenjena raziskovanju katere spremenljivke bi bilo najbolje vključiti v model, temveč pomagati raziskovalcu testirati ali spremenljivke vplivajo ena na drugo tako, kot je opisano v modelu. Odnose med spremenljivkami preverja z raziskovanjem variančno-kovariančnih matrik.

Osnovni SEM model, ki ga je predstavil Jöreskog (1973), je definiran s tremi enačbami:

$$\boldsymbol{\eta} = B\boldsymbol{\eta} + \Gamma\boldsymbol{\xi} + \boldsymbol{\zeta} \tag{1}$$

$$\boldsymbol{y} = \Lambda_{\boldsymbol{y}}\boldsymbol{\eta} + \boldsymbol{\varepsilon} \tag{2}$$

$$\boldsymbol{x} = \Lambda_{\boldsymbol{x}}\boldsymbol{\xi} + \boldsymbol{\delta} \tag{3}$$

kjer je $\boldsymbol{\eta}' = (\eta_1, \eta_2, \ldots, \eta_m) \in \mathbb{R}^{1\times m}$ vektor latentnih faktorjev, $\boldsymbol{\xi}' = (\xi_1, \xi_2, \ldots, \xi_n) \in \mathbb{R}^{1\times n}$ vektor neodvisnih latentnih spremenljivk. Ker vektorja $\boldsymbol{\eta}$ ter $\boldsymbol{\xi}$ ne moremo direktno opazovati, ju opišemo z neodvisno spremenljivko $\boldsymbol{x}' \in \mathbb{R}^{1\times q}$ ter odvisno $\boldsymbol{y}' \in \mathbb{R}^{1\times p}$. Matrike $B \in \mathbb{R}^{m\times m}$, $\Gamma \in \mathbb{R}^{m\times n}$, $\Lambda_{\boldsymbol{x}} \in \mathbb{R}^{q\times n}$ ter $\Lambda_{\boldsymbol{y}} \in \mathbb{R}^{p\times m}$ so matrike koeficientov. Poleg tega morajo biti $\boldsymbol{\varepsilon}$ ter $\boldsymbol{\delta}$ neodvisne od $\boldsymbol{\zeta}$, $\boldsymbol{\eta}$ ter $\boldsymbol{\xi}$. Še zadnja predpostavka, ki jo moramo sprejeti je, da je matrika $B - I$ nesingularna. Brez škode za splošnost smo v tem zapisu predpostavljali, da so spremenljivke merjene v odstopanju od povprečja (torej da so centrirane v 0). Model SEM namreč deluje na način, da preverja odnose med spremenjivkami s pomočjo njihovih variančno kovariančnih matrik, torej samo centriranje nanj ne vpliva.

Trditev 2.1. *Teoretična kovariančna matrika $\Sigma(\theta)$ je oblike*

$$\Sigma(\theta) = \begin{bmatrix} \Lambda_{\boldsymbol{y}}\Pi(\Gamma\Phi\Gamma' + \Psi)\Pi'\Lambda_{\boldsymbol{y}}' + \Theta_{\boldsymbol{\varepsilon}} & \Lambda_{\boldsymbol{y}}\Pi\Gamma\Phi\Lambda_{\boldsymbol{x}}' + \Theta_{\boldsymbol{\varepsilon\delta}} \\ \Lambda_{\boldsymbol{x}}\Phi\Gamma'\Pi'\Lambda_{\boldsymbol{y}}' + \Theta_{\boldsymbol{\varepsilon\delta}} & \Lambda_{\boldsymbol{x}}\Phi\Lambda_{\boldsymbol{x}}' + \Theta_{\boldsymbol{\delta}} \end{bmatrix}$$

Kjer z Φ označimo kovariančno matriko $\boldsymbol{\xi}$, Ψ kovariančno matriko $\boldsymbol{\zeta}$, $\Theta_{\boldsymbol{\varepsilon}}$ ter $\Theta_{\boldsymbol{\delta}}$sta kovariančni matriki $\boldsymbol{\varepsilon}$ in $\boldsymbol{\delta}$. Analogno še z $\Theta_{\boldsymbol{\delta\varepsilon}}$ označimo kovariančno matriko med $\boldsymbol{\delta}$ ter $\boldsymbol{\varepsilon}$. Za poenostavitev notacije z Π pišemo $(B-I)^{-1}$.

Dokaz. Oglejmo si vsak člen kovarančne matrike posebej.

$$\Sigma = \begin{bmatrix} \Sigma_{\boldsymbol{yy'}} & \Sigma_{\boldsymbol{yx'}} \\ \Sigma_{\boldsymbol{xy'}} & \Sigma_{\boldsymbol{xx'}} \end{bmatrix}$$

Najprej izračunajmo prvi člen $\Sigma_{\boldsymbol{yy'}}$:

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{yy'}) =& \mathrm{E}(\boldsymbol{yy'}) - \mathrm{E}(\boldsymbol{y})\mathrm{E}(\boldsymbol{y'}) \\ =& \mathrm{E}[(\Lambda_{\boldsymbol{y}}\Pi\Gamma\boldsymbol{\xi} + \Lambda_{\boldsymbol{y}}\Pi\boldsymbol{\zeta} + \boldsymbol{\varepsilon})(\boldsymbol{\xi}'\Gamma'\Pi'\Lambda_{\boldsymbol{y}}' + \boldsymbol{\zeta}'\Pi'\Lambda_{\boldsymbol{y}}' + \boldsymbol{\varepsilon}')] \\ =& \Lambda_{\boldsymbol{y}}\Pi(\Gamma\mathrm{E}[\boldsymbol{\xi\xi'}]\Gamma + \Gamma\mathrm{E}[\boldsymbol{\xi\zeta'}] + \mathrm{E}[\boldsymbol{\zeta\xi'}]\Gamma' + \mathrm{E}[\boldsymbol{\zeta\zeta'}])\Pi'\Lambda_{\boldsymbol{y}}' \\ &+ \Lambda_{\boldsymbol{y}}\Pi(\Gamma\mathrm{E}[\boldsymbol{\xi\varepsilon'}] + \mathrm{E}[\boldsymbol{\zeta\varepsilon'}]) + (\mathrm{E}[\boldsymbol{\varepsilon\xi'}]\Gamma' + \mathrm{E}[\boldsymbol{\varepsilon\zeta'}])\Pi'\Lambda_{\boldsymbol{y}}' + \mathrm{E}[\boldsymbol{\varepsilon\varepsilon'}] \\ =& \Lambda_{\boldsymbol{y}}\Pi(\Gamma\Phi\Gamma' + \Psi)\Pi'\Lambda_{\boldsymbol{y}}' + \Theta_{\boldsymbol{\varepsilon}} \end{aligned}$$

Prva enakost je definicija kovariance. Iz specifikacije začetnega modela vidimo $\mathrm{E}(\boldsymbol{y}) = 0$ (predpostavljali smo, da so spremenljivke centrirane v nič). Dejstvo, da je $\mathrm{E}[\boldsymbol{\xi\zeta'}] = \mathrm{E}[\boldsymbol{\zeta\xi'}] = \mathrm{E}[\boldsymbol{\varepsilon\zeta'}] = \mathrm{E}[\boldsymbol{\varepsilon\xi'}] = \mathrm{E}[\boldsymbol{\zeta\varepsilon'}] = \mathrm{E}[\boldsymbol{\xi\varepsilon'}] = 0$ sledi iz predpostavk o nekoreliranosti. Ker sta tudi $\boldsymbol{\xi}$ ter $\boldsymbol{\varepsilon}$ centrirana v nič, je varianca enaka pričakovani vrednosti produkta.
Za vrednost $\Sigma_{\boldsymbol{yx'}}$ vidimo, da je enaka:

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{yx'}) &= \mathrm{E}(\boldsymbol{yx'}) - \mathrm{E}(\boldsymbol{y})\mathrm{E}(\boldsymbol{x'}) \\ &= \mathrm{E}[(\Lambda_{\boldsymbol{y}}\Pi\Gamma\boldsymbol{\xi} + \Lambda_{\boldsymbol{y}}\Pi\boldsymbol{\zeta} + \boldsymbol{\varepsilon})(\boldsymbol{\xi}'\Lambda_{\boldsymbol{x}}' + \boldsymbol{\delta}')] \\ &= \Lambda_{\boldsymbol{y}}\Pi(\Gamma\mathrm{E}[\boldsymbol{\xi\xi'}]\Lambda_{\boldsymbol{x}}' + \Gamma\mathrm{E}[\boldsymbol{\xi\delta'}] + \mathrm{E}[\boldsymbol{\zeta\xi'}]\Lambda_{\boldsymbol{x}}' + \mathrm{E}[\boldsymbol{\zeta\delta'}]) + \mathrm{E}[\boldsymbol{\varepsilon\xi'}]\Lambda_{\boldsymbol{x}}' + \mathrm{E}[\boldsymbol{\varepsilon\delta'}] \\ &= \Lambda_{\boldsymbol{y}}\Pi\Gamma\Phi\Lambda_{\boldsymbol{x}}' + \Theta_{\boldsymbol{\varepsilon\delta}} \end{aligned}$$

Iz predpostavk vidimo, da je $\mathrm{E}[\boldsymbol{\xi\delta'}] = \mathrm{E}[\boldsymbol{\zeta\xi'}] = \mathrm{E}[\boldsymbol{\varepsilon\xi'}] = \mathrm{E}[\boldsymbol{\zeta\delta'}] = 0$. Podobno lahko izračunamo še $\Sigma_{\boldsymbol{xy'}}$. Na tem mestu sedaj izračunajmo še $\Sigma_{\boldsymbol{xx'}}$.

$$\begin{aligned} \mathrm{Cov}(\boldsymbol{x},\boldsymbol{x}) &= \mathrm{E}(\boldsymbol{xx'}) - \mathrm{E}(\boldsymbol{x})\mathrm{E}(\boldsymbol{x'}) \\ &= \Lambda_{\boldsymbol{x}}\mathrm{E}[\boldsymbol{\xi\xi'}]\Lambda_{\boldsymbol{x}}' + \Lambda_{\boldsymbol{x}}\mathrm{E}[\boldsymbol{\xi\delta'}] + \mathrm{E}[\boldsymbol{\delta\xi'}]\Lambda_{\boldsymbol{x}}' + \mathrm{E}[\boldsymbol{\delta\delta'}] \\ &= \Lambda_{\boldsymbol{x}}\Phi\Lambda_{\boldsymbol{x}}' + \Theta_{\boldsymbol{\delta}} \end{aligned}$$

Tudi tu dejstvo, da je $\mathrm{E}[\boldsymbol{\xi\delta'}] = \mathrm{E}[\boldsymbol{\delta\xi'}] = 0$, sledi iz predpostavk. Če sestavimo vse dele v kovariančno matriko, dobimo ravno matriko iz trditve. □

Naj bo $\boldsymbol{z} \equiv (\boldsymbol{y'} : \boldsymbol{x'})$ ter $Z = (\boldsymbol{z}_1, \boldsymbol{z}_2, \ldots, \boldsymbol{z}_N)$ vzororec z N opazovanji na $\boldsymbol{z}$. Kot običajno vzamememo povprečno vrednost opazovanj $\bar{\boldsymbol{z}} = \frac{1}{N}\sum_{i=1}^{N} \boldsymbol{z}_i$ ter varianco vzorca:

$$S = \frac{1}{N-1}\sum_{i=1}^{N}(\boldsymbol{z}_i - \bar{\boldsymbol{z}})(\boldsymbol{z}_i - \bar{\boldsymbol{z}})'.$$

Parametri modela ($\boldsymbol{\theta}$) so nato dobljeni tako, da minimiziramo neskladje med variančno-kovariančno matriko vzorca S ter teoretično variančno kovariančno matriko $\Sigma(\boldsymbol{\theta})$. To nas na nek način spominja na ocenjevanje s pomočjo metode najmanjših kvadratov, z razliko, da so opazovanja, ki nastopajo, posamezni izidi, medtem ko imamo pri SEM metodi opraviti z variancami ter kovariancami. Najbolj poznana ter vse do današnjih dni tudi najpopularnejša metoda ocenjevanja parametrov pri metodi SEM, je metoda največjega verjetja (ang. maximum likelihood). Funkcijo prileganja, ki jo poda Bollen (1989), zapišemo:

$$F_{\mathrm{ML}} = \ln|\Sigma(\theta)| + \mathrm{sl}(S\Sigma^{-1}(\boldsymbol{\theta})) - \ln|S| - (p+q). \tag{4}$$

V enačbi (4) je ln naravni logaritem, sl je funkcija sledi (definirana za poljubno matriko $A = (a_{ij})_{i=1,j=1}^{N,N}$ kot $\mathrm{sl}(A) = \sum_{k=1}^{N} a_{kk}$) ter $(p+q)$ število opazovanih spremenjivk. Če predpostavimo multivariatno normalno porazdelitev opazovanih spremenljivk ter pravilno specifikacijo modela, je ML ocena konsistentna, nepristranska, najbolj učinkovita ter normalno porazdeljena.
Nekoliko bolj splošna funkcija prilagajanja je dobljena preko metode najmanjših kvadratov (ang. generalized least squares) in je oblike:

$$F_{\mathrm{GLS}} = (S - \Sigma(\boldsymbol{\theta}))' W^{-1} (S - \Sigma(\boldsymbol{\theta})). \tag{5}$$

Tu je W^{-1} matrika uteži.

2.3 Natančnost prileganja

Vsak model je le slika, prav zato noben od njih ne more povsem replicirati realnosti. Vemo pa, da se različni modeli različno dobro približujejo modeliranim podatkov. Prav zato želimo, ko so parametri ocenjeni, znati povedati ali obdržati model ali pa ga spremeniti. Gre torej za statistično testiranje hipoteze, da se dan model prilega podatkom. Ker se pri metodi SEM ukvarjamo z minimizacijo kovariančne matrike, seveda ne moremo uporabiti mer, kot so R^2 ali modificiran R^2 (te namreč temeljijo na nepojasnjeni varianci - torej varianci spremenljivke napak ter totalni varianci). Prav zato je izbira primerne mere prileganja ena glavnih debat raziskovalcev vse od prve definicije modela SEM. Potrebno je izpostaviti, da tudi danes nimamo testa, ki bi bil brez težav, prav zato se navadno poda več mer prileganja.

Tako kot v večini primerov, tudi tu obstajajo številna pravila "preko palca" za primerne rezultate modela. Eno najbolj znanih je, da potrebujemo bazo z vsaj dvestotimi opazovanji, če imamo opravka z multivariatno normalno porazdeljenimi podatki. Vendar si bomo tu ogledali vsaj še nekaj natančnejših kazalcev, ki jih v grobem razdelimo na tiste, pri katerih spremljamo absolutne vrednosti, ter primerjalne, s pomočjo katerih primerjamo dva modela med sabo. Pri slednjih ne moremo reči, da model lahko sprejmemo, saj vemo le, da se obnaša bolje kot nek drugi model. Prav zato si oglejmo najprej absolutne mere.

2.3.1 Hi-kvadrat test -χ^2

Hi-kvadrat test je tradicionalna mera prileganja, saj je dobljena preprosto kot

$$\chi^2 = (N-1)(\ln|\Sigma(\theta)| + \mathrm{sl}(S\Sigma^{-1}(\boldsymbol{\theta})) - \ln|S| - (p+q)), \tag{6}$$

kjer je N velikost vzorca. Vrednosti se ob pravilni specifikaciji modela porazdeljuje po χ^2 porazdelitvi z $\frac{k(k+1)}{2} - t$ stopinjami prostoti (tu je t število parametov, ki jih ocenjujemo, ter $k = p + q$ število opazovanih spremenljivk). Z njeno pomočjo testiramo ničelno hipotezo $\mathrm{H}_0 : \Sigma = S$, torej da se model prilega podatkom. Hi-kvadrat je edina mera, za katero poznamo porazdelitev in tako lahko izračunamo p-vrednosti.

Kljub temu pa nas mera hi-kvadrat lahko hitro pripelje do napačnih zaključkov. V kolikor model priredi vrednosti, ki se prilegajo danim podatkom, potem se bosta v enačbi (6) člena $\ln|\Sigma(\theta)|$ ter $\ln|S|$ odštela med sabo. Če se empirična ter teoretična kovariančna matrika ujemata, je produkt obratne vrednosti teoretične ter empirične kovariančne matrike ravno identiteta. Vrednost funkcije sledi je v tem primeru vsota enic, torej $(p + q)$. Torej model, ki se popolnoma prilega podatkom, vrne dobre razultate. Izkaže pa se, da je pri majhnih odstopanjih od idelane lege vrednost hi-kvadrat testa skoraj N. Torej, čeprav se model razumno dobro prilega podatkom, zavrnemo ničelno hipotezo. Druga težava testa je, da se vrednosti χ^2 manjšajo z vključevanjem novih, tudi nepomembnih spremenljivk.

2.3.2 Koren povprečne kvadratne napake

Tako kot χ^2 , tudi koren povprečne kvadratne napake (RMR - ang. root mean square residual), meri slabost prileganja, torej vrednost nič predstavlja popolno prileganje. RMR ter standardiziran RMR (SRMR) izračunamo na način:

$$\mathrm{RMR} = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}(s_{ij} - \sigma_{ij})^2}{k(k+1)/2}},$$

tako kot prej je $k = p + q$ ter

$$\mathrm{SRMR} = \sqrt{\frac{\sum_{i=1}^{p}\sum_{j=1}^{i}[(s_{ij} - \sigma_{ij})/(s_{ii}s_{jj})]^2}{k(k+1)/2}}.$$

Tu smo predpostavljali $\Sigma = (\sigma)_{i=1,j=1}^{n,n}$. V kolikor se model dobro prilega podatkom, je $(s_{ij} - \sigma_{ij})$ skoraj nič, torej sta tako RMR kot tudi SRMR blizu nič.

V literaturi se v večini primerov uporablja SRMR. Za slednjega sta Hu in Bentler (1999) pokazala, da je bolj občutljiv za napačno specifikacijo modela, hkrati pa robustnejši glede za velikosti vzorca ter napačne predpostavke porazdelitve spremenljivke. To ga vsekakor favorizira pred RMR testom prileganja. Prav tako pa sta podala oceno za dobro prileganje, ki naj bi bilo dosežena za SRMR največ 0.9. Kljub vsemu pa ima SRMR težave z v osnovi nižjo vrednostjo za modele z veliko parametri ter velikimi velikostmi vzorcev.

