Faktorska analiza u kreditnom skoriranju

Ivanec, Lucija

Sveučilište Josipa Jurja Strossmayera u Osijeku
Fakultet primijenjene matematike i informatike
Sveučilišni diplomski studij matematike
Smjer: Financijska matematika i statistika

Lucija Ivanec

Faktorska analiza u kreditnom skoriranju

Diplomski rad

Osijek, 2023.

Sveučilište Josipa Jurja Strossmayera u Osijeku
Fakultet primijenjene matematike i informatike
Sveučilišni diplomski studij matematike
Smjer: Financijska matematika i statistika

Lucija Ivanec

Faktorska analiza u kreditnom skoriranju

Diplomski rad

Mentorice: prof. dr. sc. Mirta Benšić
prof. dr. sc. Nataša Šarlija

Osijek, 2023.

Sadržaj

1 Uvod

U ovom diplomskom radu objašnjava se faktorska analiza, jedna od poznatih statističkih metoda, te se primjenjuje u kreditnom skoriranju. Rad se sastoji od dva dijela. U prvom dijelu rada upoznaje se faktorska analiza, njena primjena i pretpostavke, pojmovi kreditna sposobnost i skoriranje. Drugi dio rada donosi empirijski dio, sve opisano u prvom dijelu, pokazuje se na reprezentativnom uzorku od 24 varijable i 6414 podataka malih i srednjih poduzeća u Hrvatskoj koji obuhvaća podatke financijskih izvješća i podatke o insolventnosti.

Faktorska analiza česta je multivarijatna tehnika čiji je cilj identifikacija i razumijevanje zajedničkih karakteristika više varijabli te smanjenje broja varijabli u analizi ako se radi o velikom uzorku. Ona velik broj varijabla reducira u manji broj varijabla kojima je lakše i jednostavnije upravljati. Taj manji broj dobivenih varijabli nazivamo faktorima te pomoću nekoliko faktora lakše modeliramo cio velik uzorak. Faktorska analiza ima široku primjenu u obrazovanju, bankarstvu, marketingu, računovodstvu.

Kod odobravanja kredita banke se susreću s teškom odlukom, kome odobriti, a kome ne odobriti kredit. Također, znamo da je kreditiranje glavni prihod većini banaka. Procesu odobrenja kredita treba pristupiti pozorno te osmisliti i implementirati učinkovito upravljanje kreditnim rizikom. Pri procjeni rizika banka može koristiti kreditno skoriranje, proces procjene kreditne sposobnosti klijenta. Faktoska analiza će nam pomoći u identificiranju ključnih varijabli, a model skoriranja ćemo napraviti korištenjem logističke regresije.

2 Faktorska analiza

2.1 O faktorskoj analizi

Faktorska analiza definira se kao skup matematičko-statističkih postupaka čija je zadaća da se velik broj početnih varijabla opiše manjim brojem varijabli. Na primjer, 20 početnih varijabli svede se na nove tri i te tri u potpunosti opisuju početnih 20 varijabli. Na taj se način uvelike olakšava istraživanje s manjim brojem varijabli. Nove, izvedene varijable, nazivaju se faktori iz čega i slijedi naziv faktorska analiza.

Začetnikom faktorske analize smatra se engleski psiholog Charles Edward Spearman koji je 1904. godine postavio dvofaktorsku teoriju inteligencije te za provjeru svoga empirijskog koncepta razvio metodu faktorske analize. Tek kasnije faktorska analiza našla je široku primjenu u medicini, biologiji i drugim prirodnim znanostima.

Postoje dvije vrste faktorske analize, eksplorativna (EFA) i konfirmatorna faktorska analiza (CFA). EFA ima dulju povijest, koja datira još iz Spearmanove ere, dok je CFA postala popularnija nakon proboja u računalnoj tehnologiji i metodi procjene. EFA utvrđuje faktore na temelju dobivenih podataka te ih koristi u trenutku kada faktori još nisu definirani, a CFA testira poznatu ili već postavljenu teoretsku strukturu faktora.

Postoji više razloga zašto koristiti upravo faktorsku analizu, no najvažniji je da ona omogućuje sažimanje većeg broja povezanih varijabli u manji broj tako da im objasni povezanost i utvrdi zajedničke značajke. Faktorskom analizom ne brišu se izvorne varijable, već se izvorne varijable ponovno kombiniraju u svrhu pronalaska zajedničkog faktora koji pomaže pri razumijevanju problema.

2.2 Primjena

Faktorska analiza ima široku primjenu u raznim istraživanjima. Navodimo primjer istraživanja N. K. Živadinović [9]. U svom radu napravila je istraživanje osnovnih obilježja proizvoda kava primjenjujući faktorsku analizu. Najprije je 50 studenata napisalo nekoliko pojmova vezano uz proizvod kava. Zatim je provedeno pilot istraživanje na 155 studenata kako bi se utvrdila pouzdanost upitnika. Nakon analize tih upitnika, sastavljena je konačna lista koju je činilo 15 izvornih varijabla, odnosno mišljenja, stavova i izjava o karakteristikama kave. Potom je novih 248 studenata popunjavalo upitnik. Zadatak ispitanika bio je označiti slažu li se s tom karakteristikom ili ne označavajući ih brojem od 1 do 7, gdje je 1 označavalo 'uopće se ne slažem', a 7 'u potpunosti se slažem'. Faktorskom analizom izlučena su 4 faktora koja su opisana početnim varijablama. Na primjer jedan od faktora nazvan je koncentracija, a bio je opisan sljedećim varijablama: kada popijem kavu, ne spava mi se; kavu pijem kad želim biti koncentriran/a na učenje i kava je važna za razbuđivanje. Dakle, autorica je velik broj obilježja svela na njih 4 primjenjujući faktorsku analizu.

2.3 Povezanost među varijablama

Prije izvođenja same faktorske analize potrebno je analizirati povezanost među varijablama i provjeriti jesu li uopće varijable primjerene za faktorsku analizu. Jedan od kriterija je Kaiser-Meyer Olkinova mjera ili kraće KMO (*eng. Measure of sampling adequacy - MSA*). Ona mjeri jesu li podaci prikladni za provođenje faktorske analize. Može se računati za svaku

varijablu posebno ili za sve varijable zajedno.
KMO vrijednost računamo sljedećom formulom [12, strana 445]:

$$KMO = \frac{\sum_{i \neq j} r_{ij}^2}{\sum_{i \neq j} r_{ij}^2 + \sum_{i \neq j} q_{ij}^2}, 0 \leq KMO \leq 1,$$

gdje je $r_{ij}^2 (i \neq j)$ kvadrat koeficijenta korelacije između i-te i j-te varijable, a q_{ij}^2 kvadrat koeficijenta parcijalne korelacije između i-te i j-te varijable.
Sve KMO vrijednosti nalaze se između 0 i 1. Vrijednosti iznad 0.9 smatramo izvrsnima, vrijednosti između 0.6 i 0.5 smatramo prihvatljivima i vrijednosti manje od 0.5 neprihvatljivima. Nakon što napravimo KMO vrijednosti za svaku varijablu posebno ili sve varijable, odmah uočavamo jesu li one ili nisu prikladne za provođenje faktorske analize.

Osim KMO testa prikladnost za faktorsku analizu varijabli provjerava se i Bartlettovim testom. Bartlettov test testira nul hipotezu: *ne postoji značajna korelacija između izvornih varijabli.* Test se temelji na matrici koeficijenata korelacije. Matrica koeficijenata njegove nul hipoteze je jedinična matrica. Kod jedinične matrice svi dijagonalni elementi su jedan, a svi nedijagonalni elementi iznose nula. Dakle, ako je korelacijska matrica slična jediničnoj matrici, varijable su nekorelirane i nema smisla provoditi faktorsku analizu. Mala p-vrijednost (manje od 0.05) ukazuje da faktorska analiza može biti korisna s podacima.

Faktori nisu zadani, već se procjenjuju iz podataka te se objašnjavaju i interpretiraju. Nakon što se procijene, varijable bi se trebale raščlaniti po grupama sa zajedničkim faktorima.

3 Model faktorske analize

Pogledajmo sada kako izgleda model faktorske analize. Promatramo varijable x_i, gdje $i = 1, \ldots, p$. Cilj je prikazati ih kao linearnu kombinaciju zajedničkih faktora koje označujemo sa $f_1, f_2, ..., f_k$ i slučajne greške ε.
Dolazimo do osnovnog faktorskog modela:

$$\boldsymbol{x} - \boldsymbol{\mu} = \boldsymbol{L}\boldsymbol{f} + \boldsymbol{\varepsilon} \tag{1}$$

u kojem je $\boldsymbol{x}$ vektor s očekivanjem $\mu = (\mu_1, \mu_2, \ldots, \mu_p)$ i matricom kovarijance Σ. Postoji p varijabli i od njih želimo napraviti k faktora. $\boldsymbol{L}$ je matrica dimenzije $p \times k$, a $l_{i1}, l_{i2}, \ldots, l_{p5}$ *težine zajedničkih faktora* f_j, $j = 1, \ldots, k$. $\boldsymbol{\varepsilon} = (\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_p)^T$ naziva se *vektor grešaka*, koje se još nazivaju i specifični faktori.

Broj specifičnih faktora jednak je početnom broju zadanih varijabli. Svaki od specifičnih faktora vezan je uz točno jednu početnu varijablu i objašnjava onaj dio varijabli koji nije objašnjen navedenim faktorima. Najidealniji slučaj bio bi kad bi početne varijable bile sasvim objašnjene zajedničkim faktorima, ali najčešće uz njih dodajemo i specifične faktore.

Osim matrično, faktore možemo zapisati i sljedeći način:

$$
\begin{aligned}
x_1 &= \mu_1 + l_{11}f_1 + \ldots + l_{1k}f_k + \varepsilon_1 \\
x_2 &= \mu_2 + l_{21}f_1 + \ldots + l_{2k}f_k + \varepsilon_2 \\
&\vdots \\
x_p &= \mu_p + l_{p1}f_1 + \ldots + l_{pk}f_k + \varepsilon_p
\end{aligned}
\tag{2}
$$

Uočava se, l_{i1} je težina i-te varijable u odnosu na prvi faktor. Koeficijenti l_{ij} prikazuju važnost $j-tog$ faktora f_j za i-tu varijablu x_i. Koeficijenti l_{ij} mogu se koristiti u interpretaciji f_j. Težine faktora važne su jer nam govore koliki je utjecaj nekog faktora na određenu varijablu.

Pretpostavke modela

Pogledajmo i pobliže objasnimo pretpostavke navedenog modela [10]:

$$
\begin{gathered}
E(f) = 0,\ E(\boldsymbol{\varepsilon}) = 0, \\
var(f_j) = 1,\ var(\varepsilon_j) = \psi_j \\
cov(\boldsymbol{\varepsilon}) = \Psi,\ cov(\boldsymbol{f}, \boldsymbol{\varepsilon}) = \mathbf{0}, \\
cov(\boldsymbol{f}) = \boldsymbol{I}.
\end{gathered}
$$

$\boldsymbol{\Psi}$ je dijagonalna matrica, $\mathbf{0}$ matrica nula i $\boldsymbol{I}$ jedinična matrica.
Iz pretpostavke $cov(\boldsymbol{\varepsilon}) = \Psi$ slijedi nekoleriranost grešaka, što je osnovna pretpostavka faktorskog modela.

Pretpostavljamo da su greške i zajednički faktori nekolerirani, vrijedi $E[\boldsymbol{f}\boldsymbol{\varepsilon}^T] = 0$. $cov(\boldsymbol{f}) = \boldsymbol{I}$ govori nam da su zajednički faktori međusobno nekorelirani, no ta pretpostavka nije nužna.

Uočimo,

$$
\begin{aligned}
cov(\boldsymbol{x}, \boldsymbol{f}) =& E(\boldsymbol{x} - E(\boldsymbol{x}))(\boldsymbol{f} - E(\boldsymbol{f}))^T) = E((\boldsymbol{L}\boldsymbol{f} + \boldsymbol{\varepsilon})\boldsymbol{f}^T) = \\
&= E(\boldsymbol{L}\boldsymbol{f}\boldsymbol{f}^T + \boldsymbol{\varepsilon}\boldsymbol{f}^T) = \boldsymbol{L}E(\boldsymbol{f}\boldsymbol{f}^T) + E(\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^T) = \boldsymbol{L} \cdot \boldsymbol{I} + 0 = \boldsymbol{L}
\end{aligned}
$$

to jest $cov(x_i, f_j) = l_{ij}$, što implicira da je kovarijanca između slučajnog vektora $\boldsymbol{x}$ i vektora zajedničkih faktora $\boldsymbol{f}$ u potpunosti određena matricom $\boldsymbol{L}$.

3.1 Varijanca i komunalitet

Sada možemo zapisati kovarijancu od x_i za $i \neq j$

$$
\begin{aligned}
\sigma_{ij} &= cov(x_i, x_j) = E(x_i - E(x_i))(x_j - E(x_j)) = \\
&= E(\mu_i + l_{i1}f_1 + \ldots + l_{1k}f_k + \varepsilon_i - \mu_i)(\mu_j + l_{j1}f_1 + \ldots + l_{jk}f_k + \varepsilon_j - \mu_j) = \\
&= l_{i1}l_{j1}E(f_1)^2 + \ldots + l_{i1}l_{jk}E(f_1f_k) + l_{1k}l_{j1}E(f_kf_1) + \ldots + l_{1k}l_{jk}E(f_k^2) + E(\varepsilon_i\varepsilon_j) = \\
&= l_{i1}l_{j1} + l_{i2}l_{j2} + \ldots + l_{ik}l_{jk} = l_i^T l_j.
\end{aligned}
$$

Za varijancu x_i vrijedi:

$$
\begin{aligned}
\sigma_{ii} &= var(x_i) = E(x_i - E(X_i))(x_i - E(x_i)) = \\
&= E(\mu_i + l_{i1}f_1 + \ldots + l_{ik}f_k + \varepsilon_i - \mu_i)(\mu_i + l_{i1}f_1 + \ldots + l_{ik}f_k + \varepsilon_i - \mu_i) = \\
&= l_{i1}^2E(f_1)^2 + l_{i2}^2E(f_2)^2 \ldots + l_{ik}E(f_2)^2 + E(\varepsilon_i)^2 = \\
&= l_{i1}^2 + l_{i2}^2 + \ldots + l_{ik}^2 + \psi_i \\
&= l_i^T l_i + \psi_i = h_i^2 + \psi_i,
\end{aligned}
$$

gdje je $h_i^2 = l_i l_i^T$. Uočava se da se varijanca može podijeliti na dva dijela, prvi dio

$$h_i^2 = \sum_{i=1}^{k} l_{ij}^2$$

koji nazivamo zajednička varijanca ili komunalitet i drugi dio ψ_i koji se zove specifična ili jedinstvena varijanca. Komunalitet predstavlja varijancu od x_i koja se dijeli s drugim varijablama preko zajedničkog faktora, a jedinstvena varijanca objašnjava varijabilnost koju x_i ne dijeli s drugim varijablama.

Nadalje, l_{i1}^2 je doprinos prvog zajedničkog faktora zajedničke varijance, l_{i2}^2 je doprinos drugog zajedničkog faktora zajedničke varijance i tako dalje.

Matricu kovarijanci Σ možemo izraziti pomoći k-faktorskog modela na sljedeći način:

$$\begin{aligned}
\Sigma = E(\boldsymbol{x}-\boldsymbol{\mu})(\boldsymbol{x}-\boldsymbol{\mu})^T &= E(\boldsymbol{Lf}+\boldsymbol{\varepsilon})(\boldsymbol{Lf}+\boldsymbol{\varepsilon})^T \\
&= E(\boldsymbol{Lff^TL^T} + \boldsymbol{Lf\varepsilon^T} + \boldsymbol{\varepsilon^T f^T} + \boldsymbol{\varepsilon\varepsilon^T}) \\
&= \boldsymbol{L}E(\boldsymbol{ff^T})\boldsymbol{L^T} + \boldsymbol{L}E(\boldsymbol{f^T\varepsilon^T}) + E(\boldsymbol{\varepsilon f^T})\boldsymbol{L^T} + E(\boldsymbol{\varepsilon\varepsilon^T}) \\
&= \boldsymbol{LIL^T} + E(\boldsymbol{\varepsilon\varepsilon^T}) \\
&= \boldsymbol{LL^T} + \boldsymbol{\Psi}, \qquad (3)
\end{aligned}$$

gdje je $E(\boldsymbol{\varepsilon\varepsilon^T}) = \boldsymbol{\Psi}$ po pretpostavci modela dijagonalna matrica s dijagonalnim elementima $(\psi_1, \ldots, \psi_p)$.

Pogledajmo model faktorske analize za neke p i k, neka je $p = 3$ i $k = 2$. Prema (2) dobiva se sljedeći zapis:

$$\begin{aligned}
x_1 &= \mu_1 + l_{11}f_1 + l_{12}f_2 + \varepsilon_1 \qquad (4) \\
x_2 &= \mu_2 + l_{21}f_1 + l_{22}f_2 + \varepsilon_2 \\
x_3 &= \mu_p + l_{31}f_1 + l_{32}f_2 + \varepsilon_3
\end{aligned}$$

U matričnom obliku dobiva se:

$$\begin{bmatrix} x_1 \\ x_2 \\ x_3 \end{bmatrix} = \begin{bmatrix} \mu_1 \\ \mu_2 \\ \mu_3 \end{bmatrix} + \begin{bmatrix} l_{12} & l_{12} \\ l_{21} & l_{22} \\ l_{31} & l_{32} \end{bmatrix} \begin{bmatrix} f_1 \\ f_2 \end{bmatrix}^T + \begin{bmatrix} \varepsilon_1 \\ \varepsilon_2 \\ \varepsilon_3 \end{bmatrix}$$

Kako $\boldsymbol{L}$ ima nekoliko stupaca, tada $\Sigma = \boldsymbol{LL^T} + \boldsymbol{\Psi}$ predstavlja pojednostavljenu strukturu matrice kovarijanci u kojoj su kovarijance modelirane pomoću l_{ij} jer je $\boldsymbol{\Psi}$ dijagonalna matrica. U našem primjerku $k = 2$, pa je kovarijanca jednaka produktu prva dva reda matrice koeficijenata $\boldsymbol{L}$, odnosno:

$$cov(x_1, x_2) = l_{12}l_{21} + l_{12}l_{22}$$

Ako x_1 i x_2 sadrže zajednički faktor, tada će imati i zajedničke koeficijente uz f_1 i f_2. $l_{11}l_{21}$ i $l_{12}l_{22}$ imat će visoku vrijednost. Također, ako x_1 i x_2 imaju malo toga zajedničkog, koeficijenti neće biti slični, a $l_{11}l_{21}$ i $l_{12}l_{22}$ će imati male vrijednosti.

Nadalje, mogu se pronaći kovarijance između x_i i faktora f_j preko l_{ij}, prije toga uočimo

kako je $x_1 - \mu_1 = l_{11}f_1 + l_{12}f_2 + \ldots + l_{1k}f_k + \varepsilon_1$. Iz $cov(\boldsymbol{f}) = \boldsymbol{I}$ f_2 je nekoreliran s ostalim f_j, a iz $cov(\boldsymbol{f}, \boldsymbol{\varepsilon}) = \boldsymbol{0}$ f_2 je nekoreliran s ε_1 i $var(f_2) = 1$.
Pogledajmo,

$$\begin{aligned}
cov(x_1, f_2) &= E[(x_1 - \mu_1)(f_2 - \mu f_2)] \\
&= E[(l_{11}f_1 + l_{12}f_2 + \ldots + l_{1k}f_k)f_2] \\
&= E[l_{11}f_1f_2 + l_{12}f_2^2 + \ldots + l_{1k}f_kf_2] \\
&= l_{11}cov(f_1, f_2) + l_{12}var(f_2) + \ldots + l_{1k}cov(f_k, f_2) \\
&= l_{12}
\end{aligned} \tag{5}$$

Iz prethodnog se vidi kako su težine jednake kovarijanci faktora i varijabli, vrijedi:

$$cov(x_i, f_j) = l_{ij}, i = 1, \ldots, p, j = 1, \ldots, k$$

Kako je l_{ij} $(ij) - ti$ element $\boldsymbol{L}$ možemo zapisati:

$$cov(\boldsymbol{x}, \boldsymbol{f}) = \boldsymbol{L}.$$

3.2 Invarijantnost mjerila

Ponovno skaliranje varijabli $\boldsymbol{x}$ ekvivalentno je $\boldsymbol{y} = \boldsymbol{CX}$, gdje je $\boldsymbol{C} = diag(c_i)$. Ako za k-faktorski model vrijedi $\boldsymbol{L} = \boldsymbol{L_x}$ i $\boldsymbol{\Psi} = \boldsymbol{\Psi_x}$ tada:

$$\boldsymbol{y} = \boldsymbol{CL_xf} + \boldsymbol{C\varepsilon} + \boldsymbol{C\mu}$$

i

$$Var(\boldsymbol{y}) = \boldsymbol{C\Sigma C} = \boldsymbol{CL_xL_x^TC} + \boldsymbol{C\Psi_xC}.$$

Tako k-faktorski model vrijedi za $\boldsymbol{y}$ s faktorima matrice $\boldsymbol{L_y} = \boldsymbol{CL_x}$ i specifične varijance $\boldsymbol{\Psi_y} = \boldsymbol{C\Psi_xC} = diag(c_i^2\psi_{ii})$. Primjećuje se da se faktorska matrica skalirane varijable $\boldsymbol{y}$ dobiva skaliranjem matrice težina izvornih varijabli (množimo $i - ti$ red od $\boldsymbol{L_x}$ sa c_i), a slično slijedi i za specifične varijable. Može se zaključiti da na faktorsku analizu ne utječe ponovno skaliranje varijabli.