2.3.3 RMSEA

Koren povprečene kvadratne napake aproksimacije (RMSEA - ang. root mean square error of approximation) je mera, ki zveni zelo podobno kot RMR, vendar sta tako njenen izračun kot tudi lastnosti drugačni. Čeprav sta prvo definicijo podala že Steiger and Lind (1980), ostaja do danes enden najpomembnejših idenksov, ki ga ob rezultatih SEM modela podajo vsi računalniški programi. Izračunamo jo kot

$$\text{RMSEA} = \max\{\sqrt{(\chi^2 - df)/df(N-1)}, 0\},$$

kjer so df stopinje prostosti. V kolikor je torej $\chi^2 \leq df$, se vrednost RMSEA postavi na nič, saj je le tu doseženo popolno prileganje. RMSEA favorizira enostavnejše modele, torej modele z manj parametri (z dodajanjem le teh se nam namreč veča χ^2/df). Hu in Bentler (1999) podata tudi za vrednosti RMSEA meje za prileganja in sicer za vrednosti RMSEA ≤ 0.05 pravimo, da se model dobro prilega podatkom. Pri tem indeksu prileganja je mogoče izračunati še interval zaupanja, kar nam da veliko bolj informativne rezultate. Za 90% interval zaupanja je idealno, da vključuje tudi nič (lahko je tudi samo zelo blizu) ter da zgornja meja ne presega 0.08. Kljub vsemu pa nam RMSEA poda pristranske ocene za majhne velikosti vzoorca ($N < 250$).

2.3.4 Akaike informacijski kriterij

Akaike informacijski kriterij (AIC - ang. akaike information criterion) je prvi ter edini, ki ga bomo v tem delu predstavili iz skupine primerjalnih indeksov. Edini pomen poda, ko primerjamo dva različna modela pri čemer ima boljši model nižji AIC. V tem indeksu primerjamo razliko

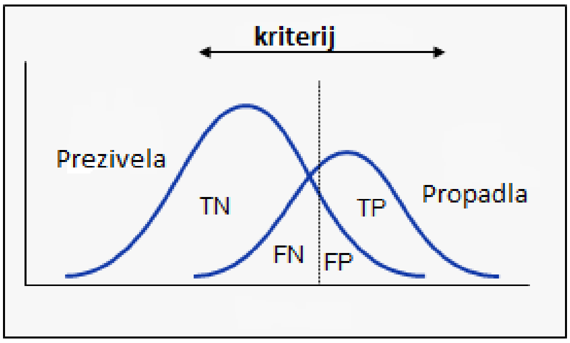
$$\text{AIC} = \chi^2 + k(k+1) - 2df.$$

Opozoriti je potrebno, da je $k(k+1) - 2df$ ravno enak dvakratniku števila prostih parametrov v modelu. Tako se torej vrednost AIC poveča z vsako dodatno spremenljivko.

2.3.5 Krivulja ROC ter Ginijev koeficient

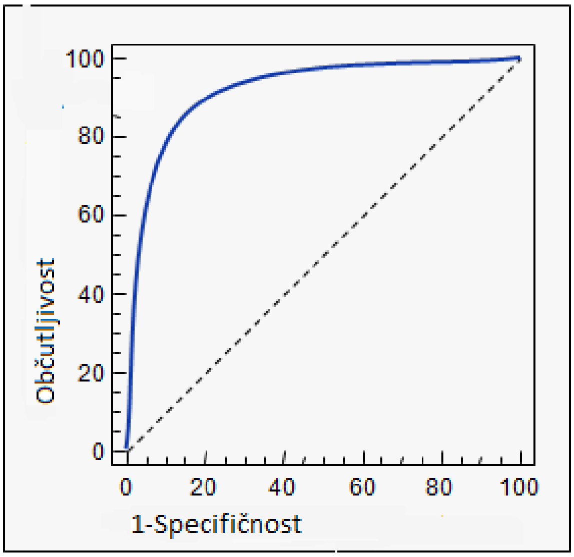
Problem mer prileganja, ki smo jih do sedaj spoznali, je v njihovi specifičnosti. Izračunamo jih lahko le v primeru, ko uporabljamo SEM model. V želji po meri, ki bo primerljiva za vse metode (od probit modela do SEM modela ter hierarhičnega SEM modela, ki ga bomo izpeljali v nadaljevanju), opozorimo na tem mestu še na mero prileganja, ki je v tem pomenu splošna. Oglejmo si krivuljo ROC (včasih imenovano tudi Lorenzova krivulja, v angleščini imenovana receiver operating characteristic) ter Ginijev koeficient. Čeprav nista tipični meri prileganja SEM modela, pa ju v literaturi pogosto zasledimo kot meri natančnosti modela z napovedovalno močjo. Glavna ideja je, da primerjata napovedano verjetnost v naslednjem obdobju z reliziranim stanjem. V našem primeru bomo primerjali verjetnost propada za naslednje leto z dejanskim stanjem, torej z binomsko spremenljivko, ki bo pokazala, ali je podjetje propadlo ali še deluje.

Krivulja ROC je bila prvič definirana v času druge svetovne vojne za pomoč analize signalov oddajnikov. Gre za dvodimenzionalni graf, ki prikazuje natančnost prileganja kreirane odvisne slučajne spremenljivke napovedi z binomsko spremenljivko, podano s trga, kar bo v našem primeru spremenljivka propada. Za spremenljivko napovedi pa dobimo bernulijevo spremenljivko. Spremenljivki, v našem primeru napovedi ter realizacije, pa morata biti neodvisni, sicer test nima pomena. V spod-



Slika 2: Grafična predstavitev vrednosti TP, FP, FN ter TN.

njem grafu vidimo primer, kjer so opazovanja porazdeljena normalno tako za propadla podjetja kot tudi za preživela podjetja. Normalnost podatkov za izvajanje testa ni zahteva, na tem mestu pa je privzeta zaradi jasnosti interpretacije. Vzorec preživelih podjetji je navadno bistveno večji od vzorca propadlih, kar vpliva na velikosti krivulj. Ta neenakost v krivuljah je tu le nakazana in je ponavadi veliko večja. V kolikor izberemo naključno neko vrednost kriterija propada, potem lahko zanjo poračunamo pravilno napovedan propad (TP ang. true positive fraction) in napačno napovedan propad (FN ang. false negative fraction). Po drugi strani pa imamo primere, ko je pravilno določeno preživetja podjetja (TN ang. true negative fraction) ter napačno izračunano preživetje podjetja (FP ang. false positive fraction), kot je to prikazano na Sliki 2. (Zweig & Campbell, 1993). Sedaj lahko izračunamo točko na ROC krivulji kot $(\frac{FP}{FP+TN}, \frac{TP}{TP+FN})$. Z izračunom vseh možnih vrednosti kriterija nato dobimo ROC krivuljo. Na ordinatni osi imamo podano občutljivost modela (delež podjetij, ki so propadle in jih model prepozna kot take), na abscisni pa vrednosti $1 -$ specifičnost modela. Specifičnost je torej delež podjetij, ki niso propadla ter jim tega tudi naš model ni napovedal. Pri modeliranju je cilj, da se ROC krivulja čimbolj približa levemu zgornjemu kotu. Optimalna krivulja je predstavljena na



Slika 3: Optimalna ROC krivulja.

Sliki 3 s polno črto, medtem ko je s črtkano črto predstavljen primer naključnega modeliranja.

Žal pa nam krivulja ne poda nobene numerične vrednosti. Prav zato se navadno izračuna skupaj z Ginijevim koeficientom, ki predstavlja razmerje med ploščino pod ROC krivuljo modela ter ROC krivuljo naivnega modela. Ginijev koeficient je pozitiven natanko tedaj, ko je večji del krivulje prileganja nad prileganjem naključnega modela, torej nad črtkano črto. To zapišemo z enačbo:

$$\text{Gini} = \frac{\int_0^1 G(t)dt - \frac{1}{2}}{\int_0^1 G_r(t)dt - \frac{1}{2}},$$

kjer je $G(t)$ funkcija ROC krivulje modela, $G_R(t)$ pa funkcija ROC krivulje naključnega modela. Zaradi poenostavitve pa se v večini uporablja enačba:

$$\text{Gini} \approx 2\int_0^1 G(t)dt - 1.$$

Maksimalna vrednost Ginijevega koeficienta je ena, kar prikazuje popolno prileganje modela. To enačbo bomo uporabljali tudi pri naših izračunih. Ker lahko Ginijev koeficient ter ROC krivujo podamo za vsako metodo, v kolikor imamo podane napovedi ter realizacijo, bosta ti dve meri prileganja ključni skozi potek celotne magistrske naloge.

2.4 Ocenjevanje vrednosti latentnih spremenjivk

Metodo SEM bomo uporabljali za konstukcijo faktorjev iz finančnih spremenljivk, ki jih najdemo v izkazih podjetij. Te faktorje bomo obravnavali kot latentne spremenjivke. Za vključitev latentne spremenljivke v GLMM metodo, ki bo naslednji korak, pa potrebujemo vrednosti te spremenljivke za vsako podjetje v vzorcu. V poglavju o specifikaciji modela smo opisali samo, kako določimo porazdelitev latentnih spremenljivk. V tem delu pa pokažemo, kako s podanimi matrikami ocenjenega modela poiščemo napovedi za vsako opazovanje. Za poenostavitev zapisa enačb (1) do (3) vpeljimo zapis:

$$\boldsymbol{Y} := [\boldsymbol{y}'\,\boldsymbol{x}']',\ \Lambda := \begin{bmatrix} \Lambda_{\boldsymbol{y}} & 0 \\ 0 & \Lambda_{\boldsymbol{x}} \end{bmatrix},\ \boldsymbol{Z} := [\boldsymbol{\eta}'\,\boldsymbol{\xi}']',\ \boldsymbol{\epsilon} := [\boldsymbol{\varepsilon}'\,\boldsymbol{\delta}']' \tag{7}$$

in $\Phi = \mathrm{E}[\boldsymbol{\xi}\boldsymbol{\xi}']$, $\Theta = \mathrm{E}[\boldsymbol{\epsilon}\boldsymbol{\epsilon}']$. Enačbi (2) ter (3) torej prepišemo v

$$\boldsymbol{Y} = \Lambda\boldsymbol{Z} + \boldsymbol{\epsilon}.$$

Rešitev po metodi največjega verjetja za $\boldsymbol{Z}$ iz modela dobimo z minimizacijo vsote kvadratov $\sum_{I=1}^{N}(\boldsymbol{Y}_i - \Lambda\boldsymbol{Z}_i)\Theta^{-1}(\boldsymbol{Y}_i - \Lambda\boldsymbol{Z}_i)$. Tu smo z $\boldsymbol{Z}_i, \boldsymbol{Y}_i$ $i = 1, \ldots, N$ označili spremenljivke posameznih opazovanj, število opazovanj pa z N. Po metodi najmanjših kvadratov dobimo oceno za $\boldsymbol{Z}_i$, podano z $\hat{\boldsymbol{Z}}_i = (\Lambda\Theta^{-1}\Lambda)^{-1}\Lambda'\Theta^{-1}\boldsymbol{Y}_i$, ki je sicer najboljša linearna neodvisna cenilka, vendar nam žal ne zagotovlja, da bo kovariančna matrika ocen enaka teoretični matriki latentnih spremenljivk.

Za zagotovilo slednjega tvorimo Lagrangeovo funkcijo

$$L = \sum_{i=1}^{N}(Y_i - \Lambda Z_i)\Theta^{-1}(Y_i - \Lambda Z_i) + \mathrm{sl}\left[X\left(\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{Z}_i' - N\Phi\right)\right]$$

kjer je X simetrična matrika Lagrangeovih multiplikatorjev. V kolikor odvajamo po $\boldsymbol{Z}_i$ ter enačimo z nič, dobimo analitično rešitev podano z

$$\boldsymbol{Z}_i = (\Lambda'\Theta^{-1}\Lambda + X)^{-1}\Lambda'\Theta^{-1}\boldsymbol{Y}_i, \tag{8}$$

ki vsebuje matriko neznanih Lagrangevih multiplikatorjev X. V nadaljevanju preoblikujemo enačbo tako, da pridemo do enačbe s samo znanimi matrikami. Prvi korak je kvadriranje enačbe (8)

$$\boldsymbol{Z}_i\boldsymbol{Z}_i' = (\Lambda'\Theta^{-1}\Lambda + X)^{-1}\Lambda'\Theta^{-1}\boldsymbol{Y}_i\boldsymbol{Y}_i'\Theta^{-1}\Lambda(\Lambda'\Theta^{-1}\Lambda + X)^{-1}.$$

Sedaj seštejemo po vseh i in delimo z N, ter tako dobimo:

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{Z}_i' = (\Lambda'\Theta^{-1}\Lambda + X)^{-1}\Lambda'\Theta^{-1}\left(\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Y}_i\boldsymbol{Y}_i'\right)\Theta^{-1}\Lambda(\Lambda'\Theta^{-1}\Lambda + X)^{-1}.$$

Iz predpostavk vemo, da je $E[\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{Z}_i] = \Sigma_{\boldsymbol{\eta\xi}}$, torej lahko enačbo prepišemo v obliko

$$\Sigma_{\boldsymbol{\eta\xi}} = (\Lambda'\Theta^{-1}\Lambda + X)^{-1}\Lambda'\Theta^{-1}A\Theta^{-1}\Lambda(\Lambda'\Theta^{-1}\Lambda + X)^{-1}$$

kjer je $A := \frac{1}{N}\sum_{i=1}^{N}E[\boldsymbol{Y}_i\boldsymbol{Y}_i']$. Množimo iz leve ter desne z $(\Lambda\Theta^{-1}\Lambda + X)$ ter dobimo

$$(\Lambda'\Theta^{-1}\Lambda + X)\Sigma_{\boldsymbol{\eta\xi}}(\Lambda'\Theta^{-1}\Lambda + X) = \Lambda'\Theta^{-1}A\Theta^{-1}\Lambda.$$

V naslednjem koraku uporabimo SVD razcep matrike $\Sigma_{\boldsymbol{\eta\xi}} = UDU'$ ter enačbo množimo na sledeč način

$$D^{\frac{1}{2}}U'(\Lambda'\Theta^{-1}\Lambda + X)UD^{\frac{1}{2}}D^{\frac{1}{2}}U'(\Lambda'\Theta^{-1}\Lambda + X)UD^{\frac{1}{2}} = D^{\frac{1}{2}}U'\Lambda'\Theta^{-1}A\Theta^{-1}\Lambda UD^{\frac{1}{2}}.$$

Sedaj vzemimo še SVD razcep matrike

$$D^{\frac{1}{2}}U'(\Lambda'\Theta^{-1}\Lambda + X)UD^{\frac{1}{2}} = VL^{\frac{1}{2}}V', \tag{9}$$

in ga vstavimo v enačbo, ki jo prepišemo kot

$$D^{\frac{1}{2}}U'\Lambda'\Theta^{-1}A\Theta^{-1}\Lambda UD^{\frac{1}{2}} = VLV'.$$

Iz enačbe (9) sklepamo, da je

$$(\Lambda\Theta^{-1}\Lambda + X)^{-1} = D^{\frac{1}{2}}U'VL^{\frac{1}{2}}V'UD^{\frac{1}{2}},$$

kar nam skupaj z enačbo (8) da željen rezultat.

Torej smo pravkar pokazali sledeče:

Trditev 2.2. *Cenilka za vrednost latentnih spremenljivk (scores) v modelu (1)-(3), ki zadošča pogoju, da je kovarinčna matrika ocen enaka teoretični kovariančni matriki lantentnih spremenljivk, torej da je*

$$\frac{1}{N}\sum_{i=1}^{N}\boldsymbol{Z}_i\boldsymbol{Z}_i' = \Sigma_{\boldsymbol{\eta\xi}}$$

je dana kot

$$\bar{\boldsymbol{Z}}_i = D^{\frac{1}{2}} U' V L^{-\frac{1}{2}} V' U D^{\frac{1}{2}} \Lambda' \Theta^{-1} \boldsymbol{Y}_i.$$

Tu je

$$\Sigma_{\boldsymbol{\eta}\boldsymbol{\xi}} = \begin{bmatrix} (I-B)^{-1}(\Gamma\Phi\Gamma' + \Psi)\left((I-B)^{-1}\right)' & (I-B)^{-1}(\Gamma\Phi) \\ (\Phi\Gamma')\left((I-B)^{-1}\right)' & \Phi \end{bmatrix}.$$

Model SEM predpostavlja neodvisnost med vsemi latentnimi ter neodvisnimi spremenljivkami, ki nastopajo v modelu. To je seveda kršeno, če imamo opazovanja razporejena po skupinah glede na določeno lastnost. Najlažje se s problemom soočimo tako, da za vsako skupino posebej izvedemo SEM model. Na ta način pa žal izgubimo odnose med skupinami, prav tako ne moremo graditi večplastnih modelov. Zato si bomo v nadaljevanju ogledali metode, ki omogočajo obravnavo podatkov po skupinah.

3 Posplošeni linearni mešani model

Večrazsežne modele regresije uporabljamo, ko so naši podatki v hierarhični obliki. To pomeni, da so elementi v prvi ravni razporejeni v skupine na drugi ravni, le ta potem v večje skupine na tretji ter tako naprej. V tem kratkem poglavju si bomo ogledali osnovno izpeljavo enega izmed večrazsežnih modelov, in sicer posplošitev multiple regresije (ang. generalized linear model - GLM), to je posplošene linearne mešane modele (ang. generalized linear mixed model - GLMM). Model bomo nato v naslednjem poglavju razširili ter dodelali.

Tako kot GLM tudi GLMM dopušča, da je odvisna spremenjivka v katerikoli izmed oblik znanih porazdelitev. Njegova glavna prednost pa je ravno večrazsežnost, ki jo dosega z vključitvijo slučajnih učinkov (ang. random effects). V vsakem nivoju zapišemo enačbo kot:

$$g(\mu_{ij}) = \nu_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sum_{m=1}^{M_j} \eta^{(2)}_{mj} z^{(2)}_{mij}. \tag{10}$$

Tu je $\mu_{ij} := E[y_{ij}|\boldsymbol{x}_{ij}, \eta^{(2)}_{mj}, z^{(2)}_{mij}]$ ter $\eta^{(2)}_j = (\eta^{(2)}_{1j}, \dots, \eta^{(2)}_{M_j j})$ vektor slučajnih učinkov v drugi ravni s pripadajočimi vrednostimi kontrolnih spremenljivk (ang. covariates) $z^{(2)}_{mij}$. Model lahko zapišemo tudi v matrični obliki za poljubno število ravni kot:

$$g(\boldsymbol{\mu}_i) = X_i\boldsymbol{\beta} + Z_i\boldsymbol{\eta}_i.$$

Za slučajne učinke $\boldsymbol{\eta}_j \in \mathbb{R}^{p\times 1}$ privzamemo, da so normalno porazdeljene z upanjem 0 ter varianco Ψ. Prav zaradi te postavke so torej slučajni učinki izraženi kot odstopanje od povprečja, ki ga zajamemo v ustaljenih učinkih $X_i\boldsymbol{\beta}$ (ang. fixed effects). Tu sta $X_i \in \mathbb{R}^{n_j \times p}$ ter $\boldsymbol{\beta} \in \mathbb{R}^{q\times 1}$. Funkcija $g(.)$ je vezna funkcija (ang. link function) s pomočjo katere modeliramo različne porazdelitve slučajne spremenljivke. Matrika Z je velikosti $n_j \times q$, pri čemer smo povsod privzeli, da je n_j število opazovanj na ravni j. Poudariti velja, da imamo opravka z regresijo glede na pogojno upanje $\boldsymbol{\mu}_i$ in ne direktno na odvisno spremeljivko $\boldsymbol{y}_i$. Pogojna porazdelitev $(\boldsymbol{y}_i|\boldsymbol{\eta})$ je neodvisna za vsak i, hkrati pa pripada eksponentni družini.