3.3 Rotacije

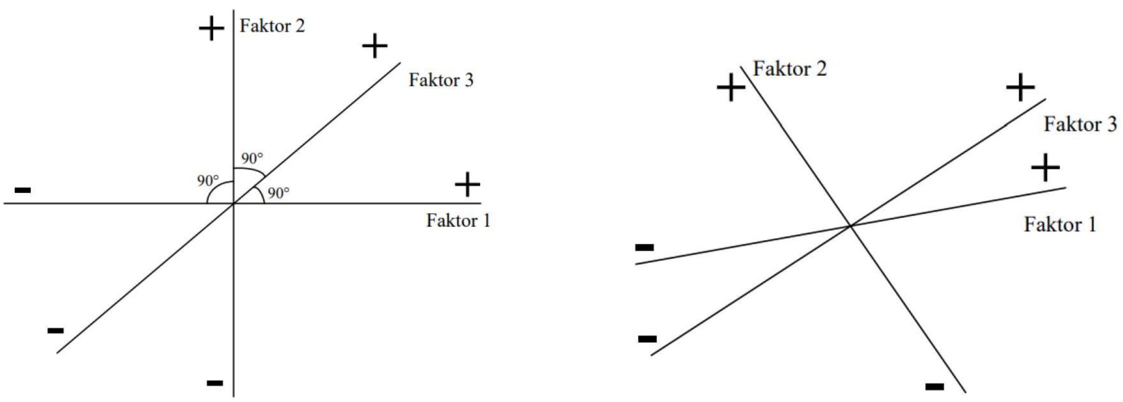
Kada se faktori ne mogu lako interpretirati, uobičajeno je transformirati ih rotacijom. Rotaciju faktora koristi se radi veće razumljivosti odnosno bolje interpretabilnosti podatka i to bez mijenjanja matematičkih svojstava. [12, stranica 373] Faktori se rotiraju u višedimenzionalnom prostoru tako da se dovedu u položaj jednostavne strukture.Jednostavna se struktura dobije ako je svaka varijabla reprezentirana sa što manje faktora. Ako su, na primjer, 3 faktora, provode se tri rotacije: prva i druga os se zarotiraju, zatim prva i treća i onda druga i treća. Žei se dobiti jednostavna struktura. Ako se ne dobije odmah, provodi se novi ciklus rotacija. [8, strana 158] Postoje dvije vrste rotacija, ortagonalna i kosa.

3.3.1 Ortagonalna rotacija

Ortagonalna rotacija koristi se kada faktori nisu povezani jedni s drugima, odnosno nisu ni u kakvoj korelaciji. U grafičkoj reprezentaciji faktora koordinatne osi predstavljaju faktore, a dimenziju određuje broj faktora. Ortagonalne rotacije su rotacije osi za 90°. Za vrijeme rotacije faktorske osi ostaju u nepromijenjenom položaju. Kut između faktorskih osi prije i poslije rotacije ostaje nepromijenjen, a postotak ukupno objašnjene varijance prije i poslije rotacije isti je. [12, strana 431] Tako da prvi faktor i dalje objašnjava najveći dio varijance, drugi manje i tako dalje. Jednako kao i prije rotacije.

3.3.2 Kosa rotacija

Ponekad nakon ortagonalne rotacije i dalje nema jasnije slike o faktorima te se tada koristi kosa rotacija. Kod kose rotacije među faktorima javlja se zavisnost, a osi ne ostaju okomito kao kod ortagonalne rotacije. [8, strana 168.] Kut između faktorskih osi može postati veći ili manji od 90○. Što je kut između faktorskih osi manji, to ukazuje na veću korelaciju među faktorima. Udaljenosti i kutovi nisu sačuvani tako da se komunaliteti prije i nakon rotacije razlikuju. [12, strana 435]
Na Slici 2 primjer je ortagonalne i kose rotacije s tri vektora, odnosno dimenzije tri [12]. Vidi se kako su osi kod ortagonalne rotacije ostale pod pravim kutom dok kod kose rotacije to nije tako.



Slika 1: Ortagonalna (lijevo) i kosa (desno) rotacija

3.4 Uvjet jedinstvenosti

Faktorski model nije jedinstven, dva različita para $(\boldsymbol{L}, \boldsymbol{f})$ i $(\boldsymbol{L}^*, \boldsymbol{f}^*)$ u rezultatu mogu dati istu matricu kovarijance $\boldsymbol{\Sigma}$. Zanima nas pod kojim uvjetima, ako postoji, se može naći dekompozicija matrice kovarijanci kao u jednadžbi (3).
Pretpostavlja se da takva dekompozicija postoji, no treba se naći pod kojim se uvjetima dolazi do jedinstvenih faktorskih koeficijenata l_{ij}. Pogledajmo kako bi odredili L i f.

Pretpostavimo neka je $\boldsymbol{\Gamma}$ otragonalna matrica, $\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T = \boldsymbol{I}$, tada osnovni faktorski model (1) možemo zapisati:

$$\begin{aligned} \boldsymbol{x} &= \boldsymbol{\mu} + \boldsymbol{L}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T\boldsymbol{f} + \boldsymbol{\varepsilon} = \\ &= \boldsymbol{\mu} + \boldsymbol{L}^*\boldsymbol{f}^* + \boldsymbol{\varepsilon} \end{aligned} \tag{6}$$

gdje je $\boldsymbol{L}^* = \boldsymbol{L}\boldsymbol{\Gamma}$ i $\boldsymbol{f}^* = \boldsymbol{\Gamma}^T\boldsymbol{f}$.

Primijetimo,

$$E(\boldsymbol{f}^*) = E(\boldsymbol{\Gamma}^T\boldsymbol{f}) = \boldsymbol{\Gamma}^T E(\boldsymbol{f}) = \boldsymbol{0}$$

$$cov(\boldsymbol{f^*}) = E(\boldsymbol{f^*} - E(\boldsymbol{f^*}))(\boldsymbol{\varepsilon} - E(\boldsymbol{\varepsilon}))^T = E(\boldsymbol{f^*})E(\boldsymbol{\varepsilon})^T = \boldsymbol{0}.$$

Vidimo da je novi k-faktorski model valjan s novim faktorima $\boldsymbol{\Gamma}^T\boldsymbol{f}$ i faktorskim težinama $\boldsymbol{L}\boldsymbol{\Gamma}$. Stoga, svaka transformacija od $\boldsymbol{f}$ dati će sličnu strukturu matrice kovarijanci $\boldsymbol{\Sigma}$. Ako se zamijeni $\boldsymbol{L}$ s $\boldsymbol{L}^* = \boldsymbol{L}\boldsymbol{\Gamma}$ imamo:

$$\begin{aligned} \boldsymbol{\Sigma} = \boldsymbol{L}\boldsymbol{L}^T + \boldsymbol{\Psi} &= \boldsymbol{L}\boldsymbol{\Gamma}\boldsymbol{\Gamma}^T\boldsymbol{L}^T + \boldsymbol{\Psi} \\ &= \boldsymbol{L}^*\boldsymbol{L}^{*T} + \boldsymbol{\Psi} \end{aligned} \tag{7}$$

Obje faktorizacije (3) i (7) imaju jednake greške $\boldsymbol{\varepsilon}$, te su rješenja jednako dobra.
Ako se iskoristi supstitucija $\boldsymbol{L}^* = \boldsymbol{L}\boldsymbol{C}$ i $\boldsymbol{f}^* = \boldsymbol{C}^{-1}\boldsymbol{f}(\boldsymbol{C}^T)^{-1}$ vidi se da i ona zadovoljava faktorski model (3). Dakle, s novim težinama, $\boldsymbol{L}^*$, dobiva se ista matrica kovarijance $\boldsymbol{\Sigma}$ kao i sa $\boldsymbol{L}$. Dolazi do problema nejedinstvenosti rješenja što onemogućava izračun jedinstvenog modela. Kako bi došli do jedinstvenog rješenja uvodi se restrikciju koja zadovoljava (3).
Primjerice, neka je matrica $\boldsymbol{L}^T\boldsymbol{\Psi}^{-1}\boldsymbol{L}$ dijagonalna ili $\boldsymbol{L}^T\boldsymbol{D}^{-1}\boldsymbol{L}$ dijagonalna, $\boldsymbol{D} = diag(\sigma_1, \ldots, \sigma_p)$.
Broj parametara u (3) bez ograničenja:

$$\begin{gathered} \text{matrica } \boldsymbol{L}(p \times k) \text{ ima } p \cdot k \text{ parametara} \\ \text{matrica } \boldsymbol{\Psi}(p \times p) \text{ ima } p \text{ parametara.} \end{gathered}$$

Stoga, ima $pk + p$ slobodnih parametara koji se umanjuju za broj nedijagonalnih elemenata matrice. Konačno, broj parametara jednak je $\frac{1}{2}k(k-1)$. Označimo sa s broj stupnjeva slobode faktorskog modela koji je jednak broju neograničenih elemenata u $\boldsymbol{\Sigma}$ umanjeni za broj slobodnih parametara u $\boldsymbol{\Sigma}$:

$$s = \frac{1}{2}p(p+1) - \{pk + p - \frac{1}{2}k(k-1)\} == \frac{1}{2}(p-k)^2 - \frac{1}{2}(p+k).$$

Kada je $s < 0$ model je neodređen i ima previše faktora, što znači da je broj faktora veći od broja početnih varijabli, postoji beskonačno mnogo rješenja. Jedinstven model dobiva se za $s = 0$, ako se ne koriste rotacije. U slučaju $s > 0$ postoji više jednadžbi nego parametara, a egzaktno rješenje ne postoji, uzimamo aproksimativna rješenja. [6] Nadalje, pogledajmo kako transformacija utječe ne komunalitet.

$$h_i^{2*} = l_i^{T*} l_i^* = l_i^T \mathbf{\Gamma}\mathbf{\Gamma}^T l_i = l_i^T l_i = h_i^2$$

Vidi se kako h_i^2 ostaje isti budući da je $\boldsymbol{L}\boldsymbol{L}^T = \boldsymbol{L}^*\boldsymbol{L}^{*^T}$. Stoga komunalitet i općenito faktorizacija Σ ostaju invarijantne u odnosu na otragonalnu transformaciju te su težine određene samo do ortagonalne transformacije matrice $\mathbf{\Gamma}$. Kako je svaka ortagonalna transformacija ekvivalentna rotaciji, vidi se da rotacije faktora ne utječu na zajedničke varijancu. Ni uz ograničenja faktorski model nije jedinstveno određen zbog mogućih rotacija koordinatnih sustava, neodređenost se obično rješava tako da se fiksira koordinatni sustav.

3.5 Metode procjene

Postoji više vrsta metoda za procjenu parametara faktorskog modela. Osnovna podjela jest na metodu glavnih komponenata i metodu maksimalne vjerodostojnosti. U ovom radu koristit će se metoda glavnih komponenata.

Metoda glavnih komponenata

1901. godine Karl Pearson prvi je opisao metodu glavnih komponenata (eng. *Principal component analysis - PCA*), a Harold Hotelling je 1933. opisao izračun. Metoda glavnih komponenata najjednostavnija je metoda u faktorskoj analizi.

Ime je dobila jer je prva glavna komponenta linearna kombinacija promatranih varijabli koja uzima najviše varijabilnosti iz podataka, a druga glavna komponenta ponovno je linearna kombinacija iz promatranih varijabli koja uzima drugi najveći dio varijabilnosti iz podataka i ortogonalna je s prvom i tako dalje. Najprije se računa prva glavna komponenta, zatim druga i tako redom. Na kraju dobivamo skup glavnih komponenata koje su poredane tako da prva ima najveći dio varijabilnosti, a zadnja najmanji dio varijabilnosti. Cilj je ove metode izvući najveću varijancu iz skupa podataka. Pogledajmo kako to izgleda [6].
Metoda glavnih komponenata započinje aproksimacijom matrice težina $\hat{\boldsymbol{L}}$, pri čemu je $\boldsymbol{S}$ procijenjena sa $\mathbf{\Sigma}$ [slijedi iz 6]

$$\boldsymbol{S} = \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\mathbf{\Psi}}. \tag{8}$$

Gdje je $\boldsymbol{S}$ uzoračka korelacijska matrica dobivena iz slučajnog uzorka $\boldsymbol{x_1, \ldots, x_n}$, a definirana sa

$$\boldsymbol{S} = \frac{1}{1-n}\sum_{i=1}^{n}(x_i - \overline{x})^T(x_i - \overline{x}).$$

Uzoračka korelacijska matrica se dijagonalizira, odnosno pronalazi spektralna dekompozicija:

$$\boldsymbol{S} = \boldsymbol{\Gamma\Lambda\Gamma}^T$$

gdje je $\boldsymbol{\Lambda}$ dijagonalna matrica na čijoj dijagonali su svojstvene vrijednosti matrice $\boldsymbol{S}$, $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_p)$. Stupci matice $\boldsymbol{\Gamma}$ normirani su svojstveni vektori matice $\boldsymbol{S}$, označeni s γ_i. Prema (8) zna se da vrijedi

$$\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T = \boldsymbol{S} - \hat{\boldsymbol{\Psi}}$$

Pretpostavimo da je broj faktora u modelu jednak k, definiramo matricu $\boldsymbol{\Lambda} = diag(\lambda_1, \ldots, \lambda_k)$ s prvih k najvećih svojstvenih vrijednosti matrice $\boldsymbol{S}$ i matricu $\Gamma = (\gamma_1, \ldots, \gamma_k)$ s pripadnim svojstvenim vektorima. Procjenjujemo matricu $\boldsymbol{L}$,

$$\hat{\boldsymbol{L}} = [\sqrt{\lambda_1}\gamma_1, \ldots, \sqrt{\lambda_p}\gamma_p],$$

dijagonalni elementi matrice $\boldsymbol{S} - \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T$ su procijenjene vrijednosti specifičnih varijanci $\hat{\psi}_i = s_{ii} - \sum_{j=1}^{k} \hat{l}_{ij}^2$. Dobiva se aproksimacija matrice jedinstvene varijance u faktorskom modelu $\hat{\boldsymbol{\Psi}} = diag(\hat{\Psi}_1, \ldots, \hat{\Psi}_p)$.
Prema definiciji, dijagonali elementi matrice $\boldsymbol{S}$ isti su kao dijagonalni elementi $\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\Psi}}$, a nedijagonalni elementi ne trebaju nužno biti procijenjeni.

Pogledajmo i rezidualnu matricu modela:

$$\boldsymbol{S} - (\hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T + \hat{\boldsymbol{\Psi}})$$

čiju veličinu elemenata možemo ograničiti

$$\sum_{i,j}(\boldsymbol{S} - \hat{\boldsymbol{L}}\hat{\boldsymbol{L}}^T - \hat{\boldsymbol{\Psi}})_{ij}^2 \leq \lambda_{k+1}^2 + \ldots + \lambda_p^2.$$

Vidi se kako je suma kvadrata s lijeve strane manja ili jednaka sumi kvadrata zanemarenih svojstvenih vrijednosti matrice $\boldsymbol{S}$.

4 Kreditna sposobnost

Kreditni poslovi spadaju u najvažnije osnovne bankarske poslove, od kamata banke ostvaruju prihode koji čine najveći udio u strukturi ukupnih prihoda banaka. Iznimno im je važno predvidjeti kreditnu sposobnost zajmotražitelja. Kreditna sposobnost poduzeća utvrđuje se subjektivno i korištenjem kreditnog skoriranja. [2]

Još prije nekoliko desetljeća banke su kod procjene kreditne sposobnosti uzimale u obzir ocjenu zajmotražitelja korištenjem subjektivne ocjene. Bankarski službenici, oslanjali su se na subjektivne procjene rizičnosti, procjenjujući kapital poduzeća, karakter osobe, stečeno znanje, iskustvo ili jednostavno osjećaj. Kreditno sposobne smatrali su osobe koje su već neko vrijeme imale račun u banci. Oni klijenti, koji su duže bili klijenti banke, imali su prednost kod dobivanja kredita. Takav proces bio je vremenski dugotrajan, skup te često nepouzdan.

Kroz godine, financijske i kreditne institucije počele su razvijati različite sustave koji pomažu u donošenju odluke: odobriti nekom poduzeću kredit ili ne. [2] Počinju se uvoditi modeli kreditnog skoriranja koji znatno olakšavaju proces donošenja odluka o odobravanju kredita, a više o modelima kreditnog skoriranja ćemo objasniti u nastavku.

4.1 Kreditno skoriranje

Model kreditnog skoringa proces je kojim se određuje kolika je vjerojatnost da klijent kasni u otplatama kredita. Preciznije, to je sustav dodjeljivanja bodova zajmotražitelju kako bi se dobila brojčana vrijednost koja pokazuje koliko je vjerojatno da zajmotražitelj, u odnosu na ostale, iskusi neki događaj, primjerice kasni u otplati kredita. [11, strana 89]

Prema Thomasu, Edelmanu i Crooku model kreditnog skoriranja jedna od najuspješnijih primjena statistike i operacijskih istraživanja u financijama i bankarstvu, a broj analitičara za skoriranje u industriji stalno raste. [4]

Primarni cilj modela kreditnog skoriranja je odvojiti poduzeća s niskim kreditnim rizikom (dobri klijenti) od onih s visokim kreditnim rizikom (loši klijenti). [11, strana 72]

4.1.1 Povijest kreditnog skoriranja

Prvi modeli bili su univarijatni, zasnivali su se na računovodstvenim podacima. Zajmodavci bi uspoređivali različite financijske pokazatelje s industrijskim normama grane djelatnosti i tako bi banka vidjela ima li razlike od standarda u toj industriji. William Beaver 1966. godine [5, strana 13] predstavio je prvi statistički pristup za predviđanje bankrota. Izdvojio je tri od 30 omjera koji najbolje prognoziraju financijski uspjeh/neuspjeh:

1. tijek novca/ukupna imovina
2. čisti prihod/ukupni dugovi
3. tijek novca/ukupni dugovi

Za svaki omjer Beaver je izračunao prag tako da se poduzeće s omjerom iznad te vrijednosti smatralo se potencijalno uspješno, a poduzeće s manjom vrijednosti potencijalno neuspješnim.

Nakon univarijatnih, razvili su se multivarijatni modeli. Multivarijatni modeli kombiniraju ključne računovodstvene varijable kako bi se proizveo skor (mjera) kreditnog rizika koja će diskriminirati uspješna i neuspješna poduzeća. Prvi takav model bio je Altmanov Z-skor model. Taj model temelji se na financijskim pokazateljima gdje svaki od njih ima odgovarajuću vrijednost, a zbrojem tih pokazatelja dobivamo Z-skor. Ukoliko bi vrijednost poduzeća bila iznad Z-skora, zahtjev za kredit se prihvaća i obrnuto. [5]

Tijekom godina industrija potrošačkih kredita proširila je uporabu i važnost kreditnog skoriranja. Danas se modeli koriste za predviđanje ne samo kreditne sposobnosti, već i potencijalnog bankrota, prihoda, profitabilnosti, gubitka, prijevare i drugog. [11, strana 25]

Velik doprinos u kreditnom skoriranju imali su Fair, Isaac Corporation, danas poznat kao FICO, a koriste ga financijske institucije. Iako postoje i drugi sustavi bodovanja, FICO Score daleko je najčešće korišten za pojedince, koristi se u 90% odluka o zahtjevima za hipoteku u Sjedinjenim Američkim Državama. Rezultati FICO skora kreću se od 300 do 850, a rezultati u rasponu od 670 do 739 smatraju se "dobrim" kreditnim rezultatima. Dakle, ako je skor veći, rizik je manji, odnosno, ako je skor manji, rizik je veći. FICO metodologija bodovanja ažurira se s vremena na vrijeme, a najnovija verzija sada je FICO Score 10 Suite, koja je najavljena 2020. [7]

Glavni razlozi za korištenje kreditnog skoriranja [13]:

- smanjenje loših potraživanja
- poboljšanje operativne učinkovitosti - budući da je proces ocjene kreditne sposobnosti kod primjene kreditnog skoriranja automatiziran, uklonjen je ručni način procjene rizika
- omogućena veća kontrola kreditnog portfelja praćenjem procesa skoringa i karakteristika portfelja
- bolje upravljanje kreditnim rizikom, pružajući preciznije i dosljednije odluke
- smanjuje operativne troškove, štedi vrijeme kreditnim analitičarima
- implementacija konzistentnog sustava za donošenje kreditnih odluka i vođenje kreditne politike na različitim lokacijama za stanovništvo.

4.2 Model kreditnog skoriranja

Osnovni su koraci kod izgradnje kreditnog skoriranog modela su [13]:

1. Studija provedivosti - određujemo ima li koristi od samog modela koji bi se radio, definiramo troškove, razmatramo provedbu skor-kartice.
2. Definicija uzorka - prikupljamo uzorak klijenata, a za svaki od njih određujemo radi li se o dobrom, lošem, neodređenom, neaktivnom ili odbijenom zajmotražitelju.

   Definicija dobrog i lošeg klijenta temelji se na dva parametra [14]:

   - **DPD** - broj dana nakon datuma dospijeća (engleski: *days past due*), klijenti koji uplaćuju na vrijeme su dobri, a oni s visokim DPD-om su deklarirani kao loši. Svaka banka postavlja svoj prag za DPD vrijednost, uzevši u obzir neke tehničke propuste ili da je zajmotražitelj zaboravio platiti.
   - **APD** - iznos dospio nakon datuma dospijeća (engleski: *amount past due*), ovdje se moraju postaviti granice, točno definirati što se smatra dugom. Neke manje iznose nema smisla smatrati nedospjelim.

Neodređeni klijenti su oni koji ne obavljaju svoje obveze, ali ne prelaze zadani DPD prag. Odbijenim se klijentima smatraju oni koji nisu dostavili potpunu dokumentaciju, a nedovoljni (*eng. insufficient*) oni klijenti koji imaju vrlo kratku kreditnu povijest (manje od godinu dana). Postoji više definicija dobrog/lošeg klijenta. Jedan od primjera definicije dobrog/lošeg klijenta jest: klijent je loš ako kasni u plaćanju svojih obveza 90 ili više dana s nedopuštenom razlikom od 50 $ ili više, inače je klijent dobar. Neodređeni i neaktivni ne uključuju se u model, ali se kasnije uključuju u statističke rezultate.

3. Prikupljanje podataka – prikupljaju se aplikacije (zahtjevi) koje će se upotrijebiti u izgradnji skoriranog modela te se upoznaje s kvalitetom podataka. Potrebno je znati dob, spol, bankarsku prošlost, brojeve telefona, godine zaposlenosti, ako se radi o poduzeću onda i grana djelatnosti, druga imovina poduzeća, broj zaposlenih i druge karakteristike.

4. Analiza karakteristika - u ovom koraku dijele se dobri od loših klijenata, najčešće se grupiraju u razrede unutar kojih se raspoređuje broj dobrih i loših te postotak dobrih i loših zajmotražitelja.

5. Zaključivanje o odbijenima – analiziraju se prikupljeni podaci i donosi zaključak o tome kakvi bi odbijeni klijenti bili da su uključeni u model.

6. Modeliranje skor-kartice – konstruira se model tako da se postave varijable i veze među njima, za koje se zna da utječu na rizik neplaćanja, primijeni se statistička metoda i testira model. Statističke su metode, koje se najčešće koriste, logistička regresija, stablo odlučivanja, linearna regresija, diskriminacijska analiza i druge.