Vsaka posplošitev pa zahteva svojo ceno. Za GLMM ne moremo analitično izpeljati minimuma funkcije najboljšega prileganja (to lahko storimo le v primeru, ko predpostavimo normalnost odvisne spremenljivke ter za vezno funkcijo identiteto), saj je težko minimizirati integral:

$$g(\boldsymbol{y}_i|\boldsymbol{\eta},\boldsymbol{\beta}) = \int_{\mathbb{R}^n} f(\boldsymbol{y}_i|\boldsymbol{\eta},\boldsymbol{\beta})\phi(\boldsymbol{\eta}_i;\Psi)d\boldsymbol{\eta}_i = \int_{\mathbb{R}} \phi(\boldsymbol{\eta};\Psi)\prod_{i=1}^{n_j} f(y_{ij}|\eta_{mj},\boldsymbol{\beta})d\boldsymbol{\eta}_i.$$

Tu je f gostota porazdelitvene funkcije, ki pa je lahko zelo kompleksna. Kot navadno je $\phi(.)$ multivariatna normalna porazdelitev ($N(0,\Psi)$). Druga enakost sledi, v kolikor predpostavimo neodvisnost med ravnmi. Ker je n navadno zelo velik, se za izračun ter minimizacijo tega integrala uporabljajo iterativni postopki.

4 Hierarhični linearni strukturni model

Predpostavke multiple analize normalnosti, homogenosti ter neodvisnosti podatkov so v realnosti skoraj vedno kršene.

1. *Normalna porazdelitev* je v nekaterih primerih primerna predpostavka, obstajajo pa tudi statistični modeli, ki so do neke mere robustni na kršenje te predpostavke (robustnost na normalnost pri SEM modelu dosežemo z GLS aproksimacijo). V kolikor poznamo porazdelitev naše slučajne spremenljike, želimo to informacijo v modelu uporabiti ter na ta način dobiti boljši model. Primer metode, ki informacije o porazdelitvah dobro upošteva, je GLMM model.

2. Za *homogenost variance* bomo na primeru pokazali, da je v večini prestroga zahteva. Ravno heteroskedastičnost je bila namreč vodilo, da smo vpeljali SEM model. Njena prisotnost vpliva tako na napačno ocenjene koeficiente slučajnih spremenljivk kot tudi na njihove standardne napake.

3. *Neodvisnost podatkov* je kršena v primeru, ko imamo hierarhično strukturo podatkov ali pa gledamo ponovitve poskusov. V GLMM modelu smo reševali te težave s pomočjo slučajnih učinkov.

SEM ter GLMM model prihajata iz različnih ozadij in se osredotočata na različna vprašanja. Oba imata svoje prednosti in slabosti, vendar so podatki navadno tako težko opazovani kot tudi grupirani. V ta namen sta Rabhe-Hasket ter Skrondal leta 2004 razvila hierarhičen SEM (HSEM) model, imenovan tudi GLLAMM (ang. generalized linear latent and mixed effect model), ki združuje prednosti obeh.

V naslednjih podpoglavjih si oglejmo tri ključne dele pri formulaciji HSEM modela:

1. modeliranje odziva,

2. strukturni model,

3. implementacija s pomočjo pogojnih verjetnosti.

4.1 Modeliranje odziva

Pogojno na latentne spremenljivke modeliramo odziv z GLMM modelom. Tu slučajni učinki ne bodo pomnoženi le s posamezno kontrolno spremenljivko, temveč z linearno kombinacijo dveh spremenljivk ter bo tako model vključeval faktorsko sktrukturo. Model odziva v dveh ravneh je torej:

$$\nu_{ij} = \boldsymbol{x}'_{ij}\boldsymbol{\beta} + \sum_{m=1}^{M} \eta^{(2)}_{mj} \boldsymbol{\lambda}^{(2)\prime}_{m} z^{(2)}_{mij}.$$

Tako kot prej prvi del enačbe opisuje ustaljene učinke, drugi pa slučajne. Vrednosti kontrolnih spremenljivk so zbrane v $z^{(2)}_{mij}$ s pripadajočimi vrednostimi faktorjev $\boldsymbol{\lambda}^{(2)\prime}_{m}$.

V treh ravneh model lahko posplošimo:

$$\nu_{ijk} = \boldsymbol{x}'_{ijk}\boldsymbol{\beta} + \sum_{m_2=1}^{M_2} \eta^{(2)}_{m_2jk} \boldsymbol{\lambda}^{(2)\prime}_{m_2} z^{(2)}_{m_2ijk} + \sum_{m_3=1}^{M_3} \eta^{(3)}_{m_3k} \boldsymbol{\lambda}^{(3)\prime}_{m_3} z^{(3)}_{m_3ijk}.$$

Sedaj lahko izpeljemo model odziva za L ravni ter M_l latentnih spremenljik na vsaki ravni. Linearni model odziva ima obliko:

$$\nu = \boldsymbol{\beta}'\boldsymbol{x} + \sum_{l=2}^{L} \sum_{m=1}^{M_l} \eta^{(l)}_{m} \boldsymbol{\lambda}^{(l)\prime}_{m} \boldsymbol{z}^{(l)}_{m}. \tag{11}$$

Tu smo zaradi poenostavitve notacije izpustili podpisane indekse. Elementi $\boldsymbol{x}$-a so neodvisne spremenljivke, ki imajo ustaljene (v vseh nivojih enake) učinke $\boldsymbol{\beta}$ na odvisno spremenljivko. Za m-to latentno spremenljivko na ravni l so slučajni učinki $\eta^{(l)}_{m}$ pomnoženi z linearno kombinacijo spremenljivk $\boldsymbol{\lambda}^{(l)\prime}_{m} z^{(l)}_{m}$. Tu še fiksiramo vrednost prvega elementa $\boldsymbol{\lambda}^{(l)}_{m}$ z 1, torej je $\boldsymbol{\lambda}^{(l)}_{m1} = 1$. Označimo z $\boldsymbol{\eta}^{(l)} := (\eta^{(l)}_{1}, \ldots, \eta^{(l)}_{M_l})'$ vektor latentnih spremenljivk na ravni l s pripadajočim vektorjem kontrolnih spremenljivk $\boldsymbol{z}^{(l)} := (\boldsymbol{z}^{(l)\prime}_{1}, \ldots, \boldsymbol{z}^{(l)\prime}_{M_l})'$. Definirajmo še vektor $\boldsymbol{\eta} := (\eta^{(2)\prime}, \ldots, \eta^{(L)\prime})'$ ter $\boldsymbol{z} := (\boldsymbol{z}^{(1)\prime}, \ldots, \boldsymbol{z}^{(L)\prime})'$. Pogojno upanje odvisne spremenljivke y glede na dane vektorje vrednosti neodvisnih spremenjivk $\boldsymbol{x}, \boldsymbol{z}$ ter $\boldsymbol{\eta}$ je povezano z linearno napovedjo ν preko povezovalne funkcije $g(.)$:

$$g(E[y|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\eta}]) = \nu. \tag{12}$$

Specifikacija je enolična z izbiro porazdelitvene družine ter pogojne porazdelitve odvisne spremenljivke glede na latentne ter neodvisne spremenljivke. Videli smo, da v modelu ni nobenih latentnih spremenljivk v prvi ravni, saj je le ta rezerviran za napako izbrane pogojne porazdelitve.

Oglejmo si naprej osnovno verzijo, kjer predpostavljamo zveznost odvisne spremenljivke y. Zaradi uporabe v kasnejših aplikacijah si bomo nato ogledali še primer, ko ima slučajna spremenljivka y binomsko obliko. V programskem jeziku Stata, ki je do sedaj še edini programski jezik, ki omogoča izvajanje HSEM modela v paketu GLLAMM, so poleg teh dveh na voljo še vse kombinacije iz tabele:

Vezna funkcija
Identiteta
Logaritem
Logit
Probit
Komplementaren log-log

Družina
Normalna
Gaussova
Gamma
Poissona
Binomska

4.1.1 Zvezna odvisna spemenljivka

V tem primeru lahko vzamemo normalno porazdelitev napake s povezovalno funkcijo identitete, torej

$$y = \nu + \epsilon,$$

z $f(\epsilon) = \frac{1}{\sqrt{2\pi\sigma^2}}\exp(-\epsilon^2/(2\sigma^2))$. Tudi neodvisna spremenljivka se porazdeljuje: $y \sim N(\nu, \sigma)$, zato je pogojna gostota:

$$f(y|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\eta}) = \sigma^{-1}\phi(\nu\sigma^{-1}), \tag{13}$$

kjer ϕ predstavlja gostoto za standardno normalno porazdelitev.

4.1.2 Diskretna odvisna spremenljivka

En od načinov definicije je preko zvezne funkcije y^*:

$$y^* = \nu + \epsilon^*,$$

ki jo diskretiziramo na način:

$$y = y_s \text{ če je } \kappa_{s-1} < y^* \leq \kappa_s,\ \kappa_0 = -\infty, \kappa_1 = 0, \kappa_S = \infty.$$

Skupno število skupin S mora biti končno, prage κ_s $s = 2, \ldots, S-1$ pa moramo določiti vnaprej. Uprabimo lahko logit ali probit vezno funkcijo, v tem primeru je gostota za napako dana z:

- $f(\epsilon) = \exp(-\epsilon)[1 + \exp(-\epsilon^2)]^{-2}$ za logit,
- $f(\epsilon) = \frac{1}{\sqrt{2\pi}}\exp(-\frac{1}{2}\epsilon^2)$ za probit.

Pogojna gostota pa je

$$f(y = y_s|\boldsymbol{x}, \boldsymbol{z}, \boldsymbol{\eta}) = F(\sigma^{-1}[\kappa_s - \nu]) - F(\sigma^{-1}[\kappa_{s-1} - \nu]), \tag{14}$$

kjer je F komulativna porazdelitvena funkcija ter je σ parameter merila (ang. scale parameter).

4.2 Strukturni model

Za vektor latentnih spremenljivk $\boldsymbol{\eta} \in \mathbb{R}^{1\times M}$ ($M = \sum_{l=1}^{L} M_l$) ima strukturni model obliko:

$$\boldsymbol{\eta} = B\boldsymbol{\eta} + \Gamma \boldsymbol{w} + \boldsymbol{\zeta}. \tag{15}$$

V matriki $B \in R^{M\times M}$ imamo izražene povezave med latentnimi sremenljivkami, $\boldsymbol{w} \in \mathbb{R}^{N\times 1}$ je vektor opazovnih neodvisnih spremenjivk, s pripadajočo matriko koeficientov $\Gamma \in \mathbb{R}^{M\times N}$. Vektor motenj latentnih spremenljivk je označen z $\boldsymbol{\zeta} \in \mathbb{R}^{M\times 1}$.

Za razliko s SEM modelom, izraženim z enačbami (1)-(3), tu vse latentne spremeljivke nastopajo v vektorju $\boldsymbol{\eta}$ z vsemi povezavami med njimi v B. V vektorju napak vsak element varira v istem nivoju kot pripadajoča latentna spremenljivka. Pomembno je izpostaviti dejstvo, da je model (15) večrazsežni strukturni model, z latentnimi spremenljivkami v različnih nivojih, kar seveda vključuje SEM model kot poseben primer.

Predpostavka, ki jo moramo dodatno sprejeti je, da so napake latentnih spremenjivk na isti ravni lahko odvisne, medtem ko morajo biti le-te na različnih ravneh neodvisne. Če nas ne zanimajo povratne informacije med latentnimi spremenljivkami, je B strogo zgornje trikotna matrika, v kolikor so elementi $\boldsymbol{\eta}$ primerno permutirani. Najlažje primerno permutacijo dosežemo z elementi $\boldsymbol{\eta}^{(l)}$, urejenimi po naraščajočih vrednostih l.

4.3 Implementacija s pomočjo pogojni verjetnosti

V želji po izpeljavi metode največjega verjetja moramo poznati bodisi porazdelitev slučajne spremenjivke napak $\boldsymbol{\zeta}$, bodisi porazdelitve latentnih spremenljivk $\boldsymbol{\eta}$. Na tem mestu bomo privzeli, da imajo latentne spremenljivke na ravni l multivariatno normalno porazdelitev s povprečno vrednostjo 0 ter kovariančno matriko Σ_l, čeprav bi načeloma lahko vzeli poljubno večrazsežno porazdelitev.

Označimo s $\boldsymbol{\theta}$ vektor, ki vključuje koeficiente neodvisnih spremenjivk, prav tako pa tudi faktorske vrednosti $\boldsymbol{\lambda}_m^{(l)}$, $m = 1, \ldots, M_l$, $l = 1, \ldots, S-1$ ter pragove κ_s, $s = 2, \ldots, S-1$ za diskretne odvisne spremenljike. Definirajmo še $y_{(l)}$ kot vektor odvisne spremenljivke ter $X_{(l)}$ matriko pojasnjevalnih spremenljivk (z vrsticami $[\boldsymbol{x}', \boldsymbol{z}', \boldsymbol{z}^1, \boldsymbol{w}]$) za vse elemente iz prve ravni, ki pripadajo določeni spremenljivki na ravni l. Analogno naj bo y vektor odvisnih spremenljivk ter X matrika pojasnjevalnih spremenljivk za vse enote. Pripadajočo pogojno verjetnost odvisne spremenljivke za prvo raven označimo $f^{(1)}(y_1|X_{(1)}, \boldsymbol{\zeta}^{(2+)}; \boldsymbol{\theta})$, kjer je $\boldsymbol{\zeta}^{(l+)} = (\boldsymbol{\zeta}^{(l)\prime}, \ldots, \boldsymbol{\zeta}^{(L)\prime})'$. Obliko pogojnih spremenljivk smo že spoznali v enačbi (13) ter (14), ki so različne glede na željeno obliko odvisne slučajne spremenljivke. Iz predpostavke vemo, da imajo latentne spremenljivke na ravni l multivariatno normalno porazdeitev, kar označimo s $h^{(l)}(\boldsymbol{\zeta}^{(l)}; \boldsymbol{\theta})$.

Model v dveh ravneh maksimiziramo po metodi največjega verjetja, torej moramo maksimizirati funkcijo:

$$\ell(\boldsymbol{\theta}, y, X) = \prod_{j=1}^{J} \int_{\mathbb{R}} \left[\prod_{i=1}^{I} f^{(1)}(y_1|X_{(1)}, \zeta^{(2+)}) \right] h(\boldsymbol{\zeta}_j|\boldsymbol{\theta}) d\boldsymbol{\zeta}_i.$$

Rekurzivno sedaj izpeljemo pogojno verjetnost za l ravni, kjer gre produkt preko vseh $l-1$ skupin kot

$$f^{(l)}(y_l|X_{(l)},\zeta^{([l+1]+)};\boldsymbol{\theta}) = \int_{\mathbb{R}} \prod f^{(l-1)}(y_{l-1}|X_{(l-1)},\zeta^{(l+)})h^{(l)}(\boldsymbol{\zeta}^{(l)}|\boldsymbol{\theta})d\boldsymbol{\zeta}^{(l)}. \tag{16}$$

Maksimizirati pa želimo produkt pogojnih verjetnosti:

$$\ell(\boldsymbol{\theta},y,X) = \prod f^{(L)}(y_L|X_{(L)},\boldsymbol{\theta}). \tag{17}$$

V GLLAMM modelu se metoda izvaja na način, da maksimizira numerično integrirano funkcijo največjega verjetja (17) s pomočjo Newtnovega-Raphsonovega algoritma.

Lema 4.1. *Integral lahko diskretiziramo z utežmi $(w_i^l)_{i=1}^{M_l}$ ter $(\alpha_i^l)_{i=1}^{M_l}$, za vsako izmed ravni $l=1,\dots,L$ na način*

$$\int_{\mathbb{R}} f(x_1,\dots,x_L)dx_1\dots dx_L \approx \sum_{i=1}^{M_1}\left[\cdots\sum_{i=1}^{M_L}\left(\prod_{k=1}^{L} w_{i_k}^k\right) f(\alpha_{i_1}^1,\dots,\alpha_{i_L}^L)\right].$$

Za dane parametre ocenimo integral glede na slučajne spremenljivke napak $\boldsymbol{\zeta}^{(l)}$ z integriranjem preko M_l neodvisnih standarno normalno porazdeljenih spremenljivk $\boldsymbol{v}^{(l)}$ s $\boldsymbol{\zeta}^{(l)} = C_l\boldsymbol{v}^{(l)}$. Matrika C_l je dekompozicija Choleskyja kovariančne matrike Σ_l. Definirajmo še $\boldsymbol{v}^{(l+)} = (\boldsymbol{v}^{(l)\prime},\dots,\boldsymbol{v}^{(L)\prime})$. Najenostvnejša ideja za poenastavitev enačbe je, da integral diskretiziramo:

$$\begin{aligned}
&\int_{\mathbb{R}} \left[\prod f^{(l-1)}(y_{l-1}|X_{(l-1)},\zeta^{(l+)})\right] h^{(l)}(\boldsymbol{\zeta}^{(l)}|\boldsymbol{\theta})d\boldsymbol{\zeta}^{(l)} \\
&= \int \phi(v_{M_l}^{(l)})\dots\int \phi(v_{M_1}^{(l)})\left[\prod f^{(l-1)}(y_{l-1}|X_{(l-1)},v^{(l)},v^{([l+1]+)},\boldsymbol{\theta})\right] dv_1^{(l)}\dots dv_{M_l}^{(l)} \\
&\approx \sum_{r_{M_l}} \pi_{M_l}\cdots\sum_{r_1}\pi_{r_1}\left[\prod f^{(l-1)}(y_{l-1}|X_{(l-1)},\alpha_{r_1},\dots,\alpha_{r_{M_l}},\boldsymbol{\theta})\right].
\end{aligned}$$

Tudi tukaj smo standardno normalno gostoto označili s $\phi(.)$, π_r ter α_r pa predstavljata uteži ter lokacijo. Prednost take poenostvitve je njegova enostavnost, vendar metoda ni vedno zanesljiva. V programu GLLAMM sta zato še dodatno na voljo metoda adaptivnih kvadratov, ki je natačno predstavjena v članku (Hesketh, Skondal 2005) ter Bayesova aproskimacija.

Zavedati pa se moramo, da tudi HSEM ne vključuje vseh možnosti, ki se nam v realnosti lahko pojavijo. Velike težave ima z manjkajočimi vrednostmi. V kolikor so manjkajoče vrednosti slučajne (ang. missing at random), jih program enostavno izpusti. Ne more pa modelirati podatkov, kjer ima manjkajoča vrednost pomen. Najenostavnejši primer, ko ima manjkajoči kazalnik pomen, so podjetja, ki delujejo v industrijski panogi, kjer določenega kazalnika ni mogoče izračunati. Seveda ne želimo zaradi tega zavreči vseh podjetij iz te industrijske panoge, hkrati pa je prav mogoče, da kazalnik želimo obdržati, saj je pomemben v modelu. Odgovora na tako problematiko nam ne da nobena od metod, predstavljenih v tej magistrski nalogi.