7. 7. Validacija skor-kartice - drugim riječima testiranje modela. Validacija je ključni dio upravljanja kreditnim rizikom, ali to je samo dijagnostički pristup koji može otkriti tek neke nedostatke modela. Model se testira primjenom različitih kvalitativnih i kvantitativnih testova, prema pravilu provodi se:

   - *out-of-sample* - testiramo model na podacima koju nisu sudjelovali u izradi modela
   - *out-of-time* - testiranje modela na podacima koji nisu iz vremena iz kojeg su podaci upotrebljeni za razvoj modela.

8. Postavljanje strategije i implementacija - kako se veze među varijablama mijenjaju, tako se i moć predviđanja modela mijenja što ima utjecaja

na kreditnu strategiju. Potrebno je nadgledati cjelokupnu situaciju i sukladno tome reagirati, zapamtiti skoriranja u određenom vremenu te njegove ulazne varijable. U ovom koraku postavljaju se granične vrijednosti kreditnog limita i ostalih parametara, provode se testiranja kako bi se utvrdili dobro postavljeni parametri. Na kraju, treba osigurati dobru komunikaciju između odjela u kojem nastaju modeli i odjela marketinga kako bi se reagiralo na promjene.

4.3 Pregled prethodnih istraživanja

Yang, Florescu,Tariqul Islam u svom radu koristili su dva algoritma kako bi otkrili koje značajke najviše utječu na kreditno skoriranje. Prvi algoritam bio je metoda glavnih komponenti, a drugi faktorska analiza. Podaci su preuzeti iz baze podataka Wharton Research Data Services za sva Standard & Poor's (S&P - američki burzovni indeks) poduzeća, svaki od skupa podataka ima svoje karakteristike koje drugačije utječu na kreditno skoriranje tako da su odlučili odabrati četiri tipična sektora: financijski, zdravstveni, energetski i potrošački sektor. Prvi algoritam bio je metoda glavnih komponenti u kojem su saželi slične značajke tako da su birali komponente s udjelom kumulativne varijance veće od 85%.
Faktorskom analizom birali su faktore s kumulativnim udjelom varijance većim od 85% koji su odabrani kao konačni faktori. Došli su do zaključka da je za većinu podataka u faktorskoj analizi trebao manji broj značajki u usporedbi s metodom glavnih komponenata kako bi se postigla gotovo ista točnost. Dakle, faktorska analiza je znatno smanjila vrijeme.
Faktori izraženi kroz financijske omjere:

- faktor 1: dug/imovina i dug/kapital
- faktor 2: zarada i profit
- faktor 3: cirkulacija gotovine
- faktor 4: likvidnost
- faktor 5: vrijednost dionica
- faktor 6: financiranje
- faktor 7: vrijednost poduzeća.

Faktori u bilanci:

- faktor 1: imovina i obveze

- faktor 2: prihod
- faktor 3: cijena opcije
- faktor 4: prilagodba nakon umirovljenja (*eng. post retirement adjustment*
- faktor 5: usklađivanje u mirovini (*eng. pension adjustment*
- faktor 6: zarada po dionici
- faktor 7: prekinuto poslovanje (*eng. discontinued operations*)
- faktor 8: povlaštene dividende (*eng. preferred dividends*)
- faktor 9: nekontrolirajući interesi (*eng. non-controlling interests*)
- faktor 10: u procesu istraživanja i razvoja.

Rezultati su pokazali da se faktorskom analizom dobije manje značajki u usporedbi s metodom glavnih komponenata s istom točnosti više od 95%.

Emela, Oralb, Reismanb i Yolalana nadogradili su kvantitativnu analizu, koju banke koriste u kreditnom skoriranju, te procijenili financijsku uspješnost klijenta banke. Njihov rad usmjeren je kao pomoć kreditnim institucijama kako bi uspješnije filtrirale poduzeća s velikim rizikom vraćanja kredita. Cijela metodologija temeljena je na analizi podataka 82 industrijska i proizvodna poduzeća koja čine kreditni portfelj jedne od najvećih banaka Turske. Pomoću financijskih pokazatelja Data envelopment analysis (DEA) napravili su jedinstvenu ocjenu financijske učinkovitosti, a rezultate su potvrdili regresijskom analizom te na temelju trenutnih saznanja o poduzeću. Njihova metodologija sastojala se od sedam koraka. Prva tri koraka bave se odabirom poduzeća i pokazateljima koji su korisni kod procjene financijskih učinaka poduzeća, nakon čega slijedi faktorska analiza. Peti korak određuje konačne financijske pokazatelje koje DEA koristi u idućem koraku gdje poduzeća dobivaju ocjenu vjerodostojnosti. Posljednji korak testira ocjene vjerodostojnosti putem regresijske i diskriminantne analize. Usmjerit ćemo se na četvrti korak u kojem se faktorskom analizom smanjuje skup podataka grupiranjem sličnih varijabli. U tom koraku faktorska analiza pomaže odlučiti koji omjeri trebaju biti uključeni u algoritam skoriranja kako bi se spriječila multikolinearnost među varijablama. Analitičarima omogućuje razmatranje više različitih financijskih dimenzija i svođenje na manji broj smislenih faktora.
Neki se omjeri odbacuju zbog savršene korelacije s drugim faktorima, dok su neki izbačeni kako bi zadovoljili zahtjeve faktorske analize.

Na početku istraživanja bilo je približno 100 poduzeća čiji su podaci bili dostupni iz različitih sektora kao što su proizvodi šumarstva, elektrotehnika, uređaji, hrana, proizvodi od papira, kemikalije, strojevi i razne druge. Ona poduzeća koja su imala nekoliko omjera koji su značajno odstupali od srednje vrijednosti bile su uklonjene. Na kraju su preostala 82 poduzeća koja su u trećem koraku rezultirala s 42 omjera. Faktorskom analizom nije nametnuto ograničenje broja faktora. Napravljena je ortogonalna varimax rotacija i dobiveno je 7 faktora:

- faktor 1: zajmovi
- faktor 2: dugotrajna imovina
- faktor 3: profitabilnost
- faktor 4: zaduženost
- faktor 5: obveze
- faktor 6: prodaja i troškovi
- faktor 7: likvidnost.

Zaključili su kako diskriminantna analiza i DEA istog skupa podataka daju slične rezultate. Konkretno, *dobre* tvrtke imaju veću likvidnost, manje bankovnih zajmova, veći kapital i bolju ravnotežu između vlastitog kapitala i dugotrajne imovine.

5 Empirijski dio: Primjena faktorske analize u kreditnom skoriranju malih i srednjih poduzeća

U ovom dijelu rada provodi se faktorska analiza na stvarnim podacima. Baza podataka sastoji se od 24 varijable i 6414 podataka, reprezentativan uzorak malih i srednjih poduzeća u Hrvatskoj koji obuhvaća podatke financijskih izvješća i podatke o insolventnosti. Kako je baza podataka velika, velik je i broj nedostajućih vrijednosti, njih 10393. Problem nedostajućih vrijednosti riješen je zamjenom srednjom vrijednosti varijable.

5.1 Opis varijabli

Najprije se upoznaju varijable. Pokazatelji likvidnosti govore koliku sposobnost ima poduzeće da podmiri svoje kratkoročne obveze. Važni su kod donošenja odluke podmirivanja dugova prema državi, kreditorima, dobavljačima i slično. Pokazatelji aktivnosti pokazuju koliko brzo cirkulira imovina u poslovnom procesu. Pokazatelji zaduženosti pokazuju strukturu kapitala i izvore financiranja. Poduzeća s visokim stupnjem zaduženosti mogu imati probleme s pronalaskom novih investitora, a može doći i do bankrota. Pokazatelji ekonomičnosti pokazuju odnos između ostvarenih prihoda i rashoda. Pokazatelji profitabilnosti pokazuju koliki dio ukupnih prihoda društvo zadržava u obliku neto dobiti, odnosno dobiti nakon poreza. Sljedeće tablice prikazuju varijable:

x_1	Koeficijent tekuće likvidnosti
x_2	Koeficijent ubrzane likvidnosti
x_3	Koeficijent trenutne likvidnosti
x_4	Gotovina prema prodaji

Tablica 1: Pokazatelji likvidnosti

x_5	Koeficijent zaduženosti
x_6	Odnos duga i kapitala
x_7	Koeficijent vlastitog financiranja
x_8	Ukupni krediti prema imovini
x_9	Ukupne obveze prema EBITDA

Tablica 2: Pokazatelji zaduženosti

x_{10}	Koeficijent obrta ukupne imovine
x_{11}	Koeficijent obrta dugotrajne imovine
x_{12}	Koeficijent obrta kratkotrajne imovine
x_{13}	Trajanje naplate potraživanja
x_{14}	Trajanje kreditiranja dobavljača
x_{15}	Prodaja prema obrtnom kapitalu

Tablica 3: Pokazatelji aktivnosti

x_{16}	Ekonomičnost ukupnog poslovanja

Tablica 4: Pokazatelji ekonomičnosti

x_{17}	Neto rentabilnost imovine
x_{18}	Neto marža profita
x_{19}	Neto dobit prema prodaji
x_{20}	Zadržana dobit prema prodaji

Tablica 5: Pokazatelji profitabilnosti

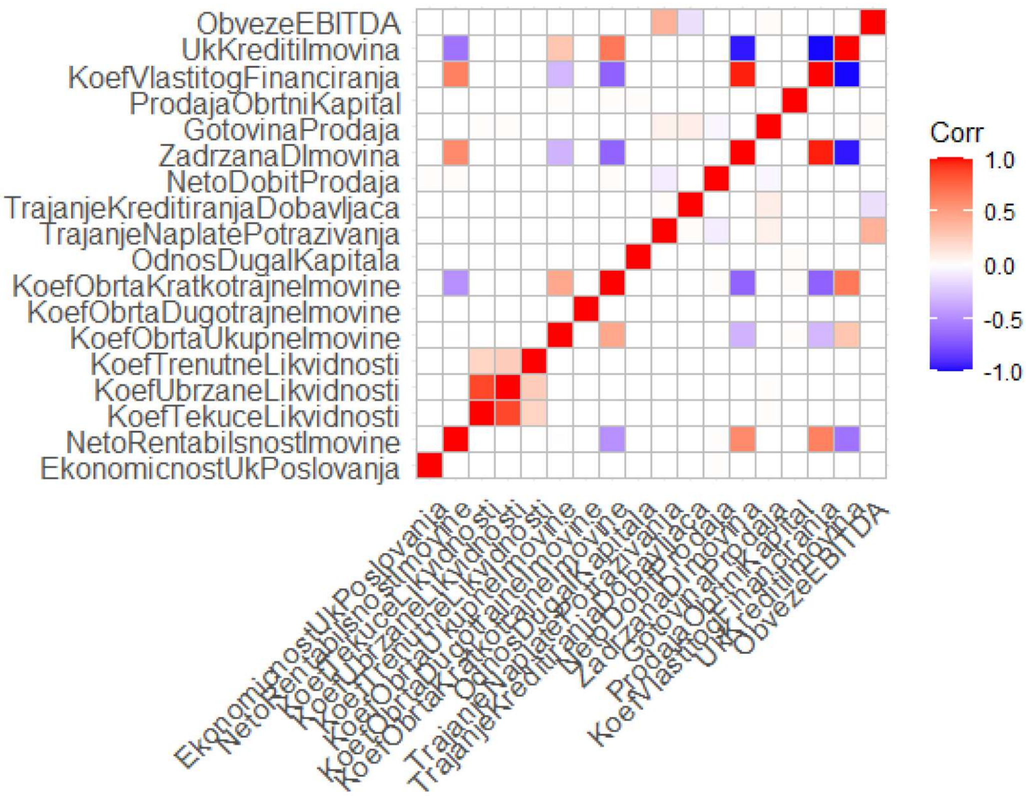
x_{21}	Je li poduzeće dobro ili loše
x_{22}	Djelatnost poduzeća
x_{23}	Županija

Tablica 6: Ostale varijable

Faktorska analiza zahtjeva samo numeričke varijable, tako da se kategorijalne varijable izbacuju iz ove primjene. Dakle, koriste se varijable od x_5 do x_{22}.