Poleg tega se pri modeliranju raziskovalci soočajo z zahtevnostjo specifikacije modela in njegovo predstavitvijo v računalniku. Prav to je ena glavnih pomankljivosti, saj pri modeliranju zahtevamo čim bolj enostaven model, ki pa se čimbolje prilega podatkom. Težave so tudi pri prileganju, saj je za specificiran model težko preveriti, kako dobro se model prilega. Glavni razlog je, da še niso razvite funkcije za merjenje natančnosti prileganja. V GLLAMM modelu je edina mera, ki jo model vrne, logaritem verjetja (ang. log-likelihood). Druga izmed možnosti za preverjanje prileganja podatkom v primeru, ko ima odvisna spremenljivka bernulijevo obliko, sta še krivulja ROC ter Ginijev koeficient, ki smo ju že spoznali in bosta, kot že rečeno, glavni meri prileganja skozi celotno magistrsko nalogo. Velika težava GLLAMM programa je tudi hitrost, saj ga ocenjujejo za enega časovno zahtevnejših algoritmov. Težavo razvijalci sicer postopno popravljajo in je v prihodnosti pričakovati boljše ter hitrejše programe. Prav vse te težave ter njegova relativna novost postavljata model HSEM ob stran dobro ustaljenim modelom v literaturi. V nadaljevanju želimo pokazati, da je model HSEM kljub svojim težavam vreden obravnave kot ena izmed možnosti pri modeliranju verjetnosti propada podjetja.

5 Kreditna analiza verjetnosti propada podjetja

Beseda kreditiranje se prvič omeni skupaj s pojavom bank okoli 2000 let pred našim štetjem. Kreditiranje je proces, pri katerem finančna institucija komitentu odobri sredstva, le ta pa se zaveže, da jih bo vrnil pod vnaprej znanimi pogoji. S tem kreditiranjem se opravlja alokacijska funkcija kapitala. Na ta način se lahko izvedejo projekti, ki brez alokacijaskega prenosa denarja ne bi bili mogoči. Tako bančništvo že od samih začetkov predstavlja pomembno vlogo v nacionalnih ekonomijah. Bančna stabilnost pa je ključna za nemoteno delovanje gospodarstva.

Osnovna dejavnost bank je torej privlačiti sredstva ter jih ponovno investirati po višji obrestni meri. Ob vsaki odobritvi kredita pa se soočajo s tveganji. Tveganje samo po sebi ni slabo, a le ob predpostavki, da banka tveganja identificira, jih kvantitativno opredeli in z njimi upravlja. Potrebno se je zavedati škodljivih posledic prevelike izpostavljenosti do manjše skupine komitentov. V primeru finančnih težav le-teh bi lahko to resno ogrozilo finančno stabilnost kreditodajalca. Nujna je torej diverzifikacija, kar pomeni, da finančne težave nekaterih komitentov niso sposobne ogroziti banke. V kolikor bi le ta sprejela preveč tveganja, bo kmalu zašla v težave, ko ne bo sposobna poplačati svojih obveznosti ter bo tako postala insolventna. Kljub zavedanju vseh tveganj pa je upravjanje z njimi relativno nova stvar, ki sega v začetke devetnajstega stoletja. Pomembno vlogo pri upravljanju s tveganji je odigrala tudi tudi vedno večja vloga računalnika. Pred tem je bila statistična obdelava večjih količin podatkov praktično nemogoča.

Statistično analiziranje podatkov je omogočilo bankam, da lahko ob enaki količini prejetih depozitov in kapitala podelijo več kreditov. Na ta način se je povečala dobičkonostnost bank, hkrati pa povečala verjetnost bančnih kriz v času izjemnih dogodkov. Ker banke ne upravljajo le z lastnim kapitalom ampak s kapitalom večjega dela prebivalstva, je finančno zdravje bank tudi ob izjemnih dogodkih postalo pomembno makroekonomsko vprašanje. Zato se je v zadnjem desetletju modeliranje tveganja znašlo pod številnimi regulativnimi postopki, med katerimi je najbolj uveljavljen Basel II. Le ta tveganje razvršča v tri večje skupine: kreditno tveganje, tržno tveganja ter operativno tveganje. Operativno tveganje predstavlja možnost izgube zaradi neustreznega ali neuspešnega izvajanja notranjih postopkov, procesov, ljudi in sistemov ali pa zaradi zunanjih dogodkov. Na ta način vidimo, da operativno tveganje ni v pravem smislu finančno tveganje, vendar ima finančne posedice. Zaradi svoje narave je specifično za vsako banko ter vsako podjetje posebej. Pri tržnem tveganju govorimo o verjetnosti zmanjšanja vrednosti sredstev ali pa celotnega portfelja zaradi razmer na trgu. Tržnih tveganj je več, vendar so za banko najbolj pomembni obrestno tveganje, ko ima banka neenakomerno strukturo fiksnih ter variabilnih obresti, valutno tveganje, v primeru ko banka posluje v različnih valutah, ter likvidnostno tveganje, ko so ročnosti sredstev in obveznosti različne. Kreditno tveganje je verjetnost, da posojilojemalec ne bo sposoben odplačati svojih obveznosti (glavnice ter obresti) in je tako najbolj očitno izmed treh, saj izhaja iz narave poslovanja bank, hkrati pa lahko banka nanj tudi vpliva. Zato se bomo v tem delu osredotočili na kreditno tveganje, oziroma bolj natančno, na modeliranje verjetnosti propada podjetja.

Namen analize kreditnega tveganja je torej oceniti verjetnost, da posojilojemalec ne bo sposoben vrniti posojenega denarja pod dogovorjenimi pogoji. Obstajajo različni razlogi za nespoštovanje kreditnih pogojev, najpogosteje pa se pojavijo zaradi finančnih težav podjetja. Za zavarovanje pred kreditnim tveganjem banka oz. vsak posojilodajalec izvaja kreditno analizo, ki je kombinacija znanosti ter umetnosti. Medtem ko v znanstvenem delu raziskovalec koplje po zgodovini posojilojemalca ter trenutnih dejstvih, se umetnost pokaže s prepoznavanjem nemerljivih informacij, ki jih je sposoben prepoznati zaradi svojih izkušenj ter občutka o dani tematiki. Zavedati se je namreč potrebno, da zgodovina ne nujno napoveduje prihodnosti ter da pretekli zaslužki ne bodo odplačali prihodnjega dolga, o katerem danes odločamo. Kot pravi star rek, zgodovino preučujemo zato, da bi razumeli sedanjost ter lažje sklepali o prihodnosti. Eliminacija kreditnega tveganja je nemogoča, zato je v trenutku, ko se mora banka odločiti o višini kredita ter njegovi obrestni meri, nujno poznavanje tveganj o nepoplačilih, s katerimi se sooča. Le na ta način lahko namreč zavaruje svoja sredstva ter dolgoročno plemeniti svoj kapital. Privzemimo še na znanje, da čeprav bomo v večini govorili o posojilodajalcu kot banki, ki tudi najbolj intenzivno kreditirajo gospodarstvo, je posojilodajalec lahko vsak posameznik, ki kupi obveznico določene družbe.

Od začetka svetovne gospodarske krize v letu 2007 je intenzivno ukvarjanje s kreditnim tveganjem spet postalo pomembno področje tako bančnikov kot tudi raziskovalcev. Skrajne razmere so namreč pokazale, da so banke precenjevale kreditno sposobnost podjetij ter posameznikov. V letih pred krizo so v želji po zagotavljanju vedno večjih dobičkov, hkrati pa v vedno ostrejši konkurenci, dodeljevale previsoke ocene kreditne sposobnosti kreditojemalcev. Težave so se začele, ker je velika večina kreditnih analitikov pozabila na objektivnost ter profesionalnost. Kljub vsemu pa vsa podjetja niso propadla, zato je prav vprašanje, kaj je razlikovalo preživela podjetja od propadlih tisto, na katerega želijo raziskovalci odgovoriti. Vedenje o načinih prepoznavanja težav podjetji nam na dolgi rok prinaša solventnost ter profitabilnost. Zaradi finančnega vzvoda se že manjše spremembe vrednosti kreditne sposobnosti komitentov lahko preslikajo v velike izgube bančnega kapitala. Tako v želji po objektivnosti tradicionalno subjektivne ocene tveganja, zamenjujejo najrazličnejši modeli, ki pa morajo biti zaradi večje varnosti finančnega sistema odobreni s strani centralnih bank.

V letih po krizi so k omejevanju ter modeliranju tveganja znatno pripomogla priporočila Baselskega odbora za nadzor bank (BCBS). Glavni namen BCBS odbora je izboljšati kvaliteto bančnega nadzora. Pri uresničevanju njihovih priporočil centralne banke ob nadzoru ocenjevanja kreditnega tveganja bankam ali drugim finančnim institucijam dopuščajo lastne modele ter v njih izbrane paramete, ki pa morajo biti odobreni ter redno preverjani.

Modeli kreditnega tveganja postanejo še bolj kompleksni, saj pričakovana izguba za banko kljub propadu podjetja ni celotna investicija. Tudi v primeru propada podjetja bo banka dobila povrnjen vsaj del kredita. Basel II pričakovano izgubo definira kot produkt deleža izgubljenih sredstev v primeru propada (ang. loss given default - LGD), verjetnosti propada podjetja (ang. probability of default - PD) ter vrednostne izpostavljenosti v primeru propada (ang. exposure at default - EAD).

Medtem ko LGD model poda delež, ki ga banka lahko izgubi v primeru propada, PD podaja verjetnost propada le tega. EAD model pa ocenjuje celotno izpostavljenost banke v trenutku, ko podjetje propade. Trgi se pri dobrih podjetjih večinoma osredotočajo na PD ocene, ki so zelo nizke. Na vrednost premoženja zato bistveno bolj vpliva čim bolj natančna PD ocena kot LGD, za katerega pa strokovnjaki vzamejo kar povprečje panoge. V primeru ko se verjetnost propada poveča in podjetje začne zahajati v težave, pa se vse več analitikov začne ukvarjati z oceno LGD. V tej magistrski nalogi si bomo ogledali, kako s pomočjo metode HSEM izpeljati enega izmed možnih PD modelov.

5.1 Ocene kreditnega tveganja

Začetki bonitetnih agencij segajo v leto 1909, ko je John Moody poleg cen železniških obveznic izdal tudi oceno tveganja le teh. Od takrat je povpraševanje po bonitetnih ocenah vedno bolj rastlo, saj so reševale problem asimetričnih informacij (odpravljale so torej podoben problem kot problem limon na trgu rabjenih avtomobilov). Te ocene nam namreč povedo, koliko tveganja moramo sprejeti pri investiciji v določeno obveznico. Na ta način lahko ločimo nizko tvegana podjetja, ki se tako lahko zadolžujejo po nižji obrestni meri. Ob tem se povečuje tudi število potencialnih investitorjev za odkup obveznic ter se veča likvidnost na trgu obvezniških papirjev, saj so tveganja bolj opredeljena, razmerje med tveganjem in donosnostjo pa je lažje oceniti. Temeljni nalogi bonitetih agencij sta torej ponuditi neodvisno ocene sposobnosti odplačila izdajatelja dolga ter ponuditi kontrolo nad uporabo denarja, ki ga je podjetje pridobilo. Slednje dosegajo s stalnim preverjanjem bonitetne ocene. Kreditne ocene izdajajo tako za države kot tudi določena podjetja ter celo določene projekte, za katere se organizira izdaja obveznic. Vsekakor so iz tega stališča neka začetna slika tudi za analitike. V tem delu bomo spoznali, kako so definirane bonitetne ocene tveganja ter zakaj so banke kljub temu še vedno motivirane za razvijanje svojih modelov tveganja.

Na trgu trenutno deluje okoli stopetdeset kreditnih agencij, vendar okoli 95% trga obvladujejo tri najpomembnejše, ki so bile vse ustanovljene v začetku dvajsetega stoletja.

1. **Fitch Ratings** izhaja iz založbe Fitch Publishing Company, katere jasen cilj je izdajati najrazličnejše finančne statistike na delnice ter obveznice, ki bi investitorjem olajšali odločitve. V letu 1924 je Fitch prvič objavil lestvico ocenjovanja od AAA do D, ki jo poznamo še danes ter je postala osnova za celotno indutrijo. Trenutno pa je Fitch najmanjši izmed velikih treh in obvladuje okoli 15% trga.

2. **Moody's** je bil ustanovljen leta 1909 ter je do leta 1970 ponujal samo ocene tveganja državnih obveznic. Danes obvladuje približno 40% trga.

3. **S & P** ima prav tako kot Moody's 40% tržni delež. Korenine izhajajo iz podjetij H.W. Poor Company ter Standard Statistics združenih v letu 1941. Danes pa poleg kreditnih ocen ponuja tudi druge finančne storitve, kot so investicijsko raziskovanje.

Njihov način razvrščanja je prikazan v tabeli 1.

Fitch Ratings	S&P	Moody's	Pojasnilo	Stopnja tveganja
AAA	AAA	Aaa	Najvišja kakovost dolžniških papirjev.	Zgornja investicijska meja (ang. upper investment grade)
AA+ AA AA-	AA+ AA AA-	Aa1 Aa2 Aa3	visoka odplačilna sposobnost.	
A+ A A-	A+ A A-	A1 A2 A3	Velika sposobnost odplačevanja, lahko se pojavijo težave v daljši prihodnosti.	
BBB+ BBB BBB-	BBB+ BBB BBB-	Baa1 Baa2 Baa3	Ustrezna sposobnost odplačevanja, v primeru neugodnih razmer pričakovane težave.	Spodnja investicijska meja (ang. lower investment grade)
BB+ BB BB-	BB+ BB BB-	Ba1 Ba2 Ba3	Visoko tvegana naložba (prihodost ni predvidljiva).	Špekulativne investicije (ang. speculative grade)
B+ B B-	B+ B B-	B1 B2 B3	Majhna verjetnost preživetja na dolgi rok.	
CCC+ CCC CCC-	CCC+ CCC CCC-	Caa1 Caa2 Caa3	Majhna verjetnost preživetja na kratki rok.	
D	D	D	Neizpolnjevanje obveznosti	Propad (ang. default)

Tabela 1: Bonitetne ocene ter njihov pomen.

Seveda se ob tem pojavlja vprašanje, zakaj banke sploh izdelujejo svoje modele, ko pa bi lahko uporabile kreditno oceno ter dodale nek dodatek ter tako postavile svojo obrestno mero za določeno podjetje. To seveda že v osnovi ni mogoče, saj vsa podjetja nimajo bonitetne ocene. Za pridobitev le te mora namreč podjetje plačati bonitetni agenciji, ki nato opravi ocenjevanje. Največkrat se podjetje za tako potezo odloči ob izdaji obveznic, ki pa jih, kot vemo, izdajajo praviloma večja podjetja. Ob izdaji obveznic želi namreč podjetje pritegniti kar največ investitorjev ter s pomočjo bonitetnih ocen povečati atraktivnost svojih obveznic. V primerih ko podjetje želi z zbranim denarjem investirati v točno določen projekt, se lahko ocena nanaša samo na projekt ter je tako drugačna od bonitetne ocene celotnega podjetja. Prav zaradi dejstva, da bonitetne ocene plačuje podjetje samo, pa številni trdijo, da ocene niso nepristranske (obstaja konflikt interesov). Vsa namigovanja so se le še okrepila med zadnjo gospodarsko krizo, ko se je izkazalo, da so bonitetne angencije prenapihovale bonitetne ocene. Ob odsotnosti jasne definicije, kaj loči dobro od slabe ocene, banka ne pozna točnega tveganja, ki bi ga v tem primeru prevzela. Poleg tega bonitetna ocena odraža preteklo stanje ter ima na ta način slabo napovedovalno moč. Velikokrat je namreč spreminjanje ocene zelo neodzivno v primerjavi s trgom (navadno traja več mesecev, preden je podjetje ponovno ocenjeno). Zaradi vseh teh razlogov je nujno, da banke izdelujejo svoje modele za predvidevanje tveganja, ki morajo imeti tudi napovedovalno moč.

5.2 Dejavniki kreditne analize

S pomočjo kreditne analize želi banka pridobiti oceno tveganosti posojila. Le ta je odvisna že od njenih predhodnjih poslovanj s stranko oziroma podjetjem ter trenutne situacije. Na podlagi dobre analize lahko tako pripravi atraktivne pogoje za najem posojila pri posojilojemalcih ter si tako zagotovi dobre stranke (torej stranke, ki bodo kredite pod danimi pogoji redno odplačevale). V kolikor je kreditna analiza preveč posplošena oziroma banka stranke ne pozna dovolj, ji bo težko odobrila konkurenčne pogoje, ki bodo hkrati profitabilni za banko. V tem primeru bodo dobri kreditojemalci odšli k drugim bankam, le ta pa bo ostala s slabimi kreditojemalci (gre torej za problem podoben problemu "slabih limon na trgu rabljenih avtomobilov"). Temeljna klasifikacija kreditnega tveganja je skozi vseh pet skupin dejavnikov tveganja metode petih C-jev.

- *Plačilna sposobnost dolžnika* (ang. capacity) se nanaša na sposobnost podjetja za odplačilo dolga. Pri tej je ključnega pomena volatilnost denarnih tokov. V prošnji za izdajo dolga mora biti namreč jasno opredeljeno kdaj ter s katerimi denarnimi tokovi bomo poplačali dolg. Telekomunikacijska industrija ter industrija osnovnih potrošnih dobrin imajo manjšo volatilnost kot visokotehnološka ali biotehnološka podjetja, zato bodo bolj atraktivna za posojilodajalce. Že iz teorije financiranja namreč izhaja, da je podjetja z višjo volatilnostjo boljše financirati z lastniškim kapitalom, ki ne zahteva vsakomesečnih odplačil. Plačilna sposobnost se nanaša tudi na kreditno zgodovino podjetja, torej kako so podjetja v zgodovni poplačevala svoje dolgove.

- *Jamstva* (ang. collateral) pri kreditu nam dodajo neko zavarovanje v primeru, da kredita ne bodo mogli poplačati. Vključuje lahko od stavb, opreme, strojev, terjatev, finančnih naložb - torej vseh sredstev podjetja ter včasih tudi premoženja lastnikov. Vključuje lahko tudi garancijo tretje stranke, da bo poplačala dolgove v primeru, da jih podjetje samo ne bo sposobno poravnati. Jamstva pri kreditih morajo biti jasno določena že ob sklenitvi kredita.

- *Kapital* (ang. capital) predstavlja lastnikov vložek ter je tako pokazatelj finančnega zdravja podjetja. Pred izdanim posojilom mora imeti podjetje zadosten delež kapitala, ki bo delovalo kot neko jamstvo v primeru težav. V primeru finančnih težav ima dolžniški kapital prednost pred lastniškim kapitalom. Dolžniški kapital ima pri poplačilu prioriteto v primerjavi z lastniškim kapitalom.

- *Razmere* (ang. conditions) opisujejo specifike podjetja ter kam se bo investiral izposojen denar. Torej, posojilodajalec želi že pred odobritvijo kredita vedeti, ali se bo ta porabil za gradnjo novih tovarn, prevzem podjetja, razvoj novih produktov in podobno. V tem delu se preverijo tudi ekonomske razmere ter gospodarski položaj panoge, v kateri posojilojemalec posluje.

- *Splošen vtis* (ang. character) je najbolj subjektivna ocena, kreditodajalec jo določi s pomočjo zgodovine s podjetjem, znanjem ter izkušnjami. Je torej nek vsesplošen vtis, ki se ga kot komponento kreditnega tveganja najhitreje pozabi v času gospodarske rasti. Meri namreč pripravljanost podjetja za izpolnjevanje finančnih zavez. Kreditojemalec jo določi s pomočjo zgodovine podjetja ter zgodovine njegovih vodinih pri poplačilu dolgov.

V času gospodarskih rasti so, zgodovinsko gledano, banke posvečale veliko težo plačilni sposobnosti dolžnikov, le malo pa ostalim točkam. Ob vsaki krizi pa se izkaže, da so vse točke enakovrednega pomena.

V želji po modeliranju izgub bi seveda želeli ohraniti čimveč zgornjih karakteristik. V samem modeliranju verjetnosti propada podjetja (PD modelu) pa se bomo direktno osredotočili na plačilno sposobnost, kapital ter razmere na trgu, indirektno pa bomo upoštevali tudi ostala dva dejavnika. V našem modelu bomo namreč uporabljali finančne kazalnike željenega podjetja ter drugih podjetij delujočih na trgu. Glavni vir informacij so namreč prav finančni izkazi podjetja, ki jih le ta v večini izdajajo na četrtletni ravni. Finančni izkazi vsebujejo številne računovodske podatke, s pomočjo katerih lahko z enostavnimi matematičnimi operacijami pridemo do finančnih kazalnikov, ki odražajo odnose med podatki o podjetju. Na ta način nam predstavijo podjetje skozi številke. Seveda nam sama številka ne pove veliko, vendar s primerjavo kazalnikov med podjetjem in njegovimi konkurenti (tudi drugačnih velikosti, saj lahko le ta vpliv eliminiramo) lahko prikažemo prednosti ter težave določenega podjetja. Kazalnik prikazuje obnašanje podjetja v danem časovnem obdobju. Primerjava kazalnikov skozi čas tako pokaže področja, kjer se podjetje izboljšuje ali pa kjer je potrebna dodatna pozornost. Prav zato se strokovnjaki za ocenjevanje kreditne sposobnosti večinoma zanašajo na kazalnike ter na njihove kombinacije. Navadno so torej finančni kazalniki izračunani v namene ocenjevanja podjetja ter jih razvstimo v sledniе skupine:

1. kazalniki likvidnosti (ang. liquidity ratios),
2. kazalniki zadolženosti (ang. debt ratios),
3. kazalniki dobičkonosnosti (ang. profitability ratios),
4. kazalniki učinkovitosti (ang. efficiency ratios).

Za vsako izmed skupin si bomo ogledali samo kazalnike, ki bodo kasneje nastopali v katerem izmed modelov ali pa so najbolj tipični predstavniki za svojo skupino. Vsi kazalniki, ki jih bomo obravnavali pri modeliranju, bodo po skupinah razvrščeni v naslednjem poglavju.

5.2.1 Kazalniki likvidnosti

Namen likvidnostnih kazalnikov je pokazati sposobnost podjetja za poplačilo kratkoročnih obveznosti. Merijo namreč plačilno sposobnost podjetja na dan, ko je potrebno poravnati obveznosti. Zaradi same narave pa je kratkoročna likvidnost zelo dinamičen proces, ki se spreminja skoraj na dnevni ravni. Različni kazalniki likvidnosti po večini primerjajo kratkoroča sredstva ali del le teh s kratkoročnimi obveznostmi. Vrednost kazalnika nad ena nakazuje, da so kratkoročne obveznosti povsem pokrite. V analizi podjetja so ključnega pomena, saj so podjetja, ki imajo težave z odplačevanjem kratkoročnih dolgov, veliko bolj tvegana. Vsak kazalnik mora biti postavljen v pravilni kontekst. Primerjati ga je potrebno med posameznimi leti, med najbljižnimi konkurenti ter s celotno industrijo. Ob tem je potrebno poudariti, da so vrednosti kazalnikov lahko odvisne tudi od pravno-ekonomskega ustroja države. Primer je različna stečajna zakonodaja med državami. Najbolj poznani so trije likvidnostni kazalniki:

1. trenutna likvidnost (ang. cash ratio),
2. količnik obratno-finančne likvidnosti (ang. quick ratio),
3. količnik splošne likvidnosti (ang. current ratio).

Kazalnik trenutne likvidnosti je v svojem bistvu najkonzervativnejši (v imenovalcu imamo zgolj najbolj likvidna sredstva) ter hkrati najpogosteje uporabljeni od treh. Izračunamo ga kot:

$$\text{trenutna likvidnost} = \frac{\text{denar + kratkoročne finančne naložbe}}{\text{kratkoročne obveznosti}}.$$

Podojilodajalci ga zato vzemajo kot mero, kako enostavno podjetje odplačuje obveznosti zgolj z najbolj likvidnimi sredstvi. Višje vrednosti zato prikazujejo lažje odplačevanje, kljub vsemu pa lahko previsoke vrednosti kazalnika trenutne likvidnosti nakazujejo na težave pri upravljanjem z denarjem. Podjetje mora namreč denar uporabiti za svoje delovanje ter ga tako plemenititi (oportunitetni stroški držanja prevelikih količin denarja na določeni točki presežejo dodatno varnost).

Osnovna definicija količnika obratno-finančne likvidnosti je, da pri merjenju trenutne likvidnostni ne upošteva zalog. Količnik obratno-finančne likidnosti je podoben kazalniku trenutne likvidnosti, vendar je v imenovalcu vrednosti denarja in kratkoročnih naložb dodana še vrednost poslovnih terjatev. Izračunamo ga kot:

$$\text{količnik obratno-finančne likvidnosti} = \frac{\text{trenutna sredstva - zaloge}}{\text{kratkoročne obveznosti}}.$$

Historično gledano vrednost kazalnika pod ena nakazuje na povečano verjetnost neizpolnjevanja kratkoročnih obveznosti. Za odplačevanje svojih obveznosti bo namreč tako podjetje izvedlo faktoring (prodaja poslovnih terjatev) ali prodalo del svojih sredstev, kar bo vplivalo na samo delovanje podjetja.

Najsplošnejši med njimi je količnik splošne likvidnosti. Ta namreč vključuje vsa kratkoročna sredstva, tudi tista, katerih prodaja je mogoča le po diskontu (ang. fire sell). Z vključitvijo vseh trenutnih sredstev dosega prikaz splošnejše slike, torej ali ima podjetje dovolj sredstev za poplačilo svojih obveznosti v nekoliko daljšem obdobju, na primer v enem letu. Formula za izračun je podobna formuli količnika obratno-finančne likvidnosti, le da v imenovalcu ne odštejemo vrednosti zalog. Izračunamo, ga torej kot:

$$\text{količnik splošne likvidnosti} = \frac{\text{trenutna sredstva}}{\text{kratkoročne obveznosti}}.$$

Za podjetja v industrijah z nizko volatilnostjo denarnih tokov so idealne vrednosti okoli 1.5, za ostala pa okoli 2. Vrednost količnika obratno finančne likvidnosti se od vrednosti kazalnika splošne likvidnosti loči predvsem po vključitvi zalog pri slednjem. Izbira primernejšega kazalnika je odvisna od industrije, v kateri podjetje deluje. Za določene zaloge obstaja zelo likviden trg, spet druga pa so lahko visoko specializirana in je tako lahko nabor kupcev zelo omejen.

5.2.2 Kazalniki zadolženosti

Za razliko od likvidnostih kazalnikov pa kazalniki zadolženosti merijo prevsem sposobnosti podjetja za odplačevanje finančnih obveznosti. Merijo torej sposobnost podjetja za odplačilo tako obresti kot tudi glavnice, hkrati pa podajajo informacije o kapitalski sestavi podjetja (npr. kolikšen je delež lastniškega proti dolžniškemu kapitalu). Na ta način določajo mero vsesplošnega tveganja, ki ga prevzemajo lastniki ali posojilodajalci. Višja stopnja zadolženosti podjetja prinaša višje tveganje propada. Za natančno analizo je tudi tu nujna primerjava med leti ter v določeni industriji.

Prvi izmed kazalnikov, ki ga bomo podrobneje spoznali je neto dolg do EBITDA (ang. net debt to EBITDA), ki prikazuje sposobnost podjetja servisiranja svojega dolga. Kazalnik primerja neto dolg z dobičekom pred obrestimi, davki in amorizacijo ter je izračunan kot:

$$\text{neto dolg do EBITDA} = \frac{\text{neto dolg}}{\text{EBITDA}}.$$

Pri tem kazalniku nižje vrednosti med podjetjema v isti industriji prikazujejo bolj konzervativno zadolženost, torej bolj stabilno podjetje. Vrednosti okoli štiri ali pet pa kažejo na velike dolžniške težave podjetja. Kazalnik namreč prikazuje pričakovano obdobje, v katerem lahko podjetje ob enakem poslovanju odplača svoj dolg.

Pokritost dolga z denarnimi tokovi iz poslovanja (ang. current cash flow to debt) nam poda kazalnik denarnih tokov iz poslovanja do celotnega dolga. Prikazuje sposobnost podjetja za odplačevanje dolga glede na generirani denarni tok. Višja vrednost kazalnika prikazuje boljše podjetje. Je eden pomembnejših kazalnikov pri odločitvah refinanciranja dolga s strani kreditodajalca. Izračunamo ga kot:

$$\text{denarni tok iz poslovanja do celotnega dolga} = \frac{\text{denarni tok iz poslovanja}}{\text{celotni dolg}}.$$

Tu je celotni dolg sestavljen iz obresti, kratkoročnega dolga ter dela dolgoročnega dolga, ki ga še moramo odplačati (pri dolgoročnih dolgovih banki navadno odplačujemo glavnico v delih). Kazalnik denarnega toka iz poslovanja do celotnega dolga je v grobem ravno obratna vrednost kazalnika neto dolga do EBITDA. Prva vsebinska razlika izhaja iz tretiranja dolga (bruto in neto dolg), glede ustreznosti pa si strokovnjaki niso enotni.

5.2.3 Kazalniki dobičkonosnosti

Generiranje dobička je enden izmed glavnih ciljev delovanja podjetij v tržnem gospodarstvu. S svojo dejavnostjo morajo podjetja namreč plemenititi denar lastnikov, hkrati pa z njihovo dobičkonosnostjo pada tveganje finančnih težav. Kazalniki profitabilnosti prikazujejo zmožnost podjetja za generiranje denarnih tokov ter dobička.

Gre za skupino, v kateri lahko zasledimo največje število kazalnikov, vsem pa je skupno dejstvo, da višja vrednost kazalnika prikazuje boljše podjetje, saj le to generira več dobička. Tudi kazalniki dobičkonosnosti nimajo večjega pomena, v kolikor jih ne pogledamo v relaciji z vrednostimi povprečnega kazalnika v industriji v kateri se podjetje nahaja. Izrazito se podjetja v različnih industrijah lahko razlikujejo že zaradi sezonskosti prihodkov.

EBITDA je vrednost dobička, preden mu odštejemo stroške obresti, davkov ter amortizacije. Čeprav ni kazalnik v najstrožjem pomenu besede, pa predstavlja približek denarnega toka podjetja in je zato ena ključnih spremenljivk pri modeliranju propada podjetja, kot bomo videli v nadaljevanju. Že njegova prva definicija in uporaba, v osemdesetih letih devetnajstega stoletja, je bila kot mera sposobnosti podjetja za odplačilo dolga. Njegova vloga pa je od tedaj le rastla na vseh področjih.

Dobičkonosnost kapitala (ang. eeturn on equity - ROE) je mera, ki pove, koliko čistega dobička je ustvarjeno z enoto vloženega kapitala v enem letu. Izračunamo ga kot:

$$\text{ROE} = \frac{\text{dobiček}}{\text{povprečni kapital}}.$$

Višja vrednost kazalnika za lastnike navadno pomeni višji donos, vendar zelo visoke vrednosti lahko nakazujejo na visoko zadolženost in zato pomenijo višje tveganje. Pri spremljanju ROE kazalnika v času moramo biti zelo previdni, saj se lahko znajdemo v situaciji, ko ROE določenega podjetja narašča zaradi manjšega proncentualnega znižanja dobička kot povprečnega kapitala. Obstajajo namreč načini, ko nižji kapital ni posledica izgube, ki se prikaže v izkazu poslovnega izida (na primer izplačila dividend lastnikom, določena prevrednostenja sredstev). ROE je ena najpomembnejših mer za lastnike podjetja, ni pa bistvenega pomena za kreditodajalce, prav zato tudi ni del nobenega izmed končnih modelov.

Drugi kazalnik, ki ga bomo pri modeliranju uporabili, je dobičkonosnost prihodkov (ang. return on sales - ROS). Glavna naloga kazalnika je meriti, koliko dobička podjetje obdrži na vsako enoto prihodka. Definiran je s formulo:

$$\text{ROS} = \frac{\text{dobiček}}{\text{celotni prihodki}}.$$

Njegova moč se prikaže pri primerjanju kazalnika med podjetji v dani industriji, saj višja vrednost kazalnika nakazuje na boljše upravljanje podjetja s stroški. Hkrati pa moramo biti pri interpretaciji pazljivi, saj je lahko nižja vrednost kazalnika posledica hitrejšega višanja prihodkov od višanja dobička. To pa za sam obstoj podjetja ni pretirano problematično.

Zelo podoben kazalnik je tudi dobičkonosnost poslovnih prihodkov (ang. operating profit margin), katerega naloga je prav tako meriti operativno učinkovitost delovanja podjetja. Definiran je s formulo:

$$\text{Dobičkonostnost poslovnih prihodkov} = \frac{\text{dobiček iz poslovanja}}{\text{celotni prihodki}}.$$

Na ta način predstavlja delež dobička iz poslovanja na vsako enoto prihodkov. Zaradi vključitve dobička iz poslovanja so določeni enkratni učinki izključeni. Na ta način pa dobimo bolj dolgoročno sliko stanja podjetja. V danem letu imamo namreč lahko izjemen dogodek, ki se ne nanaša na našo osnovno dejavnost, vendar nam izrazito dvigne dobiček. Seveda le tega ne moremo pričakovati tudi v naslednjih letih, zato nam dobičkonosnost poslovnih prihodkov podaja realnejšo sliko stanja podjetja.

5.2.4 Kazalniki učinkovitosti

V nasprotju s kazalniki dobičkonosnosti, ki so nam prikazovali, koliko denarja podjetje generira na enoto investiranega, nam kazalniki učinkovitosti podajajo sliko o tem, kako dobro podjetje upravlja s svojimi sredstvi. Zaradi svoje naloge so tako veliko bolj specifični ter ne podajajo celotne slike poslovanja. Nam pa prav zaradi tega lahko veliko bolj natančno predstavijo določen segment operativne učinkovitosti podjetja. Izboljšanje kazalnikov učinkovitosti pa se navadno že obdobje po tem odrazi tudi v izboljšani dobičkonosnosti (seveda ob predpostavki, da so vse ostale komponente enake). Prav zato lahko rečemo, da imajo napovedovalno moč, v kolikor le v primeru, da so uporabljeni v pravilnem kontekstu. Za slednje je nujno, da kazalnike primerjamo z njihovimi konkurenti. V kolikor bi jih primerjali med vsemi podjetji, bi bila zaradi šumov v podatkih kakršnakoli obdelava otežena.

Prvi kazalnik učinkovitosti, opisan v tej magistrski nalogi, je obračanje celotnih sredstev (ang. asset turnover). V svojem bistvu pojasnjuje sposobnost generiranja prihodkov na enoto sredstev v danem letu in je izračunan kot

$$\text{obračanje celotnih sredstev} = \frac{\text{prihodki}}{\text{povprečna sredstva}}.$$

Pričakovana vrednost kazalnika je odvisna od industrije, v kateri se nahaja, v splošnem pa velja, da višji kazalnik prikazuje boljša podjetja. Za konkretno podjetje pa je lahko zanimiv tudi časovni trend kazalnika.

Nekoliko drugačno sliko pa nam podaja obrat poslovnih terjatev (ang. accounts receivable turnover). Kazalnik prikazuje, koliko prihodkov iz posovanja podjetje ustvari na enoto sredstev, vezanih v poslovnih terjatvah. V primeru da število dni v koledarskem letu delimo s kazalnikom obrata poslovnih terjatev, dobimo povprečno število dni vezave poslovnih terjatev (koliko časa traja, da nam v povprečju poslovni partner plača svoje obveznosti). Kazalnik obrata poslovnih sredstev je izračunan kot:

$$\text{obrat poslovnih terjatev} = \frac{\text{prihodki od poslovanja}}{\text{povprečne terjatve iz poslovanja}}.$$

Nižja vrednost kazalnika glede na konkurenco nakazuje na težave ter zahteva natančen pregled poslovanja. Pri višji vrednosti kazalnika je potrebno posebno pozornost posvetiti večjim spremembam v primerjavi s prejšnjimi leti. Nezaželjeno zvišanje je lahko sprememba v agresivnosti kreditne politike. Rastoči prihodki iz poslovanja nam ne pomagajo prav dosti, če se ob tem bistveno zviša delež slabih terjatev.

Med kazalnike učinkovitosti spada tudi obrat poslovnih obveznosti (ang. accounts payables turnover). Kazalnik je konceptualno identičen kazalniku obrata poslovnih terjatev, vendar se pri obratu poslovnih obveznosti osedotočimo na dobavitelje. Zanima nas, kako hitro podjetje plača svoje poslovne obveznosti. Pri interpretaciji je ključna primerava znotraj panoge. Visoka vrednost kazalnika za dobro (ali dovolj veliko) podjetje nakazuje, da ima veliko pogajalsko moč in lahko zahteva daljši odlog plačila (dobavitelj mu financira obratni kapital). Hkrati pa je lahko visoka vrednost kazalnika za podjetje v težavah znak, da svojih poslovnih obveznosti ne izpolnjuje pravočasno. S tem si na kratek rok zagotovi tekočo likvidnost, po drugi strani pa uničuje odnose s partnerji, ki lahko na določeni točki ustavijo dobave podjetju. Kazalnik je definiran kot:

$$\text{obrat poslovnih obveznosti} = \frac{\text{stroški prodanega materiala}}{\text{povprečne obveznosti iz poslovanja}}.$$

Tudi tu lahko izračunamo povprečno število dni vezave poslovnih obveznosti na način, da število dni v koledarskem letu delimo s kazalnikom. Izračunana vrednost prikazje povprečno število dni potrebnih, da podjetje plača svojim dobaviteljem.