5.2 Matrica korelacije

Kao što je već spomenuto u teorijskom dijelu, da bi koristili faktorsku analizu najprije je potrebno analizirati povezanost među varijablama. Pogledajmo matricu korelacije.



Slika 2: Matrica korelacije među varijablama

Već iz matrice korelacije vidimo kako su koeficijenti likvidnosti povezani što bi naslućivalo na jedan faktor. Ostale varijable nisu previše korelirane, korelacije su većim dijelom oko 0 što i nije u korist faktorskoj analizi.

Povezanost među varijablama provjerit ćemo i Kaiser-Mayer Olkinovom mjerom, pogledajmo KMO vrijednosti u sljedećoj tablici.

Ekonomičnost ukupnog poslovanja	0.54
Neto rentabilnost Imovine	0.45
Koeficijent tekuće likvidnosti	0.72
Koeficijent ubrzane likvidnosti	0.66
Koeficijent trenutne likvidnosti	0.88
Koeficijent obrta ukupne imovine	0.59
Neto marža profita	0.45
Koeficijent obrta dugotrajne imovine	0.17
Koeficijent obrta kratkotrajne imovine	0.55
Odnos duga i kapitala	0.50
Trajanje naplate potraživanja	0.49
Trajanje kreditiranja dobavljača	0.41
Neto dobit i prodaja	0.43
Zadržana imovina	0.78
Gotovina i prodaja	0.55
Prodaja obrtni kapital	0.34
Koeficijent vlastitog financiranja	0.77
Ukupni krediti prema imovini	0.97
Ukupne obveze prema EBITDA	0.49

Tablica 7: KMO vrijednosti

Ranije je spomenuto kako KMO vrijednosti trebaju biti veće od 0.50, vidimo kako neke varijable imaju neprihvatljiv koeficijent, no mjera primjerenosti uzorka iznosi 0.73 pa uzorak smatramo dobrim za provedbu faktorske analize.

Preostaje još vidjeti p-vrijednost Barlettovog testa. Provođenjem testa dobivamo vrlo malu p-vrijednost $2.2 \cdot 10^{-16}$ i može se zaključiti da je korelacijska matrica statistički značajno različita od jedinične matrice. Odnosno, možemo primijeniti faktorsku analizu.

5.3 Odabir broja faktora

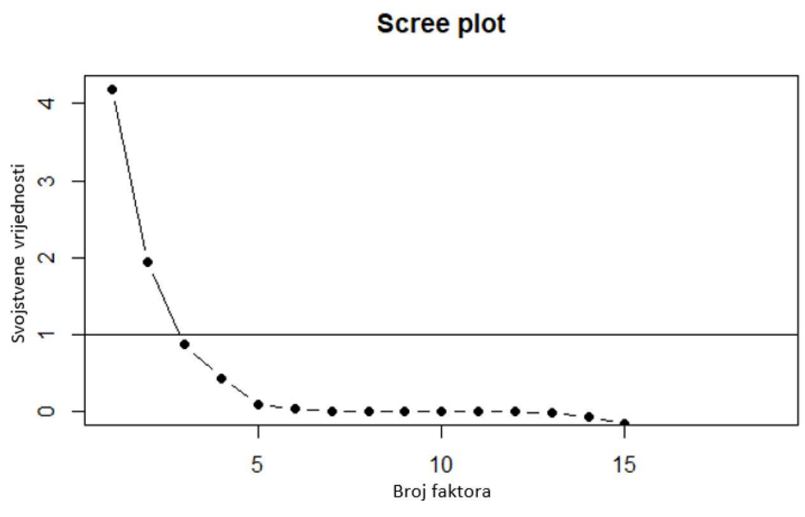
Postoji nekoliko kriterija za određivanje broja faktora, jedna od njih je Guttman-Kaiserov kriterij koji određuje broj faktora prema svojstvenoj vrijednosti matrice korelacije. Svaka komponenta čija svojstvena vrijednost veća od jedan će se zadržati. Pogledajmo u tablicu s pripadnim svojstvenim vrijednostima.

4.276384	2.942072	1.759498	1.432658e	1.090223
1.001717	1.000377	0.9993771	0.9979970	0.9864181
0.9256010	0.7581702	0.5656294	0.2036126	0.4782531
0.1010354	0.1772920			

Tablica 8: Svojstvene vrijednosti matice korelacije

Vidi se kako je sedam svojstvenih vrijednosti veće od jedan, a njih tri jako blizu jedan. Dakle, prema Guttman-Kaiserovom kriteriju imamo sedam faktora. Maknu li se ova tri jako blizu jedinici dobije se četiri faktora. Guttman-Kaiserov kriterij često daje prevelik broj faktora, pa pogledajmo i druge metode određivanja broja faktora.

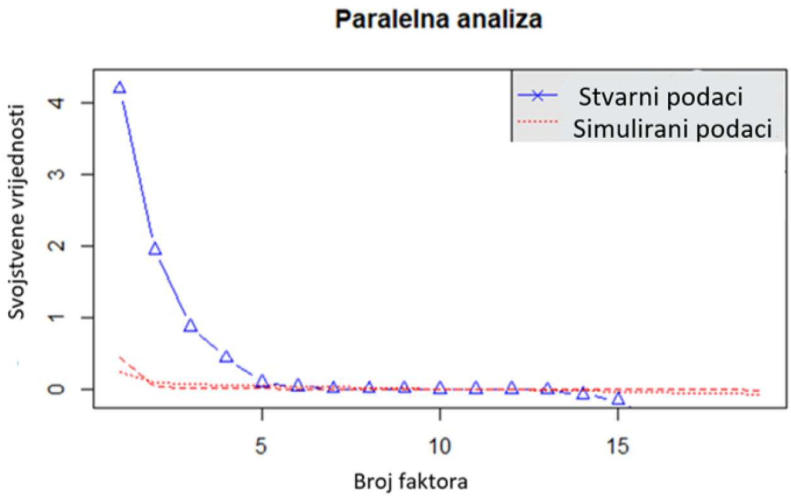
Metoda vezana uz prethodnu naziva se Catellov scree plot, jednostavna metoda koja grafički prikazuje svojstvene vrijednosti na osi ordinata i broj faktora na osi apcisa.



Slika 3: Cattelov scree test

Slika prikazuje kako su dvije komponente veće od jedan te je jedna vrlo blizu jedan. Zaključujemo da bi imali dva ili tri faktora.

Slijedi paralelna analiza. Paralelnom analizom uspoređuju se svojstvene vrijednosti sa svojstvenim vrijednostima dobivenih na osnovu slučajnih podataka, odnosno očekivanih slučajnih vrijednosti. Prema ovom kriteriju dobiva se šest faktora što se vidi na Slici 5.



Slika 4: Grafički prikaz paralelne analize

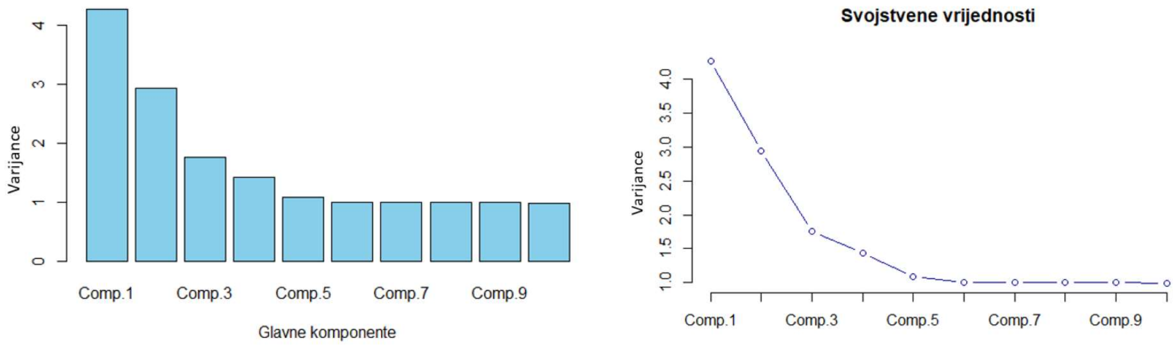
Ranije je spomenuta metoda glavnih komponenata. Pogledajmo Tablicu 9 koja nam prikazuje prvih deset glavnih komponenata.

	P1	P2	P3.	P4	P5
Standardna devijacija	2.0679419	1.7152470	1.32646065	1.19693704	1.04413746
Proporcija varijance	0.2250728	0.1548459	0.09260515	0.07540307	0.05738016
Kumulativna proporcija	0.2250728	0.3799187	0.47252387	0.54792694	0.60530710

	P6	P7	P8	P9	P10
Standardna devijacija	1.00085827	1.00018847	0.9996885	0.99899799	0.99318584
Proporcija varijance	0.05272196	0.05265142	0.0525988	0.05252616	0.05191674
Kumulativna proporcija	0.65802906	0.71068048	0.7632793	0.81580544	0.86772218

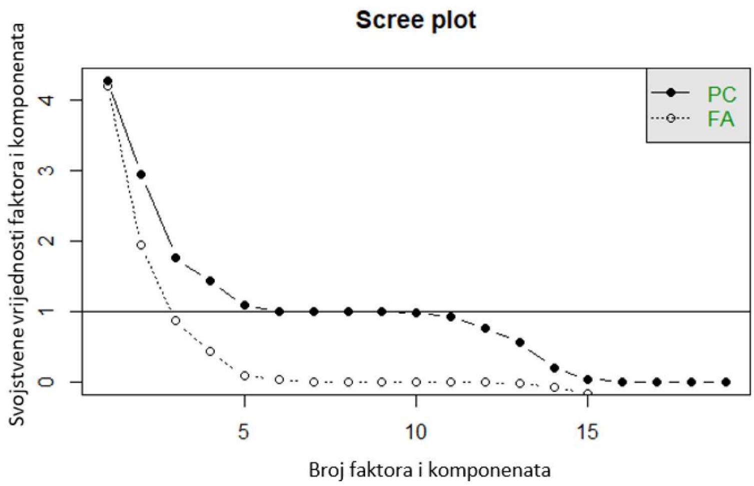
Tablica 9: Proporcije glavnih komponenti

U tablici se vidi prvih deset varijabli te njihove standardne devijacije, proporcije varijance i kumulativna proporcija. Kada bi se birao broj faktora prema metodi glavnih komponenata, zaključuje se da bi trebalo biti sedam faktora. Drugi redak tablice govori koliki je postotak objašnjenosti ukupne varijacije podataka. Tako se vidi da prva glavna komponenta objašnjava 22.51% ukupne varijacije podataka, druga objašnjava 15.48% i tako dalje. Sve glavne komponente objašnjava 100% varijacije podataka. Kako to izgleda grafički, vidi se na Slici 6.



Slika 5: Proporcije varijanci glavnih komponenata

Grafički prikazi na Slici 6 prikazuju kako prve dvije komponente imaju visoku varijancu te počnu padati od treće komponente. Slika 7 uspoređuje prethodne dvije metode scree testom.