Predstavili smo le tri najbolj pogoste kazalnike učinkovitosti, vendar smo že pri teh spoznali, kako zelo različni so lahko med sabo. To se zgodi prav zaradi osredotočanja kazalnika na točno določen segment delovanja podjetja.

5.3 Podatki

Čeprav je morda intuitivno, da bo banka za modeliranje uporabila lastno podatkovno bazo, pa se pri taki bazi pojavljajo težave zaradi pristranskosti. Izbira baze je torej odvisna od želja bank. V kolikor želijo pripraviti splošen model, ki ga bodo lahko uporabili tudi za podjetja brez zgodovine s to banko, uporabljajo podatkovne baze, kot so Bloomberg, Capital IQ ali podobne. Svojo interno podatkovno bazo pa uporabljajo v primerih, ko želijo bolj specifičen model. V zadnjih letih se zaradi vse večje regulacije na področju upravljanja s tveganji vedno več bank odloča, da razvijejo splošen model. Regulator nato model preveri ter odobri. Po potrebi se ga aplicira na notranje podatkovne baze ter preveri njegovo obnašanje.

V tem magistrskemu delu bo moja podatkovna baza sestavljena iz podjetji, ki sestavljajo indeks Russell 3000. Gre za indeks, ki je nastal leta 1984 ter zajema 3000 največjih podjetji s sedežom v Združenih državah Amerike, glede na tržno kapitalizacijo (ang. market capitalization) ter likvidnost. Indeks zajema 98% celotne tržne kapitalizacije ameriškega delniškega trga, zato se ga velikokrat uporablja kot replikacija likvidnega dela ameriškega trga. Indeks je utežen glede na velikost podjetja, kar v praksi pomeni, da 1000 največjih podjetji predstavlja približno 92% vrednosti celotnega indeksa. Indeks je vsako leto v maju spremenjen glede na novo tržno kapitalizacijo podjetij, med letom pa se nato ne spreminja. V primeru da določeno podjetje propade, je odkupljeno z borze ali na kakšen drugi način ne kotira več na borzi, ga enostavno izbrišejo iz nabora. Bolj znani sta njegovi izpeljaniki Russell 1000, ki predstavlja 1000 največjih podjetji, in ga uporaljamo kot mero spreminjanja vrednosti velikih podjetji, ter Russell 2000, ki predstavja najmanjših 2000 podjetji indeksa Russell 3000. Sledenje indeksu Russell 2000 je zelo priljubljeno med upravljalci pasivnih vzajemnih skladov, ki želijo replicirati obnašanje malih

Kategorija	Ime kazalnika	Formula
Kazalniki likvidnosti	Trenutna likvidnost	$\frac{\text{denar+ finančne naložbe}}{\text{kratkoročne obveznosti}}$
	Količnik obratno-finančne likvidnosti	$\frac{\text{trenutna sredstva- zaloge}}{\text{kratkoročne obveznosti}}$
	Količnik splošne likvidnosti	$\frac{\text{trenutna sredstva}}{\text{kratkoročne obveznosti}}$
Kazalniki zadolženosti	Neto dolg do EBITDA	$\frac{\text{neto dolg}}{\text{EBITDA}}$
	Denarni tok iz poslovanja do celotnega dolga	$\frac{\text{denatni tok iz poslovanja}}{\text{celotni dolg}}$
	Celotni dolg do EBITDA	$\frac{\text{celotni dolg}}{\text{EBITDA}}$
	Delež dolga v sredstvih	$\frac{\text{dolgoročni dolg}}{\text{celotna sredstva}}$
	Delež dolga v financiranju	$\frac{\text{celoten dolg}}{\text{celotna sredstva}}$
	Pokritost obresti	$\frac{\text{EBIT}}{\text{stroški obresti}}$
	Delež dolga v lastniškem kapitalu	$\frac{\text{neto dolg}}{\text{celoten kapital}}$
Kazalniki dobičkonosnosti	Dobičkonostnost kapitala (ROE)	$\frac{\text{dobiček}}{\text{povprečni kapital}}$
	Dobičkonostnost sredstev (ROA)	$\frac{\text{dobiček}}{\text{povprečna sredstva}}$
	Dobičkonostnost prihodkov (ROS)	$\frac{\text{dobiček}}{\text{celotni prihodki}}$
	Dobičkonosnost poslovnih prihodkov	$\frac{\text{dobiček iz prihodkov}}{\text{celotni prihodki}}$
	Bruto marža	$\frac{\text{prihodki - stroški osnovnih sredstev}}{\text{celotni prihodki}}$
	EBITDA marža	$\frac{\text{EBITDA}}{\text{celotni prihodki}}$
	EBIT marža	$\frac{\text{EBIT}}{\text{celotni prihodki}}$
Kazalniki učinkovitosti	Obračanje celotnih sredstev	$\frac{\text{prihodki}}{\text{povprečna sredstva}}$
	Obračanje fiksnih sredstev	$\frac{\text{prihodki}}{\text{fiksna sredstva (posest, stroji, ...)}}$
	Obračanje zalog	$\frac{\text{prihodki od poslovanja}}{\text{povprečne zaloge}}$
	Obrat poslovnih terjatev	$\frac{\text{prihodki od poslovanja}}{\text{povprečne terjatve iz poslovanja}}$
	Obrat poslovnih obveznosti	$\frac{\text{stroški prodanega materiala}}{\text{povprečne obveznosti iz poslovanja}}$
	Prihodki na enoto obratnih sredstev	$\frac{\text{prihodki}}{\text{(kratkoročna sredstva - kratkoročne obveznosti)}}$

Tabela 2: Kazalniki ter njihove formule.

podjetij. Dilema se morda pojavlja tudi, zakaj nismo vzeli katerega izmed indeksov S&P, saj je S&P 500 eden najpopularnejših delniških indeksov. Kljub vsemu pa največji od S&P indeksov, S&P 1500, vsebuje "le" 1500 podjetij. To pa pomeni, da izgubimo številna manjša podjetja. Prav pri teh se težave pokažejo prej ter so zato bolj atraktivna za modeliranje tveganja propada podjetja. V kolikor bi želeli izpeljati PD model za druge bodisi razvite bodisi države v razvoju, bi vzeli drugo podatkovno bazo. Sama sem se odločila za indeks ameriških podjetij, saj nam le ta zagotavlja najmanj mankajočih podatkov. Podatki so pridobljeni iz podatkovne baze Bloomberg, dne 15.1.2015.

Osnovni podatki torej zajemajo 3000 podjetij, ki so bila vključena v indeks leta 2013 ter njihove osnovne kazalnike glede na pomen, kot prikazano v Tabeli 2.

Poleg teh sta dodani vrednosti tržne kapitalizacije v letu 2013 ter 2014. S pomočjo kvocienta tržnih vrednosti, ki pomeni spremembo vrednosti delniškega kapitala v letu 2014 v primerjavi z letom 2013, bomo namreč modelirali spremenljivko, ki bo določila, ali podjetje imenujemo za "propadlo". Seveda težko katero izmed podjetij, ki kotirajo na borzi, propade v enem letu, vendar je pristop modeliranja za leto vnaprej najbolj uporaljen v bankah. Njegovo dominanco je ustalil Basel II, ki zahteva poročanje kapitalske strukture banke na letni ravni. V tem vključuje tudi vsakoletno vrednotenje vseh naložb. Prav podjetja, ki jih tu privzamemo za propadla so namreč taka, ki jih morajo banke v svojih naložbah slabiti. Zato zanjo predstavljajo podobno težavo kot propadla podjetja. Leto 2013 je na ameriškem delniškem trgu zaznamovala visoka rast vredosti delniškega premoženja. Tako je vrednost indeksa Russell 3000 v letu 2014 za 19.58% višja kot leto poprej. Za podjetja v težavah bomo oklicali vsa podjetja, katerih je tržna vrednost padla vsaj za 30%. Seveda bi morali v letih finančne krize ali stagnacije to mejo spustiti nižje, vendar se moramo zavedati, da so ta podjetja v relativnem smislu izgubila skoraj 50% vrednosti, kar jasno nakazuje na velike težave. S to definicijo konstruiramo binomsko slučajno spremenljivko, ki ima vrednost ena, v kolikor je podjetje propadlo, ter nič sicer. Od 3000 podjetij jih za propadla označimo 301, kar predstavlja 10.83% trga. Za definicijo propada podjetja bi lahko vzeli tudi ocene tveganosti bonitetnih agencij. Opisali smo že njihovo ocenjevanje ter zakaj le te ocene niso primerne za napovedi propada podjetja za banko. Prav dejstvi, da so ocene premalo odzivne na spremembe na trgu ter njihova prisotnost le pri določenih podjetjih, nas tudi tu odvrneta od uporabe ocen tveganja za modeliranje spremenjivke propada podjetja v določenem letu.

Za vsako podjetje pa je poleg vrednosti kazalnikov podan tudi vektor, ki določa kateri industriji glede na osnovno dejavnost podjetje pripada. Skupaj je tako 43 industrij, pri čemer je največ predstavnikov v industriji energetike, kamor spadajo tudi podjetja, ki se ukvarjajo s pridobivanjem nafte ter zemeljskih plinov. Prav v tej industriji je tudi bilo označenih za propadla največ, kar 34.7% podjetij, kar predstavlja 33 podjetij. Kar šest izmed industrij pa nima podjetja, ki bi propadlo, med temi je tudi najmanjša industrija z 12 podjetij, ki se ukvarjajo s finančnim svetovanjem.

Že pri izpeljavi HSEM modela sem kot eno izmed glavnih težav omenila, da model ni robusten na manjkajoče podatke. Ena od možnosti je le te dopolniti s pomočjo različnih statističnih metod. Načinov za dopolnitev je veliko, vsak pa ima svoje prednosti in slabosti. Na tem delu pa se bomo osredotočili samo na podjetja, ki imajo dane vse podatke. Čeprav se na prvi pogled zdi, da tak pristop pripelje do pristranskosti, sem ob nekoliko natančnejši analizi podatkov ugotovila, da za manjkajoče podatke res lahko rečemo, da manjkajo naključno v vseh kategorijah približno enako. Tako nam od začetnih 3000 podjetij ostane 910, kar predstavlja 30.33% začetne baze. Od teh podjetij jih je kot propadlo podjetje klasificirano 106, kar predstavlja 11.45% podjetij. Ta delež je zelo podoben prvotnim 10.83%. Čeprav je delež propadlih podjetij v končni bazi res nekoliko večji, pa nam bo to pri modeliranju prej koristilo kot škodilo. Na ta način bomo imeli namreč boljšo kontrolno skupino. Poleg tega pa so vse spremenljivke standardizirane (odšteli smo jim povprečno vrednost ter delili s standarnim odklonom). Tako imajo vse povprečje nič, na ta način pa ima tudi konstanta jasno interpretacijo. To torej postane vrednost, ko so vse spremenljivke v svojem povprečju (naravni legi). Poleg tega je standardiziranje priporočeno, da ne bi katera izmed spremenljivk samo zaradi svoje skale pretehtala drugo, ki bi morala imeti v modelu večji pomen. Vse spremenljivke so razvrščene na isto skalo s pomočjo deljenja s standardnim odklonom.

V nadaljevanju te magistrske naloge bomo namreč s pomočjo finančnih kazalnikov iz leta 2013 izpeljali modele ter jih primerjali s stanjem v letu 2014, torej ali je v letu 2014 podjetje že oklicano za propadlega. Začeli bomo z osnovnejšimi modeli (probit, logit) ter nadaljevaljevali s kombinacijo SEM modela ter GLMM, na koncu pa le tega primerjali z HSEM modelom. Pričakujem, da se bosta tako logit kot probit model spopadala s težavami multikolinearnosti. Težave bodo lahko rešene s pomočjo SEM modela s kombinacijo s probit model. Prav tako pričakujem, da bo pojasnjevalna moč bistveno večja, ko bomo vključili tudi efekt industrij, še večja pa v HSEM modelu, ko bo model estimiran s pomočjo enega modela in ne kombinacije dveh.

Celotna analiza bo izvedena v programskem jeziku Rstudio, natančneje s pomočjo paketov: lavaan za modeliranje SEM modela ter lme4 za modeliranje GLM ter GLMM modela (uporablja se funkciji glm ter glmer). Seveda pa to ni edini programski jeziki, ki izvaja SEM model. Najbolj znani poleg tega so LISREL, ki je bil tudi prvi programski jezik za izvajanje SEM modela, ter Mplus. Rstudijo je bil izbran zgolj zaradi dejstva, da gre za odprto kodni sistem. Žal pa v jeziku R še vedno ni razvite knjižnice, ki bi omogočala modeliranje HSEM modela, zato bo le ta modeliran s pomočjo gllamm odprte knjižnice, ki za svojo uporabo zahteva programski jezik Stata z verzijo 11 ali višjo. S pomočjo State in programa GLLAMM bomo vsakemu podjetju priredili njegovo verjetnost propada. Podatkovna baza bo nato ponovno prenešena v programski jezik R, kjer sem implementirala funkcijo za izračun Ginijevega koeficienta ter ROC krivuje. S pomočjo te funkcije nato določamo natančnost prileganja.

6 Izpeljava modela za napoved verjetnosti propada podjetja

6.1 Probit ter logit model

Pri modeliranju diskretne odvisne slučajne spremenljivke z zveznimi sta najpopularnejši orodji probit ter logit model. Oba sta posebna primera GLM modela, vendar je zgodovinsko gledano probit starejšega izvora. Izpeljal ga je Chester Bliss 1934, ki je predpostavljal binarno odvisno slučajno spemenljivko Y. Predpostavimo še, da so vse pojasnjevalne slučajne spremenljivke zbrane v matriki X, potem je probit definiran z enačbama:

$$Y^* = X'\beta + \epsilon,$$
$$Y = \begin{cases} 1 & \text{če je } Y^* > 0 \\ 0 & \text{sicer} \end{cases}.$$

Pri tem je β vektor koeficientov ter ϵ vektor napak, ki je porazdeljen $\epsilon \sim N(0, 1)$. Zgornjo enačbo lahko enakovredno zapišemo tudi v obliki:

$$\begin{aligned} P(Y = 1|X) &= P(Y^* > 0) \\ &= P(X'\beta + \epsilon > 0) \\ &= P(\epsilon > -X'\beta) \\ &= P(\epsilon < X'\beta) = \Phi(X'\beta) \end{aligned}$$

kjer predzadnja enakost velja zaradi simetrije normalne porazdelitve. Za pridobitev rezultatov nato uporabimo metodo največjega verjetja ali številne druge aproksimacijske postopke, ki so na voljo.

Poglejmo si sedaj primer PD modela v probit obliki. Odločili smo se, da ne bomo vključili nobenega kazalnika učinkovitosti, saj so le ti primerljivi predvsem med posameznimi podjetji znotraj industrij na kar pa s klasičnem probit modelu nimamo vpliva. Iz preostalih skupin izberemo kazalnike, ki so skupaj s svojimi keficienti ter statistikami prikazani v Tabeli 3. Poleg tega je bil v model vključena kot spremenljivka vektor standardiziranih vrednosti EBITDA. Razlog vključitve le tega je njegova sposobnost merjenja velikosti podjetja, le to pa je tudi pomembno pri verjetnosti propada (večja podjetja težje propadejo). Historčno gledano je ta boljša mera za vpliv velikosti podjetja od spremenljivke tržne vrednosti delic, ki bi bila mogoče bolj pričakovana. Razlog leži v variabilnosti cene delnic pri določenih podjetjih, ki pa je neodvisna od samega poslovanja podjetja. Na ta način torej vključimo nezavajujočo oceno velikosti samega podjetja.

Vidimo lahko, da je kazalnik neto dolg do EBITDA statično neznačilen ter ima

	koeficienti	standardna napaka	p vrednost
Konstanta	-1.308949	0.071772	$< 2e-16 ***$
Neto dolg do EBITDA	-0.007401	0.052594	0.88810
ROS	-0.279467	0.105577	0.00812 **
Trenutna likvidnost	-0.221289	0.089088	0.01299 *
EBITDA	-0.561938	0.229869	0.01450 *
Dobičkonosnost posl. prih.	0.269569	0.104044	0.00957 **

Tabela 3: Probit model.

hkrati negativen predznak koeficienta. Prav slednje nima nobenega ekonomskega pomena, saj si težko predtstavljamo, kako višji dolgovi ob konstantnem EBITDA oziroma nižji EBITDA ob konstantnih dolgovih lahko predstavlja boljše podjetje, torej manj verjetno, da propade. Še bolj zavajujoče podatke pa dobimo pri kazalniku dobičkonosnosti poslovnih prihodkov. Le ta je statistično značilen pri $\alpha = 5\%$ vrednosti. Hkrati pa njegov predznak nima ekonomskega pomena. Ni namreč zelo verjetno, da bi višji dobiček iz prihodkov ob konstantnih celotnih prihodkih prikazoval težave podjetja. Navadno v primerih, ko imamo le eno statistično neznačilno spremenljivko, postopamo tako, da jo izključimo iz modela. Le to pa nas privede do še bolj zavajujočih podatkov, prikazanih v Tabeli 4. Tu ima kazanik dobičkonosnosti

	koeficienti	standardna napaka	p vrednost
Konstanta	-1.30876	0.07174	$< 2e-16 ***$
ROS	-0.27756	0.10455	0.00794 **
Trenutna likvidnost	-0.22000	0.08846	0.01289 *
EBITDA	-0.56167	0.22981	0.01452 *
Dobičkonosnost posl. prih.	0.26815	0.10343	0.00953 **

Tabela 4: Probit model s štirimi spremenljivkami.

poslovnih prihodkov še vedno pozitiven predznak, hkrati pa so vse spremenljivke statistično značilne. Tako bi načeloma model lahko sprejeli, čeprav ne bi imel ekonomske podlage. Poleg tega pa smo na ta način zmanjšali pojasnjevalno moč modela. Izključili smo namreč edini kazalnik zadolženosti v modelu. Zadolženost pa je ključnega pomena pri modeliranju verjetnosti propada.

Zaradi zgoraj opisanih težav se odločimo poskusiti z logit modelom. Oba modela namreč predpostavljata binomsko slučajno spremenljivko. Osnoven logit model opisujeta enačbi:

$$\ln(odds(p)) = X'\beta + \varepsilon,$$

kjer je torej

$$odds(p) = \frac{p}{1-p}.$$

Tu p predstavlja verjetnost, da se dogodek (v našem primeru propad podjetja) zgodi. Vektor koeficientov je označen z β, matrika pojasnjevalnih spremenljivk z X ter vektor napak z ε. V kolikor enačbi nekoliko preuredimo, dobimo pogojno verjetnost izraženo z:

$$P(Y=1|X) = [1 + \exp(-X'\beta)]^{-1}.$$

Vidimo, da ima logit model v primerjavi s probit modelom "manjše" repe. Model je največkrat nato aproksimiran z metodo največjega verjetja. Na tem mestu je podan samo osnovni opis probit ter logit modela. Bolj podroben opis obeh modelov je izpeljan v Damodar (1995) ali številnih drugih ekonometričnih priročnikih.