Slika 6: Scree test, usporedba metode glavnih komponenata i Guttman-Kaiserovog kriterija

Može se uočiti da broj faktora može biti od dva faktora do sedam. Ranije je spomenut cilj faktorske analize, odnosno želi se pronaći manji broj faktora koji opisuju što veći dio varijabilnosti. Uočava se redom: dva faktora objašnjavaju 37.3% varijabilnosti bez preklapanja koeficijenata. Drugim riječima, svaki koeficijent pripada točno jednoj varijabli, a ta varijabla točno jednom faktoru. Tri faktora objašnjavaju 42.4% varijabilnosti s istim koeficijentima u dvije varijable. Kod četiri faktora varijabilnost iznosi 50.2%, no jedna varijabla ponavlja se u tri faktora. Faktorska analiza korištenjem pet, šest i sedam varijabli ne daje rezultat. Iako se s četiri faktora opisuje veća varijabilnost, faktorska analiza s tri faktora jednoznačno je određena. U nastavku će se vidjeti faktorska analiza s tri faktora.

5.4 Opis faktora

U poglavlju 3.3 spomenute su rotacije faktorskih koeficijenata koje se rade radi lakše interpretacije. Jedna od najčešćih je varimax rotacija [8] iskorištena u ovom radu. Nakon varimax rotacije svaka izvorna varijabla teži biti povezana s jednim, ili manjim brojem faktora. Dakle, u rezultatu se dobiju visoki faktorski koeficijenti za manji broj istaknutih varijabli te niske faktorske koeficijente za ostatak varijabli te nam tako olakšava interpretaciju. Uz varimax rotaciju i pretpostavku da imamo tri faktora provodimo faktorsku analizu.

Dobiveni su sljedeći faktori:

Faktor 1

Koeficijent zaduženosti	0.989
Zadržana dobit prema prodaju	-0.998
Ukupni krediti i imovina	0.987

Tablica 10: Faktor 1 i pripadni faktorski koeficijenti

Faktor 2

Koeficijent tekuće likvidnosti	0.957
Koeficijent ubrzane likvidnosti	-0.207
Koeficijent trenutne likvidnosti	-0.281

Tablica 11: Faktor 2 i pripadni faktorski koeficijenti

Faktor 3

Neto rentabilnost imovine	0.957
Koeficijent obrta ukupne imovine	-0.207
Koeficijent obrta kratkotrajne imovine	-0.281

Tablica 12: Faktor 3 i pripadni faktorski koeficijenti

Ova tri faktora objašnjavaju 42.4% ukupne varijance.

Prvi faktor orijentiran je na imovinu, financiranje i profit. Govori kako poduzeće financira svoju imovinu, koliko se financira iz tuđih izvora te u kojoj mjeri poduzeće vraća uložen kapital. Drugi je faktor likvidnost poduzeća. Taj je faktor logičan, naslućen još iz matrice kovarijanci. Treći faktor orijentiran je na imovinu i obrtaje, a opisuje kakva je cirkulacija imovine u poslovnom procesu.

Koristeći se faktorskom analizom i dobivenim rezultatima identificirat će se predstavnika faktora.

5.5 Model skoriranja logističkom regresijom

Nakon odabira faktora, napravit će se model skoriranja kako bi se procijenila vjerojatnost insolventnosti pri čemu se kao procesiranje koristila faktorska analiza. Model skoriranja u radu napravljen je metodom logističke regresije. Popularna i česta metoda za klasifikaciju jedinice za dani set nezavisnih varijabli koje mogu biti numeričke ili kategorijalne.

Ova metoda modeliranja jednostavna je za primjenu, daje dobre rezultate, koji su razumljivi i lako se mogu interpretirati, te pretpostavke za njenu primjenu nije teško zadovoljiti. Jedna od pretpostavke je zavisna varijabla $0-1$, odnosno slučajna varijabla koja poprima samo dvije vrijednosti. U ovom radu to je varijabla *Dobar Loš* koja poprima vrijednost 1 za poduzeća koja kasne s plaćanjem, u nastavku *loša poduzeća*, te vrijednosti 0 ako poduzeće ne kasni s plaćanjem, *dobra poduzeća.*
Nadalje, podaci za svaku jedinicu moraju biti prikupljeni neovisno, kako model ne bi sadržavao multikolinearnost te da imamo velik uzorak. Ranije je spomenuto kako skup sadrži 6414 podataka poduzeća različitih djelatnosti. Dakle, veliki uzorak i nezavisnost su zadovoljeni, preostaje još pretpostavka o multikolinearnosti.

Uzorak čini jednak broj dobrih i loših poduzeća, točnije njih 3207.
Kod odabira modela mora se vidjeti ima li uopće smisla napraviti takav model. Isprobavajući više proizvoljnih modela, odabran je model koji zadržava pokazatelje zaduženosti, likvidnosti i profitabilnosti, a zadovoljava pretpostavke logističke regresije. Scoring model napravljen je od četiri varijable:

- Koeficijent zaduženosti
- Koeficijent tekuće likvidnosti
- Neto rentabilnost imovine
- Gotovina prema prodaji.

Ispitat ćemo problem multikolinearnost koristeći faktor inflacije varijance (*eng. Variance Inflation Factor*). Za varijable s VIF vrijednosti manjom od pet zaključujemo da nisu međusobno kolinearne.

Koeficijent zaduženosti	1.081027
Koeficijent tekuće likvidnosti	1.000006
Neto rentabilnost imovine	1.081026
Gotovina prema prodaji	1.000005

Tablica 13: Multikolinearnost

Promotri li se tablica multikolinearnosti, vidi se kako nema izražena problema multikolinearnosti, dakle odabrane varijable mogu činiti jedan model.

Sljedeća tablica prikazuje regresijske koeficijente.

	Estimate	Std. Error	z-value	Pr(>|Z|)
(Intercept)	-0.007252	0.02508	-0.289	0.772
Koeficijent zaduženosti	0.0005915	0.0002621	2.257	0.024
Koeficijent tekuće likvidnosti	-0.000001856	0.000002936	-0.632	0.527
Neto rentabilnost imovine	-0.00000004707	0.00000005522	-0.853	0.394
Gotovina prema prodaji	0.000009608	0.00002293	0.419	0.675

Tablica 14: Regresijski koeficijenti

Iz Tablice 14 može se vidjeti ako se koeficijent zaduženosti poveća za jedan, logaritam šansi da poduzeće postane loše povećava se za 0.0005915. Dobiven je logičan slijed, odnosno, što je veći odnos duga i imovine, povećava se i financijski rizik od nevraćanja dugova. Poduzeće koje u prevelikoj mjeri koristi zaduživanje kao oblik financiranja, s vremenom neće biti u mogućnosti podmirivati potraživanja. Drugim riječima, poduzeće će postati loše.
Povećanjem koeficijenta tekuće likvidnosti za jedan, smanjuje se logaritam šansi da poduzeće postane loše. Ponovno se dobije razumljiv slijed. Poduzeće, koje poveća sposobnost vraćanja kratkoročnih obveza, postaje stabilnije te se smanjuje mogućnost da postane loše. Nadalje, koeficijent uz varijablu neto rentabilnost imovine negativan je, što znači da se jediničnim povećanjem neto rentabilnosti imovine smanjuje šansa da poduzeće postane loše. Neto rentabilnost imovine omjer je dobitka i imovine. Povećanje tog koeficijenta značilo bi i povećanje dobiti. Rastom dobiti može se zaključiti da poduzeće dobro posluje, a dobrim poslovanjem poduzeće neće imati problema pri vraćanju kredita. Zadnja je varijabla gotovina prema prodaji čijim se jediničnim povećanjem povećava šansa da poduzeće postane loše. Gotovina prema prodaji omjer je gotovine i prihoda od prodaje. Povećanje tog omjera može se protumačiti da poduzeće zadržava više gotovine nego što zaradi od prodaje, a tada ima problem s profitabilnošću. S druge strane, možda poduzeće skladišti gotovinu jer ne zna u što bi uložilo. Ponekad poduzeća s visokim omjerom gotovine prema prodaji sakupljaju novac kao mjeru predostrožnosti zbog nemogućnosti vanjskog finan-

ciranja ili kredita. Dakle, visok omjer gotovine prema prodaji može značiti da poduzeće nema profit, da se suočava s financijskim poteškoćama ili ima ograničen pristup kapitalu. Sve nabrojeno čini ga osjetljivim na insolventnost te poduzeće postaje loše.

5.6 Kvantitativna validacija modela skoriranja

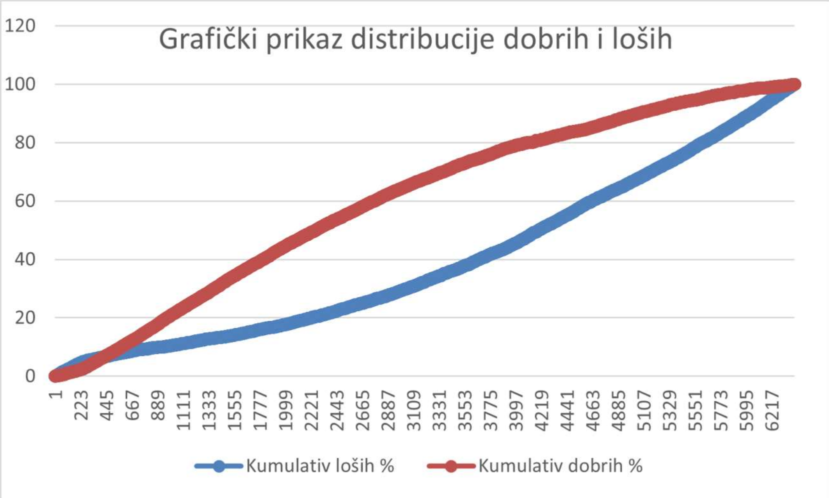
Nakon izgradnje modela, potrebno ga je i validirati različitim kvantitativnim metodama. U nastavku ćemo pogledati preciznost klasifikacije modela kroz Kolmogorov–Smirnov statistiku (KS), matricu konfuzije, ROC krivulju, greška tipa 1, greška tipa 2 i GINI koeficijent.

Cut-off vrijednost u modelu postavljen je na 0.50181 pri čemu poduzeća s vrijednošću manjom od 0.50181 procjenjujemo kao dobra (0), dok se poduzeća s dodijeljenom vrijednošću većom od 0.50181 procjenjuju kao loša (1).

Kolmogorov-Smirnov statistika (KS) označava maksimalnu apsolutnu vrijednost, a radi se tako da se napravi apsolutna razlika kumulativa dobrih i kumulativa loših. Što je ta razlika veća, to model bolje diskriminira dobre i loše. Prema vrijednosti KS statistike zaključuje se koliko je model dobar na sljedeći način [13]:

- KS $<$ 20: model vjerojatno neće raditi
- 20 - 40: osrednji model
- 41 - 50: dobar model
- 51 - 60: jako dobar model
- 61 - 75: odličan model
- KS $>$ 75: model predobar da bi bio istinit

Sljedeća slika prikazuje kumulativne frekvencije dobrih i loših poduzeća pri čemu crvena linija predstavlja kumulativne frekvencije dobrih, a plava linija kumulativne frekvencije loših poduzeća. Što je udaljenost među grafovima veća, bolja je kakvoća modela.



Slika 7: Distribucija dobrih i loših poduzeća

Kumulativna frekvencija loših raste brže od kumulativne frekvencije dobrih zato što je više loših pronađeno među niskim vrijednostima skorova. Možemo zaključiti da model dobro klasificira dobre i loše klijente. Vrijednost KS statistike u ovom modelu iznosi 35.55%, odnosno prema prethodnim kriterijima rekli bismo da model radi osrednje.

Matrica konfuzije (*eng. confusion matrix*) sastoji se od četiri mjere, a pokazuje broj poduzeća koja su ispravno ili ne ispravno klasificirana.

	predviđanja	
stvarno	loši	dobri
loši	TP	FP
dobri	FN	TN

Tablica 15: Općenita matrica konfuzije

Tablica 15 prikazuje matricu konfuzije. Potrebno je razumjeti četiri kategorije. Prva je istinito pozitivan (*eng. true positive, TP*) - broj poduzeća koja su se pokazala kao dobra, a stvarno i jesu dobra. Istinito negativna (*eng. true negative, TN*), poduzeća koja su klasificirana kao loša i takva stvarno i jesu. Osim istinitih vrijednosti postoje i greške. Ako je predviđeno da će poduzeće biti dobro, a zapravo nije, nazivamo lažno pozitivno (*eng. false positive, FP*). Na kraju, lažno negativno (*eng. false negative, FN*), kada je predviđeno da će poduzeće biti loše, a zapravo nije. Na temelju vrijednosti iz matrice konfuzije mogu se izračunati neke korisne mjere koje pomažu pri kakvoći modela.
Postotak uspješne klasifikacije mjeri se kroz:

- total hit rate - omjer ukupnog broja poduzeća ispravno klasificiranih i ukupnog broja poduzeća

- good hit rate - omjer dobrih poduzeća ispravno klasificiranih i stvarno dobrih poduzeća

- bad hit rate - omjer loših poduzeća ispravno klasificiranih i stvarno loših poduzeća

Greška tipa 1 je vjerojatnost odobravanja kredita lošem poduzeću i označava se s α, a greška tipa 2 je vjerojatnost ne odobravanja kredita dobrom poduzeću i označava se s β. Banke i druge kreditne institucije više obraćaju pažnju na grešku tipa 1 jer je ona povezana s gubitkom, dok je greška tipa 2 povezana s propuštenom mogućnošću za dobitkom. α i β se procjenjuju kao omjeri elemenata iz matrice konfuzije.

$$\alpha = \frac{FP}{TP + FP}, \beta = \frac{FN}{FN + TN}$$

α je omjer onih koji su predikcijom dobri, a zapravo su loši i ukupnog broja loših poduzeća, dok je β omjer onih koji su predikcijom loši, a zapravo su dobri i ukupnog broja dobrih poduzeća.
Pogledajmo kako to izgleda na podacima.

	predviđanja	
stvarno	loši	dobri
loši	2004	1203
dobri	874	2333

Tablica 16: Matrica konfuzije

Od ukupno 3207 dobrih poduzeća, model je njih 2004 točno procijenio kao dobra, a 1203 pogrešno kao loša. 2333 dobrih poduzeća svrstano je u dobre, a njih 874 svrstano je u loša poduzeća.
Uz matricu konfuzije lako se izračuna postotak uspješne klasifikacije.

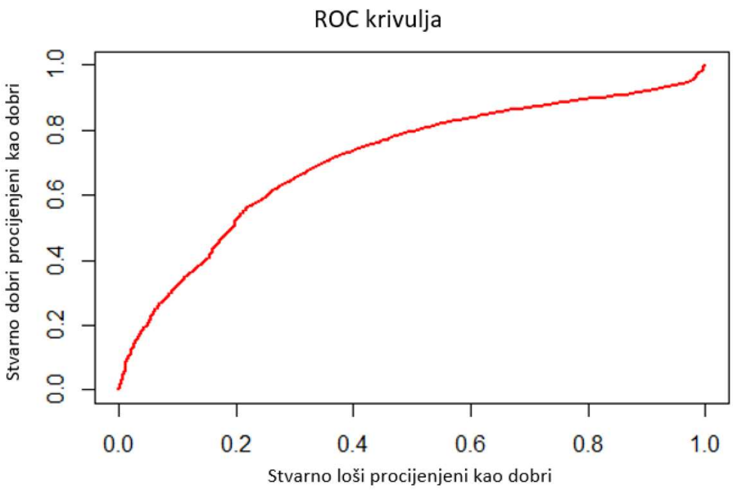
total hit rate	66.67%
good hit rate	72.75%
bad hit rate	62.52%

Tablica 17: Postotak uspješne klasifikacije

Dakle, iz Tablice 17 može se vidjeti da je od stvarno loših model pogodio 66.67% loših i od stvarno dobih, model je pogodio 72.75% dobrih. Omjer loših poduzeća ispravno klasificiranih i stvarno loših poduzeća, odnosno bad hit rate iznosi 62.52%.

Greška tipa 1 iznosi 35.51%, a greška tipa 2 jednaka je 27.25%.

ROC krivulja (*eng. Receiver Operating Characteristic curve*) pokazuje nam koliko dobro model klasificira dobra i loša poduzeća. Krivulja treba biti što konkavnija, ako je ona dijagonalna tada uopće ne razlikuje dobre i loše klijente. Odnosno, za niske vrijednosti vjerojatnosti da će poduzeće biti loše treba puno dobih i malo loših vrijednosti, a kod visokih vrijednosti vjerojatnosti da će poduzeće biti loše, mnogo loših i malo dobrih. Slika 9 prikazuje ROC krivulju.



Slika 8: ROC krivulja

Prema ROC krivulju možemo zaključiti da model dobro klasificira dobra i loša poduzeća. Računamo površinu ispod ROC krivulje koju nazivamo AUC (*eng. Area Under ROC curve*). Kada je AUC=1, klasifikator točno razlikuje sva dobra i loša poduzeća, no ako je AUC=0, tada sva loša poduzeća klasificira kao dobra i sva dobra poduzeća kao loša. Za vrijednost površine 0.5 klasifikator ne razlikuje dobra i loša poduzeća, odnosno pogađa koje je poduzeće dobro, a koje loše. Što je veća površina ispod krivulje, model je bolji.

AUC vrijednost na podacima iznosi 0.71, što nije idealno, ali je prihvatljivo.

Preostaje izračunati GINI indeks, a računa se po formuli $(2 \cdot AUC) - 1$, a poprima vrijednosti između 0% i 100%. Gdje 0% ne razlikuje dobre i loše, a 100% savršeno ih razlikuje, u ovom uzorku GINI indeks iznosi 41.62%, što je vrijednost prosječnog skoriranog modela.

6 Zaključak

Tema diplomskog rada je primjena faktorske analize u kreditnom skoriranju. Rad se sastoji od teorijskog i empirijskog dijela. U teorijskom dijelu rada objašnjen je model faktorske analize, opisano je kreditno skoriranje te je dan pregled prethodnih istraživanja u kojima je korištena faktorska analiza u kreditnom skoriranju. U empirijskom dijelu rada primijenjena je faktorska analiza na reprezentativnom uzorku malih i srednjih poduzeća u Republici Hrvatskoj u cilju identificiranja varijabli odnosno financijskih omjera koji su važni za opisivanje financijskih karakteristika poduzeća. Nakon toga je napravljen model kreditnog skoriranja upotrebom logističke regresije čija je svrha procjena insolventnosti poduzeća. Dobiveni model je validiran primjenom standardnih testova, a rezultati su pokazali dobru kvalitetu modela. Kroz ovaj rad smo prikazali da se faktorska analiza može koristiti u postupku izrade modela kreditnog skoriranja kako bi se identificirale ključne varijable odnosno faktori značajni za predikciju insolventnosti poduzeća.

Literatura

[1] T. W. ANDERSON, *An Introduction to Multivariate Statistical Analysis*, New Jersey, 2003.

[2] Z. BOHAČEK, N. ŠARLIJA, M. BENŠIĆ, *Upotreba kredit skoring modela za ocjenjivanje kreditne sposobnosti malih poduzetnika*, Ekonomski pregled, 2003.

[7] A. BURAK EMELA, M. ORALB, A. REISMANB, R. YOLALANA *A credit scoring approach for the commercial banking sector*, Socio-Economic Planning Sciences, Ujedinjeno Kraljevstvo, 2003.

[4] P. HAJEK, K. MICHALAK, *Feature selection in corporate credit rating prediction[J]*, Knowledge Based Systems, 2013.

[5] A. HANIČ, E. ZUNIĆ, A. DŽELIHODIĆ,, *Scoring models of bank credit policy managment*, Institute of Economic Sciences, Belegrade, School of Economics and Business, Internacional Burch University, Sarajevo, 2013.

[6] W. HÄRDLE, L. SIMAR, *Applied Multivariate Statistical Analysis*, Springer, Berlin, 2007.

[7] A. HAYES, Investopedia, dostupno na: `https://www.investopedia.com/terms/f/ficoscore.asp`, 2013.

[8] R. KHATTREE,*Multivariate Data Reduction and Discrimination with SAS Software*, Oakland University, 2000.

[9] N. KURNOG ŽIVADOVIĆ, *Utvrđivanje osnovnih karakteristika proizvoda primjenom faktorske analize*, Ekonomski pregled, Zagreb, 2004.

[10] K. V. MARDIA, J. T. KENT, J. M. BIBBY, *Multivariate analysis*, Academic Press, London, 1995.

[11] E. MAYS, *Handbook of Credit Scoring*, Glenlake Publishing Company, Ltd., Chicago, 2001.

[12] A. C. RENCHER, *Methods of Multivariate Analysis*, John Wiley i Sons, Brigham Young University, Kanada, 2002.

[13] N. ŠARLIJA, *Upotreba i primjena kredit skoring modela*, Predavanja iz kolegija 'Upravljanje kreditnim rizicima', 2008.

[14] *Credit scoring model validation and best practices, The teach brief*, dostupno na: https://cigniti.com/wp-content/uploads/tech-briefs/Credit-Scoring-Model-Validation-and-Best-Practices-ai-ml.pdf, Cigniti, strana 3.

Sažetak

Ovaj diplomski rad sadrži dva dijela, najprije teorijski dio u kojem se objašnjava faktorska analiza i kreditno skoriranje. Drugi je dio empirijski na kojem se na reprezentativnom uzorku podataka malih i srednjih poduzeća u Republici Hrvatskoj napravljena faktorska analiza. Opisano je kako pronaći broj faktora, rotacija faktora te na kraju i sami faktori. Slijedi skorirani model dobiven na podacima metodom logističke regresije i kvantitativna validacija modela.

Ključne riječi: faktorska analiza, kreditna sposobnost, kreditno skoriranje, logistička regresija.

Summary

This thesis contains two parts: the largest theoretical part, which explains factor analysis and credit scoring and the empirical, in which a factor analysis on a representative sample of data from small and medium-sized enterprises in the Republic of Croatia was made. There are also described methods of finding the number of factors, rotation of factors and finally the factors themselves. The following is the scored model obtained from the data using the logistic regression method and the qualitative validation of the model.

Keywords: factor analysis, creditworthiness, credit scoring, logistic regression.

7 Životopis

Rođena sam 16. siječnja 1996. godine u Virovitici. Osnovnu školu Petra Preradovića pohađala sam u rodnom selu Kladarama te više razrede u Pitomači. Nakon završetka osnovne škole upisujem Gimnaziju Petra Preradovića Virovitica, opći smjer. 2014. nastavljam školovanje na Prirodoslovno-matematičkom fakultetu u Zagrebu, gdje sam završila preddiplomski studij matematike - nastavnički smjer. Diplomski sveučilišni studij upisujem na smjeru Financijska matematika i statistika na Fakultetu primijenjene matematike i informatike u Osijeku.