V kolikor zgornji model ponovimo v logit obliki, dobimo rezultate prikazane v Tabeli 5. Tudi logit model pa težav ne odpravlja, še več, napake se pojavljajo na istih me-

	koeficienti	standardna napaka	p vrednost
Konstanta	-2.29265	0.15356	$< 2e-16$ ***
Neto dolg do EBITDA	-0.01799	0.09448	0.8489
ROS	-0.49092	0.18683	0.0086 **
Trenutna likvidnost	-0.42664	0.17843	0.0168 *
EBITDA	-1.24562	0.51668	0.0159 *
Dobičkonosnost posl. prih.	0.47620	0.18855	0.0116 *

Tabela 5: Logit model

stih kot pri probit modelu. Na ta način bi lahko sprejeli model, ki bi imel ekonomsko povsem nelogične koeficiente ter velike težave s statističnimi neznačilnostimi. Glavni razlog vseh težav pa se skriva v visoki korelaciji med kazalnikoma ROS ter profitne marže, kot je prikazano v Tabeli 6. Gre torej za avtokorelacijo med slučajnimi spremenljivkami. Tako probit kot tudi logit model pa predpostavljata neodvisnost vseh slučajnih spremenjivk. Vidimo pa, da sta si kazalnika podobna, saj je korelacijski koeficient med njima kar 0.8321. To doprinese še težave z multikolinearnostjo, torej napačno specifikacijo modela. Vendar vidimo, da tudi z izključitvijo kazalnika neto dolga do EBITDA, ki je statistično najmanj značilen, težav ne rešujemo. Potrebno bi se bilo odločati med izklučitvijo ali kazalnika profitne marže ali pa kazalnikom ROS. Na glavni vir težav pa nas nista opozorila niti probit niti logit model. Modela sta si podobna tudi v njuni pojasnjevalni moči. Pri probit modelu je Ginijev pojasnjevalni koeficient enak 0.2502177, pri logit modelu pa 0.253079. Oba torej pojasnita približno 25% več kot naivni model.

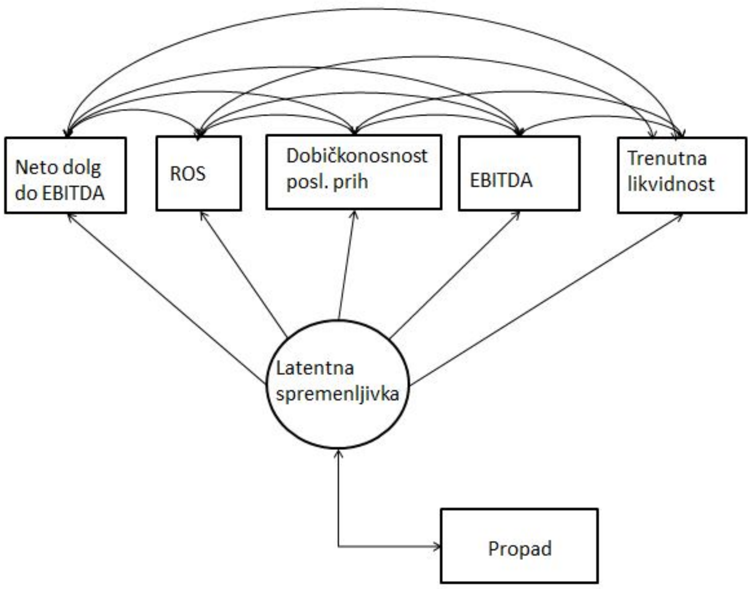
Neto dolg do EBITDA	1	-0.0675	-0.1002	-0.0114	-0.0133
ROS	-0.0675	1	0.1667	0.2407	0.8321
Trenutna likvidnost	-0.1002	0.1668	1	0.0102	0.1502
EBITDA	-0.0114	0.2407	0.01021	1	0.2332
Dobičkonosnost posl. prih.	-0.0134	0.8321	0.1502	0.2332	1

Tabela 6: Korelacijski koeficienti med spremenljivkami.

6.2 SEM model

V začetku smo SEM model izpeljali prav z motivacijo, saj lahko modelira tudi spemenljivke z visoko koreliranostjo. Žal pa SEM model ne more direktno modelirati diskretnih odvisnih spremenljivk. Prav to je bil enden od razlogov, da smo izpeljali tudi formule za oceno latentnih spremenljivk.

Najprej bomo iz vseh pojasnjevalnih spremenljivk izpeljali novo latentno spremenljivko, le-to pa bomo nato vstavili v probit model. S pomočjo analize po poti bi lahko model predstavili z diagramom 4. Tu so jasno predstavljene vse relacije med spremenljivkami. Model torej predpostavlja, da variančno-kovariančna matrika vseh pojasnjevalnih spremenljivk ni enaka identiteti, torej že pri modeliranju predpostavljamo korelacijo med spremenljivkami. S pomočjo Rstudia ter paketa laavan



Slika 4: Prikaz relacij v SEM modelu.

dobimo rezultate o koeficientih, hkrati pa tudi minimizirano variančno-kovariančno matriko. Slednje na tem mestu ne bomo prikazali, saj njena interpretacija na tej točki nima dodane vrednosti. Vidimo, da so vse spremenljivke statistično značilne,

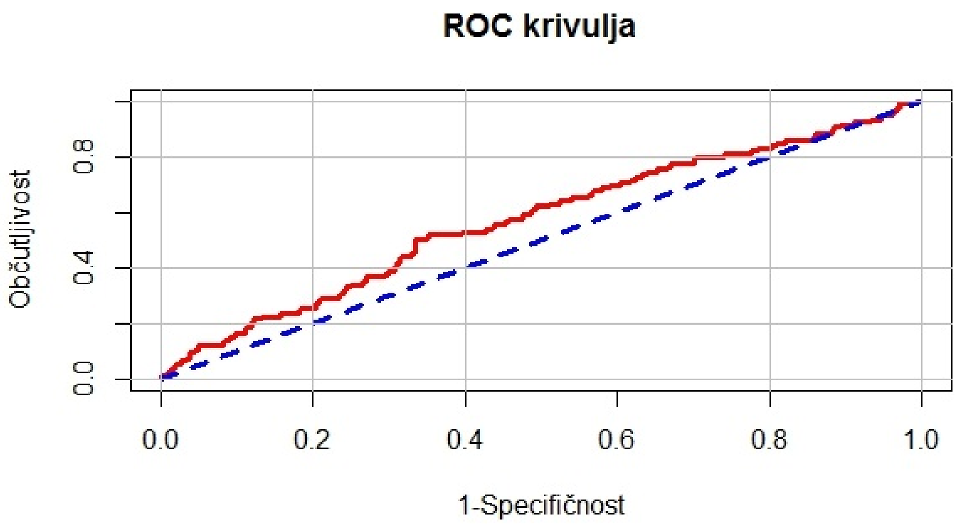
sprem =	koeficienti	standardna napaka	p vrednost
EBITDA	0.256	0.034	0.000
Dobičkonosnost posl. prih	0.873	0.039	0.000
Neto dolg do EBITDA-ja	-0.058	0.034	0.089
ROS	0.952	0.040	0.000
Trenutna likvidnost	0.174	0.034	0.000

Tabela 7: SEM model.

	koeficienti	standardna napaka	p vrednost
Konstanta	-1.2007	0.0548	$< 2e - 16 ***$
Sprem	-0.1210	0.0585	0.0388 *

Tabela 8: Probit model.

edino težava pri značilnosti ostaja pri neto dolgu do EBITDA. Kazalec je značilen le v primeru, ko predpostavljamo $\alpha = 10\%$. Kljub tem manjšim težavam pa so vsi predznaki koeficientov podprti z ekonomsko teorijo. Koeficienti pri kazalnikih profitabilnosti ter likvidnosti pozitivno vplivajo na novo spremenljivko, le ta pa nato negativno vpliva na verjetnost propada. Tudi spremenljivka EBITDA ima negativen vpliv, kar prikazuje, da večja podjetja z višjim denarnim tokom težje propadejo in je le to pričakovano. Kazalnik zadolženosti pa po pričakovanjih pozitivno vpliva na samo verjetnost propada. Tako se vsaj na prvi pogled zazdi, da lahko takemu modelu zaupamo. Vsekakor pa je za uporabo modela potrebno še nekoliko več kot zgolj

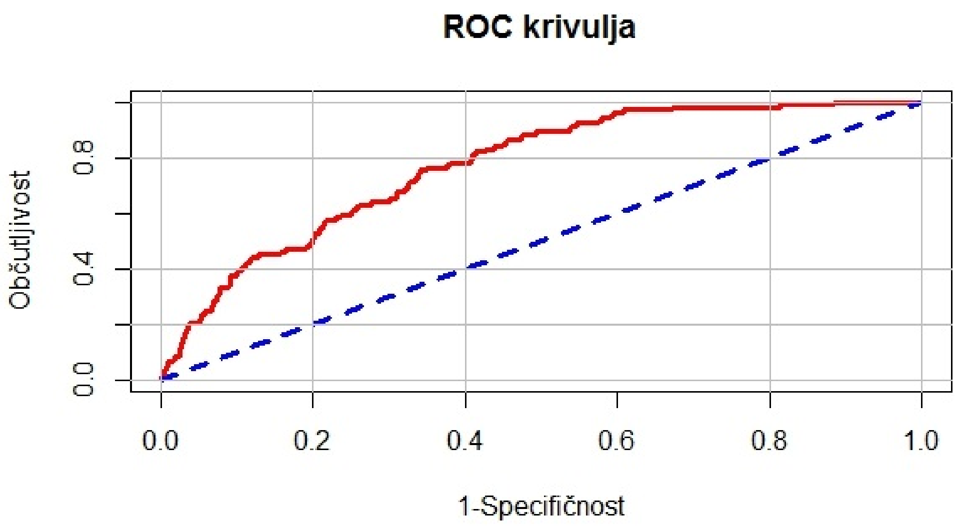


Slika 5: Roc krivulja za SEM model v kombinaciji s probit modelom.

predznak koeficienta. Pregledati je potrebno tudi njegovo prileganje podatkom ter napovedovalno moč. Pri izračunu ROC krivulje ter Ginijevega koeficienta opazimo, da je napovedovalna moč skoraj ničelna. Tu je ROC krivulja predstavljena s polno črto, za primerjavo pa je dodano prileganje naivnega modela s črtkano črto. Na Sliki 5 vidimo, da model le malenkostno bolje napove propad od naivnega modeliranja, to dokazuje tudi Ginijev koeficient, ki znaša 0.1286. Le to je bistveno nižje od probit ali pa logit modela, saj smo sedaj vse koeficiente združili v eno spremenljivko. Težava združevanja v le eno spremenljivko bo odpravljena z vpeljavo HSEM modela.

Pri iskanju modela, ki se bolje prilega našim podatkom, bomo začeli podobno kot prej. Najprej bomo s pomočjo SEM modela kreirali novo latentno spremenljivko, le to pa bomo nato postavili v GLMM model s probit vezno funkcijo. Hierahično strukturo podatkov lahko enostavno podpremo z enkonomsko teorijo. Vsako podjetje namreč pripada eni izmed vnaprej določenih 43 industrij, zato želimo slučajne efekte razločevati glede na ta podatek. Na ta način bomo podatke modelirali na dveh ravneh ter na ta način doprinesli k boljši specifikaciji modela.

Na Sliki 6 je vidno, da je prileganje sedaj bistveno boljše. Ginijev koeficient tako



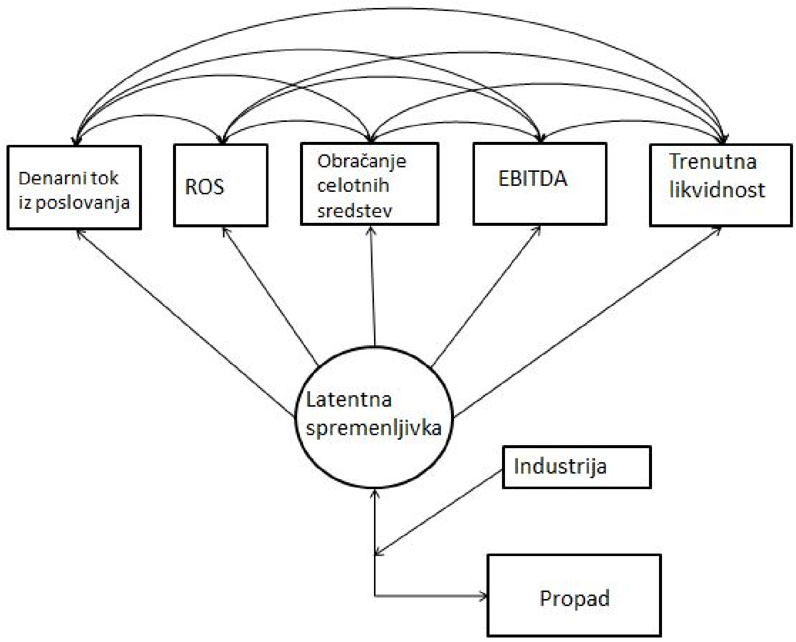
Slika 6: Roc krivulja za SEM model v kombinaciji s GLMM modelom.

doseže vrednost 0.4797, kar je bistveno več kot za primer bodisi kombinacije SEM modela s probit funkcijo bodisi probit ali logit modela. Izpeljali bomo seveda še bolj natančen model, ki bo v vključeval kazalnike iz vseh štirih skupin. Na tem mestu smo želeli samo pokazati, kako nezanesljiva sta lahko probit ter logit model, ki zato ne bi smela biti predmet uporabe pri modeliranju PD modela. V nadaljevanju pa smo prikazali pomen večrazsežnega vzorčenja za boljše prileganje.

6.3 Primerjava HSEM modela s kombinacijo GLMM ter SEM modela

V tem delu si bomo ogledali, kolikšen je doprinos vpeljave direktnega HSEM modela v primerjavi s kombinacijo GLMM ter SEM modela. Prav tako kot pri GLMM modelu lahko tudi v HSEM modelu ločimo podjetja glede na industrijo, v kateri deluje. Hkrati pa nam HSEM model omogoča, da kazalnike vključimo direktno v model, ne pa prek posredne latentne spremenljivke, ki je bila do sedaj generirana s pomočjo SEM modela.

Shematično bi modela predstavili s Sliko 7 ter Sliko 9, kjer so jasno prikazane relacije med spremenljivkami. V želji po ohranitvi razmeroma enostavnega modela



Slika 7: GLMM ter SEM model

je bilo vanj vključenih pet spremenljivk iz vsej štirih kategorij. Iz enakih razlogov kot prej je bila vključena spremenljivka EBITDA. Le ta pa spada tudi v kategorijo dobičkonosnosti, iz katere imamo torej dve spremenljivki, saj se izkaže, da prav kazalniki dobičkonosnosti najbolj doprinašajo k izboljšanju prileganja modela. Zavedati pa se moramo, da smo modelirali podatke za leto 2013, ko sta ameriški trg zaznamovala visoka gospodarska rast in rastoč optimizem. Prav zato je izstopanje doprinosa kazalnikov dobičkonosnosti pričakovano, ne sme pa nas uporaba le ene skupine kazalnikov zavesti. Ključnega pomena so kazalniki vseh skupin, saj se nekateri izkažejo za primernejše v času gospodarske ekspanzije, spet drugi pa v času gospodarskih kontrakcij. Pri obeh modelih dodamo vpliv industrij šele na drugi ravni. Podjetja so torej v prvi ravni osnovne, ki so nato na drugi ravni razvrščene v skupine glede na industrijo, v kateri delujejo.

S kombinacijo SEM ter GLMM modela na ta način dobimo rezultate, prikazane v tabelama 9 ter 10. Vse spremenljivke so statistično značilne, v kolikor vzamemo $\alpha = 10\%$. Štiri izmed spremenljivk negativno vplivajo na verjetnost propada. Podjetje z višjim denarni tokom iz poslovanja do skupnega dolga ima namreč lahko višji denarni tok iz poslovanja ali pa nižji dolg kot primerjano podjetje ter je tako manj tvegano (seveda ob predpostavki, da je druga količina konstantna). Kazalnik ROS ter mera EBITDA sta kazalnika profitabilnosti, tako višja vrednost predstavlja višjo profitabilnost ter nižje tveganje propada. Tudi višja trenutna likvidnost zagotavlja

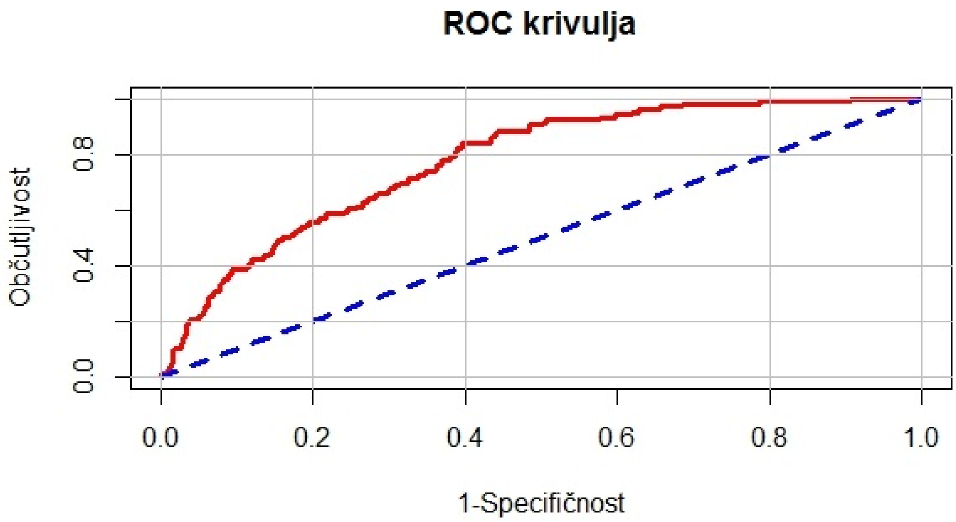
sprem =	koeficienti	standardna napaka	p vrednost
Denarni tok iz posl. do celotnega dolga	0.021	0.005	0.0651
EBITDA	0.289	0.072	0.000
ROS	0.826	0.185	0.000
Obračanje celotnih sredstev	-0.374	0.051	0.000
Trenutna likvidnost	0.199	0.056	0.000

Tabela 9: SEM model za konstrukcijo nove sprmenljivke.

	koeficienti	standardna napaka	p vrednost
Konstanta	-1.37961	0.09303	$< 2e - 16$
Sprem	-0.30294	0.07319	3.49e-05

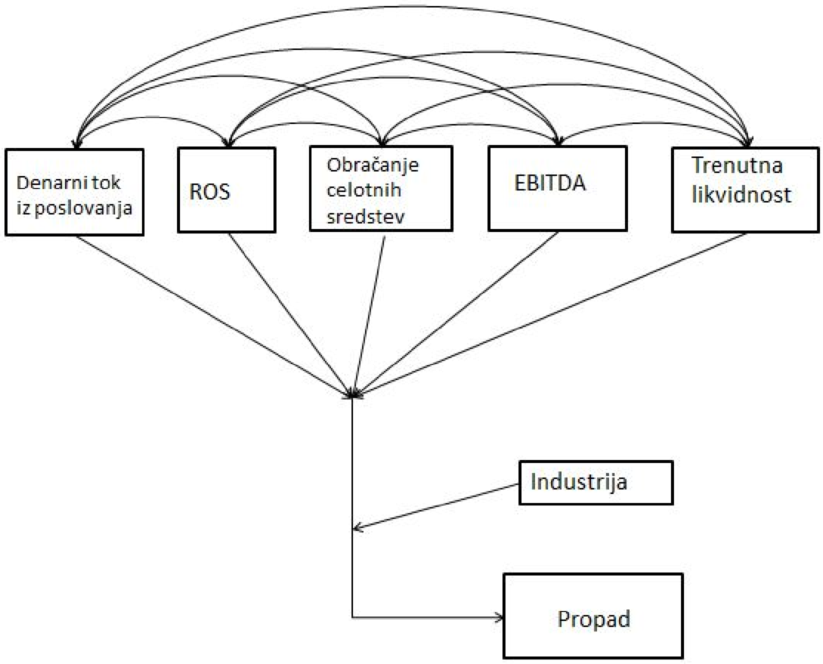
Tabela 10: GLMM model s probit vezno funkcijo.

lažje poplačevanje dolgov. Zato lahko rečemo, da so predznaki koeficientov skladni z ekonomsko teorijo. Na prvi pogled pa nas lahko preseneti pozitivna korelacija med verjetnostjo propada ter kazalnikom obračanja celotnih sredstev. Vendar se moramo zavedati, da je tudi ta kazalnik sestavljen iz dveh delov; prihodkov ter celotnih sredstev. Seveda višji dohodki prinašajo nižje tveganje, vendar nižja povprečna vrednost (kar pomeni višja vrednost kazalnika) prinaša večje tveganje. Najverjetneje je v absolutnem smislu v modelu premagala ta komponenta. Seveda bi bili ob tem primorani pogledati še koeficinete slučajnih učinkov glede na industrijo. Zaradi preglednosti pa le ti niso podani v nobenem izmed programskih jezikov. Slučajni učinki pa so upoštevani pri konstrukciji nove slučajne spremenljivke, ki podaja verjetnost propada. Ginijev koficient nam za dani model poda vrednost 0.4892, kar je hkrati



Slika 8: ROC krivulja za GLMM ter SEM model

ena najvišjih vrednosti, ki jo lahko z danimi podatki dosežemo. Prileganje modela je prikazano tudi s pomočjo ROC krivulje na Sliki 8. Vidimo, da je model občutno boljši od naključnega modela.



Slika 9: HSEM model

Pri modeliranju HSEM modela uporabimo analogen način, torej v prvi ravni obravnavamo podjetja kot posameze enote, v drugi pa kot del skupine glede na industrijo. To je prikazano tudi na Sliki 9. Bistvena razlika je torej zgolj v tem, da ne kreiramo nove spremenljivke, temveč modeliramo direktno. Žal nam program GLLAMM v Stati za izvajanje HSEM modela ne poda p vrednosti, saj v splošnem ne pozna porazdelitve slučajnih spremenljivk. V prejšnjem primeru smo namreč privzeli, da so spremenljivke normalno porazdeljene in smo na ta način izvedli SEM model za kreiranje latentne spremenljivke. Le ta je bila prav tako normalno porazdeljena ter nato vključena v GLMM model s probit vezno funkcijo. V splošnem pa HSEM model za spremenjivke ne zahteva točno določene porazdelitvene funkcije. Podano imamo torej matematično upanje ter disperijo slučajnega vzorca. V kolikor privzamemo normalnost porazdelitve slučajnih spremenljivk, lahko izračunamo testno T-statistiko za testiranje pričakovane vrednosti (μ_0) slučajnega vzorca $X_1, \ldots X_n$ s pomočjo formule:

$$T = \frac{\bar{X} - \mu_0}{\sqrt{\sigma^2/n}}.$$

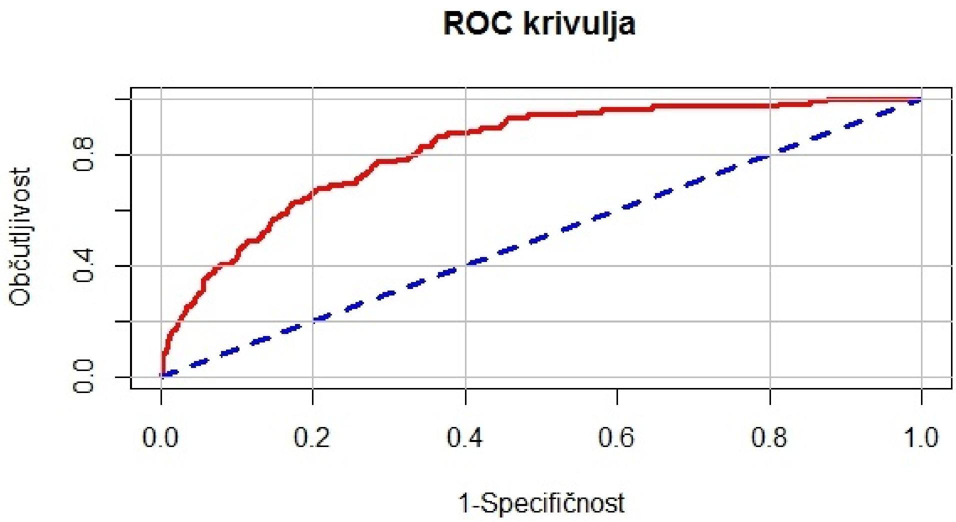
Vrednosti koeficientov ter standardnih napak HSEM modela so podane v Tabeli 11, poleg njih pa so dodane še izračunane vrenosti T-statistike za preverjanje domneve $\mu_0 = 0$, torej ali so koeficienti statistično različni od nič.

Kritična vrednost standardne normalne porazdelitve pri statistični značilnosti $\alpha = 0.01$ je 2.576. Vsi koeficienti so torej statistično značilni, saj so absolutne vrednosti vseh T-statistik krepko nad kritično vrednostjo. Prav tako so vse spremen-

	koeficienti	standardna napaka	T-statistika
Denarni tok iz posl. do celotnega dolga	-1.3275	1 .1791	-33.96
EBITDA	-0.65042	0.2367	-82.89
ROS	-0.0996	0.06228	-48.23
Obračanje celotnih sredstev	-0.0378	0.06077	-18.78
Trenutna likvidnost	-0.1635	0.09683	-50.93

Tabela 11: HSEM model.

ljivke negativno povezane z verjetnostjo propada podjetja. To velja tudi za kazalnik obračanja celotnih sredstev, ki je bil v prejšnjem modelu pozitivno povezan s spemenljivko propada. Tako kot prej pa je koeficient zelo blizu nič in statistično značilen. Sklepamo torej lahko, da v kolikor modeliramo neposredno s pomočjo HSEM modela prevlada, učinek višanja prihodkov nad učinkom nižanja celotnih sredstev. Zanima nas tudi, ali smo z novim modelom pridobili tudi na natančnosti prileganja. Izkaže se, da je nov Ginijev koeficient 0.5625 ter ROC krivulja podana na Sliki 10. Prileganje se je torej izrazito izboljšalo samo s tem, ko smo modeli-



Slika 10: ROC krivulja za HSEM model

rali neposredno (kar za 13%), v primerjavi s osnovnim probit modelom pa kar za 125%. To jasno nakazuje, da bi moral biti HSEM model preferiran pred kombinacijo GLMM ter SEM modela ter osnovnega probit modela. S spremeninjanjem vstopnih slučanjih spremenljivk lahko model oblikujemo po naših potrebah. Dane spremenljivke so bile izbrane iz vseh kategorij tako, da dosežemo dobro prileganje. Izkaže se tudi, da v kolikor uporabimo druge spremenljivke, lahko dosežemo boljše prileganje GLMM ter SEM modela, vendar pa Ginijev koeficient ne preseže vrednosti 0.495. Doprinos modeliranja obeh ravni skupaj (HSEM model) je v primerjavi z modelom, pri katerem sta ravni modelirani posebej (kombinacija GLMM ter SEM modela), se vedno prisoten, a gre le redko za zelo izrazita odstopanja. Model bi lahko še izboljšali, v kolikor bi vpeljali tudi časovno komponento. Na ta način bi lahko tvorili dinamičen model, ki ga zahteva tudi prihajajoči Basel III. Ena izmed možnosti za razširitev HSEM modela bi bila, da bi imeli v prvi enoti še vedno podjetja, razlikovana za posamezno leto, v drugem nivoju skupino podatkov za dano podjetje, ne glede na leto podatkov, v tretjem pa efekt industrij. Vse to pa presega namen te magistrske naloge, katere glavni cilj je bil pokazati, da je vpeljava HSEM modela smiselen korak, ki ob danih podatkih izboljša napovedovalno moč.

7 Zaključek

V tem magistrskem delu smo spoznali SEM model ter njegovo razširitev, HSEM model. Videli smo, da že po sami specifikaciji modela upoštevata koreliranost podatkov, s pomočjo latentnih spremenljivk pa še dodatno zagotavljata vključitev vseh direktnih ter indirektih povezav. Prav zato modela podata ekonomsko pričakovane predznake koeficientov ter boljšo statistično značilnost koeficientov. SEM model je bil tudi izjemno intuitiven, saj smo model lahko predstavili s pomočjo metode po poti. Tako lahko zahtevne matematične procese v ozadju izpustimo, konkreten model pa definiramo samo s pomočjo diagrama. Žal pa linearni strukturni model ne omogoča modeliranja podatkov, ki so razvrščeni v skupine. V našem primeru so bile te skupine industrije, v katerih podjetja delujejo. Zato smo izpeljali njegovo razširitev, hierarhični linearni strukturni model, ki je neke vrste hibrid med SEM ter GLMM modelom.

V nadaljevanju smo si njuno delovanje ogledali na primeru podatkov iz indeksa Russell 3000. Spoznali smo, da je uporaba SEM modela v ekonometiriji nujna, saj sta nam tako probit kot tudi logit model podala zavajujoče rezultate. Še več, ekonomsko neutemeljeni predznaki koeficientov so se pojavjali pri obeh modelih na istih mestih. Hkrati pa noben izmed modelov ni opozoril na spremenljivki, ki sta zaradi svoje koreliranosti povzročali multikolinearnost. Nato smo se osredotočili na SEM model za kreiranje latentne spremenljivke ter nato probit model. Opazili smo, da so težave z interpretacijo koeficientov izginile. Prav tako so bili vsi koeficienti statistično značilni. Pri taki konstrukciji modela pa smo izgubili skoraj vso napovedovalno moč; prilegal se je le nekoliko bolje od naivnega modela. Velik preobrat v prileganju (prileganje postane več kot tri krat boljše) pridobimo že samo z razvrstitvijo podjetij glede na industrijo, v kateri delujejo. Razliko v odstopanju od naivnega modela smo merili s pomočjo Ginijevega koeficienta ter ROC krivulje. Na ta način smo pokazali pomen vključenosti efekta industrij v našem modelu. Seveda je tak učinek pričakovan, saj je vrednost finančnih kazalnikov v veliki meri odvisna od industrije. Tu vidimo, da nam teorija kreditne analize lahko pomaga tudi kot vir smernic za izboljšanje prileganja modela. Še bolj natančno prileganje pa smo dobili, v kolikor smo vzeli kazalnike iz vseh osnovnih štirih skupin (kazalniki likvidnosti, dobičkonosnosti, zadolženosti in učinkovitosti) ter HSEM model. Pri HSEM modelu simuliramo v dveh nivojih; v prvem so osnovni objekti podjetja, ki so v drugem nivoju razvrščeni v skupine glede na industrijo, v kateri delujejo. Na ta način se je prileganje brez kakršnegakoli dodajanja podatkov povečalo za 13%. Ob tem smo ohranili enostavnost modela, saj ima še vedno linearno obliko (linearen je v spremenljivkah). Upoštevali smo torej osnovno dogmo statističnega modeliranja, s katero smo začeli. Našli smo relativno enostaven model, ki je glede na vhodne podatke specificiran pravilno.

Literatura

[1] A. Ansari in K. Jedidi, *Baysaian factor analysis for multilevel binary observation*, Psychometrika **65** (2000) 475–496.

[2] T. Asparouhov in B.O. Muhten, *Pseudo maximum likelihood estimation of mean- and covariance structures with missing data*, Jurnal of the American Statistical Association **85** (2004) 195-203.

[3] K.A., Bollen, *Structural Eqations with Latent Variabes*, Wiley, New York,1989.

[4] R. Bunemann in G. Williams, *Generalised Linear Mixed Models (GLMM)*, [25.3.2015], dostopno na `http://www.coloss.org/beebook/I/statistical-guidelines/5/2`.

[5] S. Blochwitz, T. Liebig in M. Nyberg, *Benchmarking Deutsche Bundesbank's default risk model, the KMV Private Firm Model and common financial ratios for German corporations. In Workshop on Applied Banking Research*, BASEL COMMITTEE ON BANKING SUPERVISION, 2000, [26.4.2015], dostopno na `http://www.bis.org/bcbs/oslo/liebigblo.pdf`.

[6] R.J. Carroll, L.A Stefanski in D. Ruppert, *Measurment error in Nonlinear Models*, Chapman&Hall, London, 1995.

[7] *Credit & Financial Risk analysis*, [ogled 31. 3. 2015], dostopno na `http://credfinrisk.com/index.html`.

[8] G. Damodran, *Basic econometrics*, international ed. **3rd**, McGraw-Hill, New York,1995.

[9] *Finančni slovar... vaš vstop v svet financ*, 2009-2011, [7.4.2015], dostopno na `http://www.financnislovar.com/index.html`.

[10] H. Goldenstein in R.P. McDonald, *A general model for the analysis of multilevel data* , Psyhometrika **53** (1988) 455-467.

[11] A. Gelman, *Scaling regression inputs by dividing by two standard deviations*, Statist. Med **27** (2008) 2865-2873.

[12] L. Hu in P.M. Bentler, *Fit indexes in covariance structure modeling: Sensitivity to underparameterized model misspecification.*, Psychological Methods **3(4)** (1999) 424-453.

[13] D. Hooper, J. Coughlan in M. Mullen, *Structural equation modelling: Guidelines for determining model fit.* , Articles **2** (2008) .

[14] D. Iacobucci, *Iacobucci, Dawn. Structural equations modeling: Fit indices, sample size, and advanced topics.*, Journal of Consumer Psychology **20** (2010) 90-98.

[15] K.G. Jörgeskog, *A general model for estimating a linear structural equation system*, Structural equation models in Social sciences (1973) 85-112. New York: Seminar .

[16] K.G. Jörgeskog, *Simultaneous factor analysis in several populations*, Psyhometrika **53** (1971) 409-426.

[17] K.G. Jörgeskog, *Latent variable scores and their uses*, on-line paper, [10.1.2015], dostopno na `http://www.ssicentral.com/lisrel/techdocs/lvscores.pdf`(2000).

[18] K.G. Jörgeskog, *Simultanious factor analysis in several populations*, Psyhometrika **36** (1971) 409-426.

[19] G. Kabir, I. Jahan, H. Chisty in M. A. A. Hasin, *Credit Risk Assessment and Evaluation System for Industrial Project* , International Journal of Trade, Economics and Finance **1.4** (2010) 331–341.

[20] D.A. Kenny, *Structural equation Modeing*, [20.3.2015], dostopno na `http://davidakenny.net/cm/causalm.htm`.

[21] M. Kovačič, *On-line slovarček statističnih pojmov*, [15.4.2015], dostopno na `http://www.ljudmila.org/matej/statistika/mva.html`.

[22] F. Klinker, *Generalized Linear Mixed Models for Ratemaking: A Means of Introducing Credibility into a Generalized Linear Model Setting* , [25.3.2015], dostopno na `https://www.casact.org/pubs/forum/11wforumpt2/Klinker.pdf`.

[23] R. Kuja-Halkola , *Hierarhical Linear Models and Structural Equation Modelling for the Children of Siblings model*, Uppsala Universitet, Project Report (2008).

[24] P.W Lei in Q. Wu, *Introduction to structural equation modeling: Issues and practical considerations.*, Educational Measurement: Issues and Practice **26.3** (2007)33-43.

[25] M. C. McCulloch, *Generalized linear mixed models*, Institute of Mathematical Statstics and the American Statistical Association, United States of America, 2003.

[26] J.N. Ory in P. Raimbourg , *Credit Rating Agencies' Function on Bond Markets: Price Stability Vs Information Transmission*, 21st Australasian Finance and Banking Conference. 2008.

[27] J. C. Pinherio in D. M. Bates, *Mixed effects Models in S and S-PLUS*, Springer, New York, 2000.

[28] *POJASNILO V ZVEZI Z UGOTAVLJANJEM PRIZNANE OBRESTNE MERE PRI POSOJILIH MED POVEZANIMI OSEBAMI, verzija 19. 11. 2007, [ogled 8. 4. 2015], dostopno na* `http://www.durs.gov.si/si/davki_predpisi_in_pojasnila/davek_od_dohodkov_pravnih_oseb_pojasnila/davcna_osnova/odhodki/pojasnilo_v_zvezi_z_ugotavljanjem_priznane_obrestne_mere_pri_posojilih_med_povezanimi_osebami`.

[29] S. Rabe-Hesketh, A. Skrondal, *Generalized Linear Latent and Mixed Models with Composite links and Exploded Likelihoods*, Statistical Modeling (2004) 27-39.

[30] S. Rabe-Hesketh, A. Skrondal in A. Pickles , *Generalized Multilevel structural equation modeling*, Psyhometrika **2** (2004) 167-190.

[31] S. Rabe-Hesketh, A. Skrondal in A. Pickles, *GLLAMM Manual*, **1**, Department of Biostatistics and Compiuting, Institute of Psychiatry, King's Collage, University of London, dostopno na `http://www.gllamm.org`.

[32] S. Rabe-Hesketh, A. Skrondal, *Latent Variable Modeling: A Survery*, Scandinavian Jurnal of Statistics **34.4** (2007) 715-745.

[33] S. Rabe-Hesketh, A. Skrondal in A. Pickles, *Maximum likelihood estimation of limited and discrete dependent variable models with nested random effects.*, Jurnal of Econometrics **128** (2005) 301-323.

[34] S. Rabe-Hesketh, A. Skrondal, *Prediction in multilevel generalized linear models*, Journal of the Royal Statistical Society **172** (2009) 659-687.

[35] *Ready ratios*, [30.4.2015], dostopno na `http://www.readyratios.com/`.

[36] T. Van Gestel in B. Baesens, *Credit Risk Managment: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital: Basic Concepts: Financial Risk Components, Rating Analysis, Models, Economic and Regulatory Capital.*, Oxford University Press, New York, 2009.

[37] M.H. Zweig in G Campbell, *Receiver-operating characteristic (ROC) plots: a fundamental evaluation tool in clinical medicine*, Clinical Chemistry **39** (1993) 561-577.